Sequence-Specific ^{1}H NMR Assignment and Secondary Structure of the Arc Repressor of Bacteriophage P22, As Determined by Two-Dimensional ^{1}H NMR Spectroscopy†

J. N. Breg, R. Boelens, A. V. E. George, and R. Kaptein*
Department of Chemistry, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands
Received May 4, 1989; Revised Manuscript Received July 20, 1989

ABSTRACT: The Arc repressor of bacteriophage P22 is a DNA binding protein that does not belong to any of the known classes of such proteins. We have undertaken a ^{1}H NMR study of the protein with the aim of elucidating its three-dimensional structure in solution and its mode of binding of operator DNA. Here we present the ^{1}H nuclear magnetic resonance (NMR) assignments of all backbone protons and most of the side-chain protons of Arc repressor. Elements of secondary structure have been identified on the basis of networks of characteristic sequential and medium-range nuclear Overhauser enhancements (NOEs). Two α-helical regions have been found in the peptide regions 16–29 and 35–45. The ends of the helices could not yet be firmly established and could extend to residue 31 for the first helix and to residue 49 for the second. Immediately before the first helix, between residues 8 and 14, a region is present with β-sheet characteristics dominated by a close proximity of the α-protons of residues 9 and 13. Because of the dimeric nature of the protein there are still two possible ways in which the NOEs in the β-sheet region can be interpreted. If the NOEs are intramonomer, this requires a tight turn involving residues 10–12. Alternatively, if the NOEs are intermonomer, then an antiparallel β-sheet would be implicated comprising two strands of different Arc monomers. While the data presently do not allow an unambiguous choice between these two possibilities, some evidence is discussed that favors the latter (β-sheet between monomers). Since the N-terminal region of Arc is responsible for the sequence-specific recognition of its operator, our findings suggest the existence of a DNA binding motif in which a β-sheet region is present.

The Arc and Mnt repressors of the *Salmonella* bacteriophage P22 control the expression of a third regulatory protein, P22 antirepressor (Susskind, 1980; Sauer et al., 1983; Vershon et al., 1985). Arc repressor is a dimer (53 residues/monomer) and is about 40% homologous to Mnt repressor (82 residues/monomer). It is therefore likely that they are structurally related (Sauer et al., 1983; Vershon et al., 1985). CD measurements have indicated that both proteins have a high α-helix content (Vershon et al., 1985). However, neither Arc nor Mnt displays strong homology with the conserved helix–turn–helix regions that form the DNA binding sites of many phage and bacterial DNA binding proteins (Pabo & Sauer, 1984). After cloning the *arc* and *mnt* genes of bacteriophage P22 into overproducing *Escherichia coli* strains (Vershon et al., 1985), it was possible to start biophysical studies toward their respective structures by X-ray crystallography and NMR.[1] Crystals of Arc have been grown, but the three-dimensional structure has not yet been solved (Jordan et al., 1985).

Arc and Mnt appear to be members of a new class of sequence-specific DNA binding proteins different from the helix–turn–helix proteins, as suggested by a number of findings (Knight et al., 1989). First, Arc scores very low in a systematic search for sequence homology with DNA binding proteins that do contain the helix–turn–helix structural motif [Dodd & Egan, 1987; also cf. Pabo and Sauer (1984)]. Second, from an analysis of a large number of mutant Arc proteins, it has been shown that mutations in the N-terminal 10 residues of Arc cause a large decrease in binding to operator DNA, without affecting the ability of Arc to fold into a stable three-dimensional structure (Vershon et al., 1986; Bowie & Sauer, 1989). Also, an Mnt mutant with proline substituted for histidine at position 6, which corresponds to position 9 in Arc, has lost the ability to bind the wild-type operator but is able to bind a mutant operator with a change of the third operator base pair (Youderian et al., 1983). Moreover, a hybrid protein containing the nine N-terminal residues of Arc fused to the C-terminal region of Mnt has the operator binding specificity of Arc (Knight & Sauer, 1989). From these findings it appears that a significant part of the DNA binding surface of Arc repressor is formed by residues at the N-terminus, which lack homology with helix–turn–helix proteins.

We have embarked on an NMR study of Arc repressor with the aim, first, to determine the three-dimensional structure of the protein and in particular its DNA binding site.[2] The ultimate goal is to elucidate the mode of sequence-specific DNA recognition of Arc and Mnt repressors. For NMR studies of repressor–operator complexes [cf. Boelens et al. (1987)], the ^{1}H resonance assignments of both protein and DNA are a prerequisite. In the present paper we present the sequence-specific ^{1}H NMR assignment of all backbone protons and of the major part of the side-chain protons of Arc repressor. These assignments provide sufficient basis for interpretation of the sequential and medium-range NOEs in

† This work was supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Research (NWO).
* Address correspondence to this author.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; 2D, two dimensional; NOESY, 2D NOE spectroscopy; COSY, 2D shift-correlation spectroscopy; DQF-COSY, double-quantum-filtered COSY; TPPI, time-proportional phase increment; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; CD, circular dichroism.

[2] A related NMR study on a mutant of Arc repressor ($Pro^8 \rightarrow Leu$) has recently been carried out by D. J. Patel and his co-workers (Zagorski et al., 1989).

0006-2960/89/0428-9826$01.50/0

terms of the secondary structure elements of the protein. It is shown that the N-terminus of Arc has an extended backbone structure, including a region with β-sheet characteristics, while toward the C-terminus two α-helices are present, separated by a loop. Our results are in agreement with secondary structure predictions based on the analysis of mutants of Arc that are neutral with respect to protein folding (Bowie & Sauer, 1989).

In NMR studies of dimeric proteins such as Arc repressor there is an essential difficulty related to the interpretation of NOEs. A priori, an observed NOE can be interpreted as either between protons in the same monomer or between different monomers, and no easy distinction can be made. However, for the sequential assignment (Wüthrich, 1986) this does not seem to be a problem. For instance, the probability of an NOE between an amide and a C_α proton being a sequential NOE (vs a long-range one) is not affected by the oligomeric nature of the protein. Also, in regions of regular secondary structure such as α-helices, observations of networks of short-range (d_{NN}, $d_{\alpha N}$) and medium-range connectivities, such as $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, and $d_{\alpha\beta}(i,i+3)$, make the interpretation of these NOEs as intramonomer extremely likely. However, for long-range NOEs the ambiguity remains. For example, in the case of the Arc respressor we observed a strong NOE between the C_α protons of residues Gln^9 and Arg^{13}. If this is an intramonomer NOE, it would imply a tight turn involving residues 10–12, but if it would be intermonomer, the NOE would lead to an extended β-sheet involving residues of different monomers. Although at present no clear distinction can be made, some evidence will be presented that favors the latter possibility.

Materials and Methods

Sample Preparation. An initial sample of 10 mg of Arc repressor was obtained as a gift from Prof. R. T. Sauer (MIT, Cambridge). The complete sequential assignments were performed with this sample. Subsequently, Arc repressor was isolated from *E. coli* strain X90 containing plasmid pTA200, as described by Vershon et al. (1987), and was extensively dialyzed against sodium phosphate buffer at pH 5.0 and then lyophilized. The repressor was dissolved in D_2O or in H_2O containing 5% (v/v) ^2H_2O for the deuterium lock and 50 mM sodium phosphate, at a concentration of 4 mM (monomer), and the pH was adjusted to 4.7–5.0. The protein was lyophilized between measuring in H_2O and D_2O and vice versa.

NMR Spectroscopy. All two-dimensional NMR spectra were recorded on a Bruker AM 500 or a Bruker AM 600 spectrometer, the latter at the SON hf NMR facility, Department of Biophysical Chemistry, Nijmegen University, The Netherlands. Both spectrometers were interfaced with an Aspect 3000 computer. The data were processed on a μVAX II with the "2D NMR" software library written in FORTRAN 77, developed in our laboratory. All two-dimensional spectra were recorded in the pure-phase absorption mode by application of TPPI (Marion & Wüthrich, 1983). The HDO signal was suppressed by presaturation during the recycle delay and during the mixing time of the NOESY experiments. Unless otherwise indicated, the recording temperature was 30 °C.

HOHAHA spectra (Braunschweiler et al., 1983; Davis & Bax, 1985) at 500 and 600 MHz were recorded with a MLEV-17 mixing sequence (Bax & Davis, 1985), with total mixing times of 20–70 ms. Usually a data set of 512K × 2K was obtained, which was processed with a sine-bell window shifted over $\pi/4$ and zero filling to obtain a spectral data matrix of 1K × 1K. Separate base-line corrections in each domain were applied after the double Fourier transformation was completed. The NOESY spectra at 500 and 600 MHz were recorded with a 32-step phase cycle (States et al., 1982) and with various mixing times between 50 and 200 ms. At 30 ms before the last 90° pulse the decoupling was interrupted for a 7-ms homospoil pulse, which resulted in an efficient suppression of the broad components of the H_2O signal. At both spectrometer frequencies a series of data sets of 512K × 2K was obtained for the various mixing times. Each set was processed with a sine-bell window shifted over $\pi/4$ and zero filling to obtain a spectral data matrix of 1K × 1K. After complete Fourier transformation of the spectrum, separate base-line corrections were applied in both dimensions.

The cross-peaks that are normally supressed together with the water line were made observable in separate 2D experiments at 500 MHz by using the SCUBA method (Brown et al., 1988) in which the decoupler is switched off prior to the start of the 2D experiments. The H_2O relaxation delay (set to 80 ms), and also the mixing time in the NOESY experiment, was interrupted in the middle by a composite 180° pulse. The resulting data sets were processed in the same way as regular NOESY and HOHAHA spectra. DQF-COSY spectra (Piantini et al., 1982) in D_2O and H_2O were recorded at 600 MHz at 37 °C. A time domain data set of 600K × 2K points was obtained, with equal spectral widths in both dimensions (7240 Hz), which amounts to a maximum t_1 of 40 ms. For t_2 a Gaussian multiplication was applied (line broadening 12 Hz), while for t_1 a sine-squared-bell window shifted by $\pi/100$ was used. Zero filling prior to the second Fourier transformation resulted in a spectrum of 1K × 2K real data points, with digital resolution 7.1 and 3.5 Hz in ω_1 and ω_2, respectively. After the first Fourier transformation separate base-line corrections were performed for the spectrum on the left- and right-hand side of the H_2O line.

Slowly exchanging protons have been identified by first lyophilizing a fully protonated Arc sample in H_2O and subsequently recording a NOESY spectrum of this sample immediately after dissolving in D_2O (~9 mM protein concentration at pH 4.65). After a time period of 12 h nearly all amide protons were fully exchanged, and the presence of residual amide signals in the spectrum, recorded within a period of 5 h, was taken as indicative for slow exchange. A NOESY spectrum was preferred to a HOHAHA spectrum, since the latter was less informative due to the weakness or absence of a number of NH–C_αH cross-peaks. In the NOESY spectrum the amide protons were easily recognized by the presence of multiple cross-peaks with aliphatic protons.

Results and Discussion

Preliminary Analysis. The 500-MHz ^1H NMR spectrum of Arc in H_2O at a protein concentration of 4 mM is presented in Figure 1. Apart from a slightly increased line width this spectrum is identical with that taken at a concentration of 0.2 mM. It had been established before that under these conditions Arc exists as a dimer (Vershon, 1985). Therefore, it appears that almost no further aggregation occurs above the level of a dimer. The assignment procedure was facilitated by a favorable chemical shift dispersion for most parts of the spectrum. This can be seen in the spectrum of Figure 1 and is also illustrated by the methyl region of the HOHAHA spectrum in D_2O (see Figure 2). In Figure 2 a number of spin systems are indicated that are recognized from a series of HOHAHA spectra with varying mixing times (assignments to specific residues were obtained at a later stage). These spin systems included four valines, recognized by pairs of signals with equal intensities for the methyl groups at the positions of C_βH and C_αH and two alanines recognized as unique spin

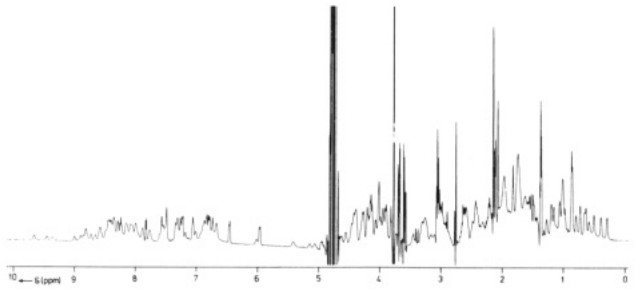

FIGURE 1: 500-MHz ^{1}H NMR spectrum of Arc repressor (4 mM) at 30 °C and pH 5.8. The spectrum contains a number of sharp signals that stem from small molecular weight contaminants, for instance, around 3.8 ppm.

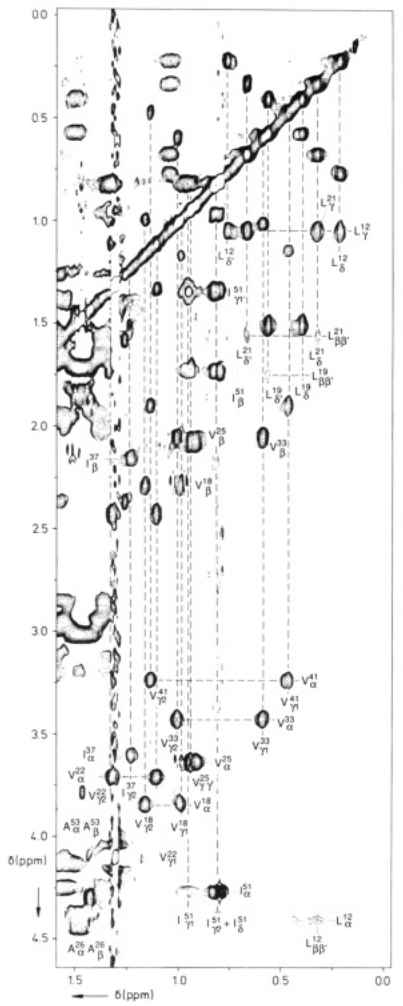

FIGURE 2: Part of a 500-MHz ^{1}H NMR HOHAHA spectrum of Arc repressor in D_2O, with mixing time of 70 ms, containing the methyl signals. The assignment of the identified resonances of the alanine, valine, leucine, and isoleucine residues is indicated.

systems with single methyl groups. The spin system for a fifth valine residue (Val^{25}) with nearly coinciding methyl signals is also included in Figure 2. From the HOHAHA spectra the methyl signals for the two leucines were also identified in a straightforward way by pairs of cross-peaks with equal intensity, by coherence transfer from the δ and δ' protons up to the β-protons (see Figure 2). For these residues no connectivities in the HOHAHA spectra could be observed between the α-protons and the methyls. The leucine spin systems could only be completed in the sequential analysis (see below). The identification of the Ile^{51} spinsystem was hampered by the overlap of the C_γ and C_δ methyl signals and could be found by comparison of the HOHAHA spectrum with the DQF-

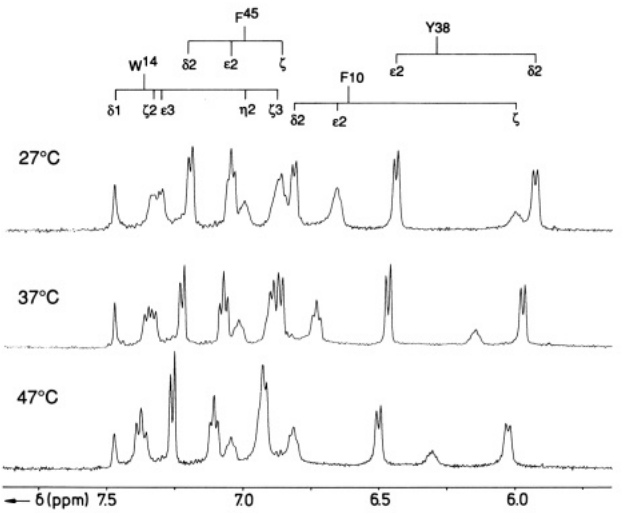

FIGURE 3: Part of the 500-MHz ^{1}H NMR spectrum of Arc repressor (~0.4 mM, pH 5.8) showing the aromatic side-chain resonances at three different temperatures. The assignment of the aromatic residues Phe^{10}, Phe^{45}, Trp^{14}, and Tyr^{38} is indicated.

COSY spectrum of the Arc repressor. For Ile^{37} only part of the spin system has been identified as yet. The present assignments for the isoleucine residues have been included in Figure 2.

Three serine spin systems and the aromatic side chain protons were identified from a HOHAHA spectrum with a short mixing time. The interpretation of the aromatic side-chain signals in the initial phase of the assignment procedure was aided by a temperature series of 1D spectra of Arc in D_2O (see Figure 3). The two phenylalanines were identified sequentially on the basis of further analysis of NOESY spectra in H_2O. The strong temperature dependence for the Phe^{10} signals is noteworthy. The highest temperature indicated in Figure 3 is close to the melting temperature of Arc. The ^{1}H NMR spectra indicated a refolding of the protein at lower temperature after melting at about 55 °C (Vershon et al., 1985). In general, the temperature dependence of the chemical shifts, together with a slight dependence on pH in the region 4.5–5.0, was used several times to remove ambiguities in the assignment. The NOESY spectra in D_2O also provided probable positions for the C_α- and C_βH's of the aromatic residues Phe^8, Phe^{45}, Tyr^{38}, and Trp^{14} from cross-peaks with aromatic ring protons. The further assignment of the spectrum was based on sequential connectivities in NOESY spectra in H_2O, together with spin system connectivities from amide to C_αH's and side-chain protons observed in the HOHAHA spectra in H_2O. The ^{1}H NMR assignments that have been obtained so far are compiled in Table I.

Sequence-Specific Assignment of Amide-Based Backbone Connectivities. A complete 600-MHz NOESY spectrum of Arc repressor is presented in Figure 4. This spectrum is a sum of three separate NOESY data sets with mixing times of 100, 150, and 200 ms. This spectrum and a similar spectrum obtained at 500 MHz were used throughout the sequential analysis, although in some cases the individual spectra were used. The sequential assignment was performed according to procedures described by Wüthrich (1986) and by Englander and Wand (1987). The NOESY spectrum in H_2O revealed a large number of strong d_{NN} connectivities that were used as starting points. These connectivities, together with sequential $d_{\alpha N}$, $d_{\beta N}$, $d_{\gamma N}$, and $d_{\delta N}$ NOEs, led to recognition of short stretches of adjacent residues. The actual position of these stretches within the total sequence of Arc was inferred from the nature of the side chains for a limited number of

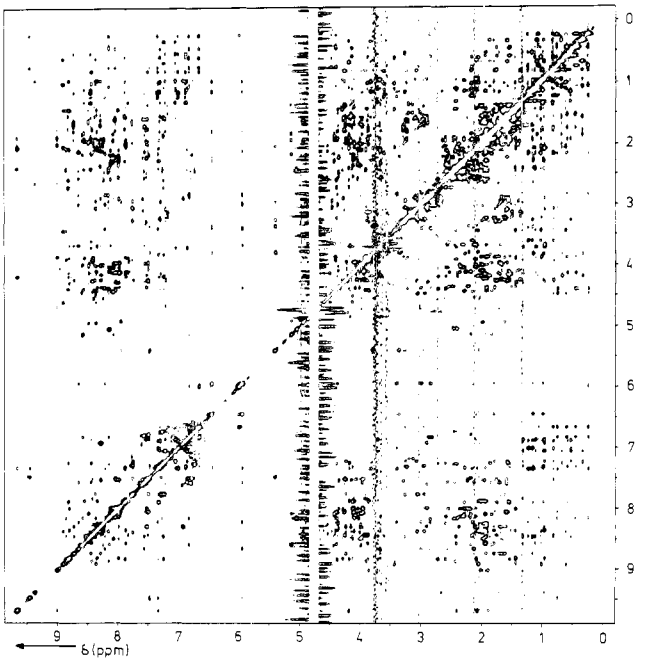

FIGURE 4: NOESY spectrum of the Arc repressor in H_2O at 600 MHz. The spectrum is a sum of three data sets for NOESY experiments with mixing times 100, 150, and 200 ms.

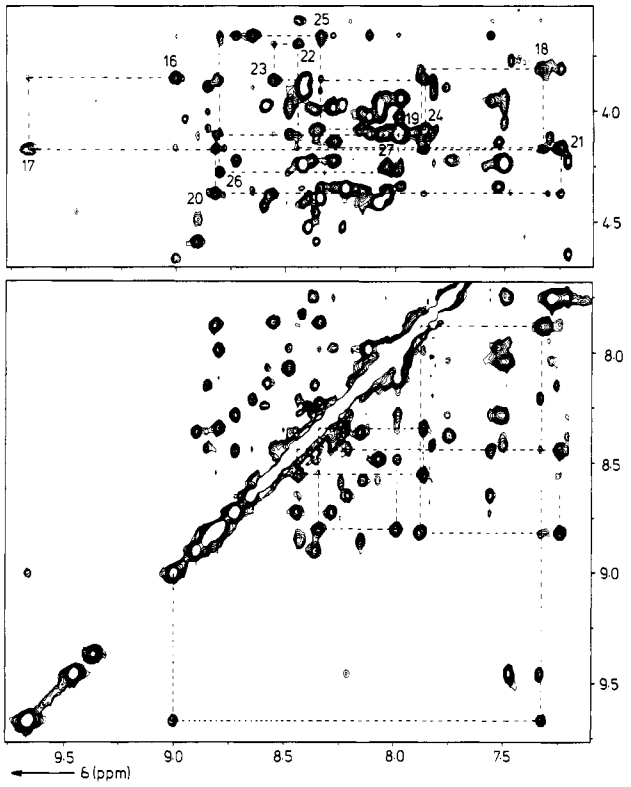

FIGURE 5: Sequential assignment in a 600-MHz NOESY spectrum recorded in H_2O of backbone signals for the residues Arg^{16} to Glu^{27} in the first α-helix of Arc. The sequential assignment is indicated in the $C_\alpha H$–NH part of the spectrum (top panel), with the residue number positioned in the spectrum near the intraresidue contact. The lower panel shows the NH–NH part of the spectrum with cross-peaks connected by dashed lines corresponding to the upper panel.

residues. By use of this strategy, the signals for the backbone protons of the amino acids Arg^{16} to Glu^{27} were characterized by the presence of three Val residues at positions 18, 22, and 25 and by an Ala residue at 26 (see Figure 5). Another set of backbone connectivities was identified by the unique tyrosine at position 38 and a valine residue at position 41. This yielded

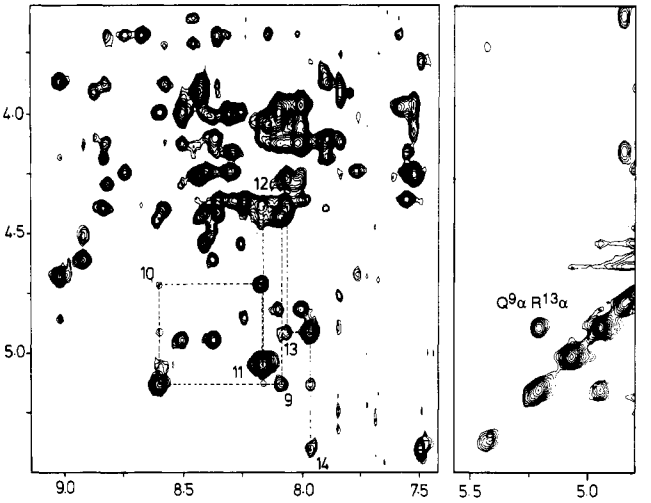

FIGURE 6: Part of the 500-MHz SCUBA NOESY spectrum in H_2O showing sequential $d_{\alpha N}$ connectivities of residues 8–14 (left panel). The right panel shows a small portion of a NOESY spectrum in D_2O, in which the Glu^9 $C_\alpha H$–Arg^{13} $C_\alpha H$ cross-peak is observed.

assignments for residues 35–46. Phe^{45} was included in this segment, although its $C_\alpha H$ signal, like that of Phe^{10}, was very close to the H_2O line as suggested by the NOESY spectra in D_2O. These signals could only be observed unperturbed in the spectra in H_2O by using the SCUBA method (see Figure 6). The latter spectra showed the sequential $d_{\alpha N}$ NOEs for both Phe^{45} and Phe^{10}. The identification of both stretches 16–27 and 35–46 was further supported by sequential side-chain NOEs $d_{\beta N}$, $d_{\gamma N}$, and $d_{\delta N}$, for most of the residues, and by the observation of weak $d_{NN}(i,i+2)$ contacts in the NOESY spectra with the longest mixing times. Residues 30–34 were identified by sequential NOEs and by the presence of the fifth valine (Val^{33}) in this segment. Due to near overlap of the backbone amide protons of Glu^{27} and Asn^{29} the sequential assignment of the backbone protons for residues 27–29 relied mostly on side-chain sequential contacts and on the $d_{\alpha N}$ cross-peak from Asn^{29}. The latter NOE was obscured by the proximity of the H_2O signal and therefore only observed in the SCUBA NOESY spectrum. The second alanine, at position 53, was the starting point for the identification of sequential contacts for a stretch of amino acid residues starting from the C-terminus. Strong d_{NN} NOEs and sequential contacts for the C_α and $C_\beta H$'s were found for residues 45–53. The latter segment included a cross-peak close to the diagonal between the amides of Lys^{46} and Lys^{47} and, in the SCUBA NOESY spectrum, a weak $d_{\alpha N}$ cross-peak connecting Phe^{45} and Lys^{46}. It should be stressed here that the slight variation of the chemical shifts depending on the sample conditions was essential for the firm identification of NOE contacts in the amide region of the spectrum.

Starting from residues Phe^{10} and the unique tryptophan at position 14, the segment from 9 to 14 is found by using sequential d_{NN} NOEs of relatively weak intensity and strong $d_{\alpha N}$ NOEs (see Figure 6). This includes the position of Pro^{15} C_δ protons, found from the strong NOEs with Trp^{14} $C_\alpha H$, in accordance with a trans configuration for the Trp^{14}-Pro^{15} peptide linkage (Wüthrich et al., 1984). Similarly, the position of Pro^8 $C_\alpha H$ was identified from a strong NOE of this proton to the amide proton of Glu^9. Under the standard measuring conditions, the small chemical shift difference of the amide signals of Asn^{11} and Leu^{12} was sufficient to observe the two individual spin systems but precluded a distinction between the intraresidue $d_{N\alpha}(i,i)$ NOE of Asn^{11} and the sequential $d_{\alpha N}$ NOE with Leu^{12}. The overlap of the latter cross-peaks,

Table I: ^{1}H Chemical Shifts[a] and Assignments of P22 Arc Repressor at pH 4.7 and 30 °C

residue	HN	Hα	Hβ/β′	Hγ/γ′	others
Met[1]					
Lys[2]	8.96	4.44	1.83	1.47	
Gly[3]	8.59	3.98			
Met[4]	8.24	4.52	2.02/2.10	2.54/2.60	
Ser[5]	8.39	4.40	3.86/3.91		
Lys[6]	8.34	4.35	1.77	1.42/1.49	3.01 (ε/ε′)
Met[7]	8.22	4.84	1.90	2.45/2.58	
Pro[8]		4.42	1.92/2.27		3.67/4.17 (δ/δ′)
Gln[9]	8.07	5.13	1.81/1.92	2.17	6.79/7.48 (ε/ε′)
Phe[10]	8.58	4.75	2.83		6.83 (δ_2), 6.64 (ϵ_2), 5.97 (ζ)
Asn[11]	8.14	5.04	2.39/2.44		6.72/7.25 (δ2/δ2′)
Leu[12]	8.13	4.38	0.42	1.04	0.22/0.73 (δ_3/δ'_3)
Arg[13]	8.04	4.91	1.59/1.65	1.35/1.47	3.12/3.18 (δ/δ′), 7.17 (ε)
Trp[14]	7.94	5.39	2.94/3.25		7.47 (δ1), 9.45, (ε1), 7.29 (ε3), 7.33 (ζ2), 6.87 (ζ3), 6.99 (η)
Pro[15]		4.67			3.35/3.78 (δ/δ′)
Arg[16]	9.00	3.85	1.87/1.99	1.70/1.75	3.27 (δ/δ′), 7.53 (ε)
Glu[17]	9.66	4.18	2.05	2.39	
Val[18]	7.32	3.81	2.26	0.96/1.13[b]	
Leu[19]	7.88	4.13	1.72	1.52	0.43/0.56 (δ_3/δ'_3)
Asp[20]	8.82	4.38	2.58/2.71		
Leu[21]	7.24	4.17	1.54/1.84	1.06	0.31/0.66 (δ_3/δ'_3)
Val[22]	8.44	3.70	2.40	1.09/1.30[b]	
Arg[23]	8.55	3.86	1.91	1.60/1.78	3.21 (δ/δ′), 7.30 (ε)
Lys[24]	7.86	4.09	1.98	0.94	2.87 (ε/ε′)
Val[25]	8.34	3.66	2.07	0.95	
Ala[26]	8.80	4.28	1.42		
Glu[27]	7.98	4.10	2.17/2.25	2.35/2.53	
Glu[28]	8.11	4.03	2.08/2.30	2.34/2.56	
Asn[29]	7.98	4.80	2.88		6.91/8.27 (δ2/δ2′)
Gly[30]	8.06	3.98			
Arg[31]	8.48	4.93	1.59/1.64	1.96	3.03/3.26 (δ/δ′), 7.76, (ε)
Ser[32]	8.35	4.60	4.12/4.50		
Val[33]	8.90	3.42	2.03	0.59/1.00[b]	
Asn[34]	8.36	4.46	2.93/3.05		7.27/7.75 (δ2/δ2′)
Ser[35]	8.15	4.36	4.01/4.08		
Glu[36]	8.85	3.89	2.07	2.41	
Ile[37]	8.43	3.59	2.14		1.21 ($\gamma 2_3$)
Tyr[38]	8.20	3.38	2.84/2.98		5.93 (δ_2), 6.43 (ϵ_2)
Gln[39]	8.64	3.66	2.11	2.42/2.55	6.79/7.56 (ε2/ε2′)
Arg[40]	7.56	3.96	1.69/1.84		2.64/2.93 (δ/δ′), 9.36 (ε)
Val[41]	8.44	3.25	1.91	0.51/1.14[b]	
Met[42]	8.72	4.23	1.67	1.77	
Glu[43]	8.28	3.99	1.99/2.09	2.35	
Ser[44]	7.50	4.23	3.94/4.06		
Phe[45]	7.74	4.66	3.06/3.23		7.20 (δ_2), 7.03 (ϵ_2), 6.84 (ζ)
Lys[46]	8.37	4.23	1.94	1.37/1.52	2.93 (ε/ε′)
Lys[47]	8.27	4.15	1.94		1.49/1.58 (δ/δ′), 2.98 (ε/ε′)
Glu[48]	7.52	4.35	2.21	2.46/2.54	
Gly[49]	7.97	3.94/4.09			
Arg[50]	7.98	4.26	1.56		2.89/3.08 (δ/δ′), 7.58 (ε/ε′)
Ile[51]	7.49	4.24	1.73	0.95/1.33	0.81 ($\gamma 2_3$), 0.82 (δ_3)
Gly[52]	8.41	3.90			
Ala[53]	7.82	4.10	1.30		

[a] Chemical shifts in ppm from 4,4-dimethyl-4-silapentane-1-sulfonate with accuracy ±0.01 ppm. [b] Stereospecifically assigned as γ1/γ2 according to the IUPAC–IUB conventions (IUPAC–IUB Commission on Biochemical Nomenclature, 1970).

however, was removed at pH 4.2. The sequential and intraresidue NOEs from Phe[10] $C_\alpha H$, and also the $d_{\alpha N}$ NOE from Pro[15] to Arg[16], were only observable in the SCUBA NOESY spectrum. Met[7] $C_\alpha H$ was identified from its strong sequential NOEs to the Pro[8] C_δ protons in a NOESY spectrum in D_2O, also in accordance with a trans isomer for this peptide linkage. Since the former Met[7] signal was close to the H_2O signal, the position of its NH and the sequential NOE to Lys[6] were obtained from the SCUBA experiments. Subsequently, the backbone contacts for Gly[3] to Lys[6] were obtained by using sequential backbone NOEs. For Lys[2] only a $d_{\alpha N}$ cross-peak to Gly[3] was present. For the N-terminal methionine residue no assignments were thus far obtained.

Assignment of Side-Chain Proton Resonances. The assignment of side-chain protons was based on the HOHAHA spectra with various mixing times in combination with the NOESY spectra in H_2O. Nearly all C_β protons were found as well as most of the C_γ protons. The broadness of the signals, in accordance with the dimeric nature of Arc, resulted in a limited quality of the HOHAHA spectra. This, together with overlap for the HOHAHA spectra and the NOESY spectra, precluded the complete assignment for the longer side chains such as lysine, arginine, proline, and isoleucine. The poor quality of the HOHAHA spectra is probably due to the broadness of the signals, in accordance with the dimeric nature of Arc. For all arginines the $N_\epsilon H$ signal was identified by connecting these signals to the respective amide resonances, via their side-chain signals, by using the combined information from HOHAHA and NOESY spectra. For the Ile[37] side chain the assignment is limited to $C_\beta H$ and the C_γ methyl signal. Leu[12] $C_\beta H$ and H′ were identified at very high-field positions (0.42 ppm). This was deduced not only from a HOHAHA spectrum with short mixing time but also from a cross-peak in a DQF-COSY spectrum. The discussion of the unusual

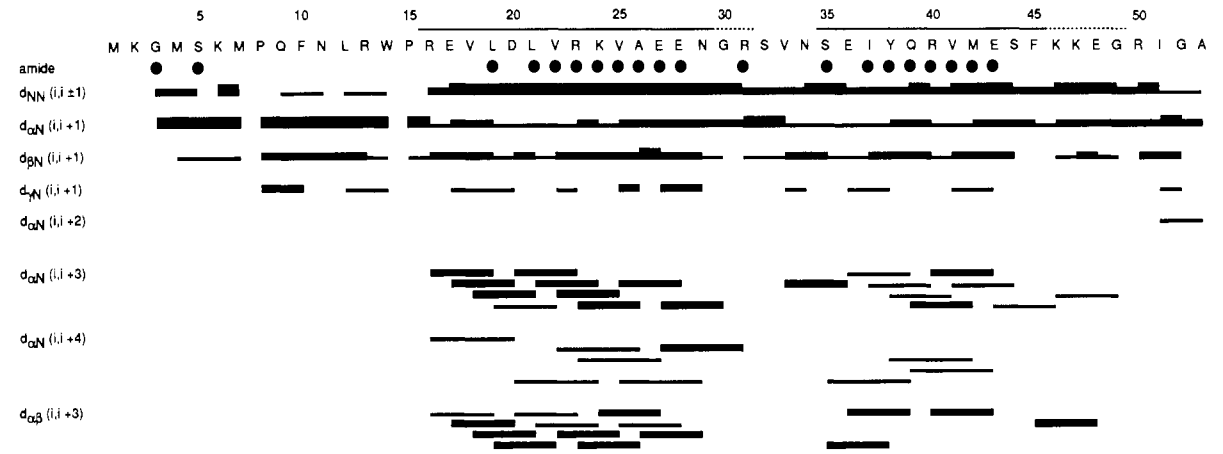

FIGURE 7: Sequence of Arc repressor protein together with a schematic summary of the short- and medium-range NOEs determining the secondary structure of Arc. The thickness of the connecting lines is a measure for the intensity of the NOEs. The black circles on the row labeled "amide" denote residues that contain slowly exchanging amide protons. The line above the amino acid sequence denotes the regions of α-helical secondary structure. The uncertainty in the end points of both helices is indicated by dashed lines.

position of these signals will be included in the tertiary structure analysis, on which work is in progress.

For the prolines, some of the C_β and $C_\gamma H$'s are not identified yet due to overlap. The signals of the side-chain amide protons of the Asn and Gln were all identified from NOEs to their C_β and C_γ protons, respectively. Since NOEs to the methyl groups of the methionines have not been found yet, nor those to arginine NH_2 and lysine amino groups, no assignments for these groups have been included in Table I.

The methyl resonances of all valines are assigned stereospecifically according to the method of Zuiderweg et al. (1985). To this end a DQF-COSY spectrum in D_2O at 37 °C was recorded at 600 MHz. This spectrum contained only part of the expected cross-peaks, which is due to the large line width for most signals. The cross-peaks for the valines were, however, clearly observable with a large $^3J_{\alpha\beta}$ in accordance with a trans orientation for these protons, allowing stereospecific assignment of the methyl groups via differential $d_{\gamma N}$ NOE intensities.

Secondary Structure. α-Helices are characterized by strong sequential d_{NN} NOEs and the presence of medium-range $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, and $d_{\alpha\beta}(i,i+3)$ NOEs (Wüthrich, 1986; Zuiderweg et al., 1983). Furthermore, since secondary structures are stabilized by hydrogen bonds between backbone amide protons and carbonyl oxygens, α-helices are often characterized by a sequence of slowly exchanging amide protons starting with the fourth residue in the helix. In a NOESY spectrum recorded within 5 h after dissolving the protein in D_2O, about 40% of the amide protons were still observable. Figure 7 shows the short- and medium-range NOEs and the slowly exchanging amides of Arc repressor. The combination of strong d_{NN} and medium-range NOEs clearly shows that an α-helix is present starting at Arg[16] and extending at least to Asn[29] and possibly to Arg[31]. This is supported by a series of slowly exchanging amides in this peptide segment. The end point of this helix is difficult to identify due to overlap of a number of $d_{\alpha N}(i,i+3)$ NOEs with sequential $d_{\alpha N}$ cross-peaks. The helix may continue as far as Arg[31], as judged from the NOE from Glu[27] $C_\alpha H$ to Gly[30] and Arg[31] NH, respectively, while a possible NOE between Glu[28] $C_\alpha H$ and Arg[31] NH coincides with the Gly[30] $d_{\alpha N}$ cross-peak. Slow H–D exchange observed for a series of amide protons starting at Leu[19] is in accordance with an α-helix that starts at residue 16.

The pattern of medium-range NOEs and slowly exchanging amides indicates the presence of a second helix (see Figure 7). The start of this helix is probably at Ser[35], in accordance with a $d_{\alpha N}(i,i+4)$ and a $d_{\alpha\beta}$ NOE starting from this residue, and the absence of a cross-peak from Asn[34] $C_\alpha H$ to Ile[37] NH. A possible cross-peak between Ser[35] $C_\alpha H$ and Tyr[38] NH is obscured due to overlap with the intense cross-peak Met[7] NH–Lys[6] $C_\alpha H$. The presence of an intense NOE from Val[33] $C_\alpha H$ to Glu[36] NH suggests that the second helix may even start at Val[33]. Again, the end of the second helix is dificult to define, due to overlap of $d_{\alpha N}(i,i+3)$ contacts from Met[42] and from Ser[44], while the $d_{\alpha N}(i,i+3)$ contact from Phe[45] is not detected. When, however, the $d_{\alpha\beta}(i,i+1)$ connectivity from Phe[45] is taken into account, the second helix may continue as far as Gly[49]. The data on slowly exchanging amides do not provide conclusive evidence as to the end of the second helix. The slow exchange observed for the amides of Ser[35] and Arg[31] may be due to tertiary structure interactions. The position of the two α-helices have been indicated in Figure 7. The observation of relatively intense NOEs for the C-terminal amino acids indicates a well-defined structure for this part of the molecule. The $d_{\alpha N}(i,i+2)$ NOE from Ile[51] suggests this residue to be involved in a tight turn. The precise structure of the C-terminus will have to await the analysis of the tertiary structure of the Arc repressor protein.

The α-helices are in principle also characterized by $^3J_{N\alpha}$ constants of 3–4 Hz, in contrast to a value of about 9 Hz for a β-sheet. DQF-COSY spectra in H_2O showed cross-peaks in accordance with assignments obtained from HOHAHA spectra. The apparent splittings of these cross-peaks however was completely determined by the line width (Neuhaus et al., 1985), and no reliable information about the $^3J_{N\alpha}$ could be derived to support the α-helices or the β-sheet secondary structures (see below).

An extended structure is characterized by strong sequential $d_{\alpha N}$ contacts and weak $d_{N\alpha}(i,i)$ and d_{NN} NOEs. For all amino acids of the segment Pro[8] to Trp[14], strong sequential $d_{\alpha N}$ contacts are indeed observed. For Phe[10] this contact is obscured by the proximity of the H_2O line, but in a SCUBA NOESY experiment the sequential NOE is as strong as the others in this segment. In the same experiment a weak intraresidue $d_{N\alpha}$ NOE is observed for Phe[10], as is also present for the other residues in this segment. The intraresidue and sequential $d_{\alpha N}$ contacts for Asn[11] were only observed at a reduced pH but were weak and strong, respectively (see above). In the latter spectrum the sequential d_{NN} contact between Asn[11] and Leu[12] was still too close to the diagonal to be observed.

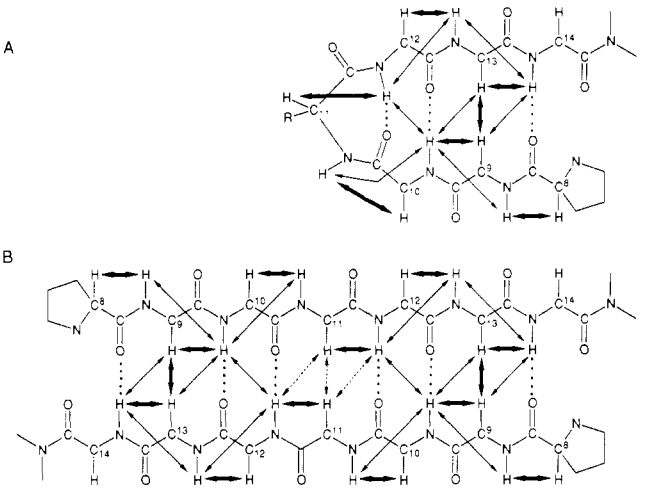

FIGURE 8: Schematic representation of the β-sheet part for residues 8–14 in Arc. The long-range NOEs allow two possibilities: an intramonomer β-sheet with a γ-turn or inverse γ-turn (A) or an intermonomer β-sheet with strands from different Arc monomers (B). Included are observed NOEs with the intensities represented by the thickness of the connecting arrow. The broken arrows in the intermonomer β-sheet (B) represent NOEs between the two strands that cannot be measured for reasons of symmetry, either because they would be between protons of the same residue in different monomers (Asn^{11} NH–Asn^{11} NH) or because the NOE cannot be distinguished from an intramonomer one (Asn^{11} C_αH–Leu^{12} NH).

A number of long-range NOEs involving the extended part between residues 8 and 14 were observed that are characteristic for an antiparallel β-sheet. Thus, a strong NOE between the C_α protons of Gln^9 and Arg^{13}, medium-intensity cross-strand $d_{\alpha N}$ NOEs between residues 13 and 10 and between 9 and 14, and a medium-intensity NOE between the amide protons of residues 10 and 12 are present. Since at present it is not known whether these NOEs are intra- or intermonomer, there are still two interpretations possible for the secondary structure. If the NOEs are between protons within an Arc monomer, then this implies a loop structure involving residues 10–12. In that case the most likely conformation would be a γ-turn (or inverse γ-turn) as shown in Figure 8A (Némethy & Printz, 1972; Rose et al., 1985). However, if the NOEs are between different monomers, then an intermonomer antiparallel β-strand is implied (see Figure 8B). For this situation the dyad axis of the Arc dimer would lie between the C_α protons of Asn^{11} so that the $d_{\alpha\alpha}$ NOE between these protons could not be observed. Furthermore, the cross-strand $d_{\alpha N}$ NOE between residues 11 and 12 could not be distinguished from the sequential one (cf. Figure 8B). While the present data do not allow a clear distinction between these cases, there are some arguments in favor of the intermonomer β-sheet. The first follows from a comparison between the observed NOE intensity pattern and the corresponding distances expected for a straight β-sheet and for the γ-turn, normal and inverse (see Table II). Inspection of Table II shows that the observed NOE intensities correspond completely with an intermonomer β-sheet and neither with a γ-turn nor with an inverse γ-turn. Thus, for a γ-turn the intraresidue $d_{N\alpha}(i,i)$ for Asn^{11} is expected to be strong, while the observed NOE is weak. Further, a strong sequential $d_{\alpha N}$ NOE for residues 11 and 12 is observed, whereas in the γ-turn this distance would be relatively long. Similarly, the inverse γ-turn is not likely due to discrepancies between observed and expected NOE intensities for the sequential $d_{\alpha N}(10,11)$ and the intraresidue $d_{N\alpha}(12,12)$.

The second argument is based on the absence of slowly exchanging amide protons in the peptide region 9–14, while slow exchange is observed in the α-helical region (cf. Figure 7). This is unusual, since normally the amide protons involved in hydrogen bonding in β-sheets are very persistent toward H–D exchange [see for instance Wagner (1983) for the situation in BPTI]. Since it is known that Arc monomers exchange between dimers (R. T. Sauer, personal communication), the relatively fast H–D exchange of amide protons could be related to the disruption of the β-sheet during dissociation of the Arc dimer. Both of these arguments are not definitive. For instance, a loop structure different from a γ-turn or inverse γ-turn could be present, and the β-sheet region could be destabilized for other reasons. Thus, while we presently have a slight preference for the intermonomer β-sheet, the question which of the two possibilities exists and also the complete structure determination of Arc repressor must await the resolution of the NOEs in terms of their intra- or intermonomer character. We should note that there are several examples of stabilization of protein dimers by intermonomer β-sheet regions such as, for instance, in insulin (Blundell et al., 1972) and in cro repressor of phage λ (Anderson et al., 1984).

It is remarkable that the two α-helices and also the β-sheet region each contain a distinct polar and apolar side (see Figure 9). In the 2D NOE spectra many NOEs were present between the apolar residues of both α-helices and the β-sheet region (data not shown). Therefore, the amphiphilic character of these elements appears to be essential for the stability of the protein. This was also noted by Bowie and Sauer (1989)

Table II: Selected Proton–Proton Distances and Observed NOEs for Residues 9–14 of Arc Repressor in Three Different Conformations (Full Antiparallel Dimer β-Sheet, γ-Turn, and Inverse γ-Turn)

proton pair	distance			obsd NOE[b]
	β-sheet	γ-turn[a]	inverse γ-turn[a]	
Glu^9N–Phe^{10}N	4.3			weak
$Gln^9\alpha$–Phe^{10}N	2.2			strong
Phe^{10}N–$Phe^{10}\alpha$	2.8	2.5–2.8	2.5–2.8	weak
Phe^{10}N–Asn^{11}N	4.3	3.6–3.8	3.5–3.7	weak
Phe^{10}N–Leu^{12}N	3.3[c]	3.2–3.6	3.5–4.0	medium
$Phe^{10}\alpha$–Asn^{11}N	2.2	2.2–2.4	3.2–3.4	strong
Asn^{11}N–$Asn^{11}\alpha$	2.8	2.2–2.4	2.6–2.8	weak
Asn^{11}N–Leu^{12}N	4.3	3.8–4.2	3.7–4.0	nd[d]
$Asn^{11}\alpha$–Leu^{12}N	2.2	3.5–3.7	2.5–2.7	strong
Leu^{12}N–$Leu^{12}\alpha$	2.8	3.0–3.2	2.3–2.5	medium
Leu^{12}N–Arg^{13}N	4.3	4.4–4.6	4.6–4.8	weak
$Leu^{12}\alpha$–Arg^{13}N	2.2	2.4–2.6	2.4–2.6	strong
$Arg^{13}\alpha$–Trp^{14}N	2.2			strong
Arg^{13}N–$Arg^{13}\alpha$	2.8			weak

[a] Distances estimated from model building on the basis of the ideal γ-turn and inverse γ-turn (Némethy & Printz, 1972); cf. Figure 8. [b] Observed NOEs classified according to three different intensities. [c] Distance between protons of two Arc monomers. [d] Not determined.

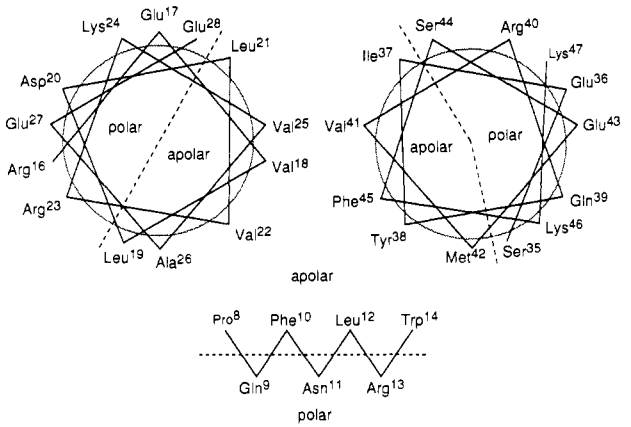

FIGURE 9: Schematic representation of the three secondary structure elements in Arc repressor. Both helices and the β-sheet have very clear polar and apolar sides, as indicated.

in an analysis of structurally neutral substitutions for Arc repressor. The latter analysis predicted a β-sheet for residues 9–14 and α-helices for residues 16–28 and 35–47, which corresponds strikingly with the NMR data. The NMR study leaves some ambiguity as to the ends of the two α-helices and to the intra- or intermonomer nature of the β-sheet. Sauer and his co-workers [see Knight et al. (1989)] have identified the N-terminal residues of Arc as the DNA binding domain. The present NMR study, irrespective of the interpretation of the long-range NOEs, shows that this region of the protein is in an extended backbone conformation and forms at least partially a β-pleated sheet.

Further work will be aimed at resolving the ambiguity in the long-range NOEs, which is a prerequisite for the determination of the solution structure of Arc repressor and of its complexes with operator DNA.

ACKNOWLEDGMENTS

We thank Dr. R. T. Sauer for a gift of the initial sample of Arc repressor and for many helpful discussions.

REFERENCES

Anderson, J., Ptashne, M., & Harrison, S. C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1307–1311.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 335–366.

Blundell, T., Dodson, G., Hodgkin, D., & Mercola, D. (1972) *Adv. Protein Chem. 26*, 279–402.

Boelens, R., Scheek, R. M., van Boom, J. H., & Kaptein, R. (1987) *J. Mol. Biol. 193*, 213–216.

Bowie, J. U., & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2152–2156.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–558.

Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166–169.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2821–2822.

Dodd, I. B., & Egan, J. B. (1987) *J. Mol. Biol. 194*, 557–564.

Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958.

Jordan, S. R., Pabo, C. O., Vershon, A. K., & Sauer, R. T. (1985) *J. Mol. Biol. 185*, 445–446.

Knight, K. L., & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 797–801.

Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D., & Sauer, R. T. (1989) *J. Biol. Chem. 264*, 3639–3642.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Némethy, G., & Printz, M. P. (1972) *Macromolecules 5*, 755–758.

Neuhaus, D., Wagner, G., Vásak, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257–273.

Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.

Piantini, U., Sorensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.

Rose, G. D., Gierasch, L. M., & Smith, J. A. (1985) *Adv. Protein Chem. 37*, 1–109.

Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P., & Susskind, M. M. (1983) *J. Mol. Biol. 168*, 699–713.

Susskind, M. M. (1980) *J. Mol. Biol. 138*, 685–713.

Susskind, M. M., & Youderian, P. (1982) *J. Mol. Biol. 154*, 427–447.

Vershon, A. K., Youderian, P., Susskind, M. M., & Sauer, R. T. (1985) *J. Mol. Biol. 260*, 12124–12129.

Vershon, A. K., Bowie, J., Karplus, T., & Sauer, R. T. (1986) *Proteins: Struct., Funct., Genet. 1*, 302–311.

Wagner, G. (1984) *Q. Rev. Biophys. 16*, 1–57.

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

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

Youderian, P., Vershon, A., Bouvier, S., Sauer, R. T., & Susskind, M. M. (1983) *Cell 35*, 777–783.

Zagorski, M. G., Bowie, J. U., Vershon, A. K., Sauer, R. T., & Patel, D. J. (1989) *Biochemistry* (preceding paper in this issue).

Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837–5841.

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601–611.