Katoh, H., Yoshinaga, M., Yanagita, T., Oghi, K., Irie, M., Beintema, J. J., & Meinsma, D. (1986) *Biochim. Biophys. Acta 873*, 367–371.
Knecht, R., & Chang, J.-Y. (1987) *Anal. Chem. 58*, 2375–2379.
Kornfeld, R., & Kornfeld, A. (1985) *Annu. Rev. Biochem. 54*, 631–664.
Kraft, N., & Shortman, K. (1970) *Aust. J. Biol. Sci. 23*, 175–184.
Kumagai, H., Kato, H., Igarashi, K., & Hirose, S. (1983) *J. Biochem. 94*, 71–77.
Kurachi, K., Davie, E. W., Strydom, D. J., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry 24*, 5494–5502.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Lee, F. S., Fox, E. A., Zhou, H.-M., Strydom, D. J., & Vallee, B. L. (1988) *Biochemistry 27*, 8545–8553.
Lee, F. S., Auld, D. S., & Vallee, B. L. (1989a) *Biochemistry 28*, 219–224.
Lee, F. S., Shapiro, R., & Vallee, B. L. (1989b) *Biochemistry 28*, 225–230.
Maes, P., Damart, D., Rommens, C., Montreuil, J., Spik, G., & Tartar, A. (1988) *FEBS Lett. 241*, 41–45.
Morita, T., Sanda, A., Takizawa, Y., Ohgi, K., & Irie, M. (1987) *Agric. Biol. Chem. 51*, 2751–2761.
Palmieri, M., Carsana, A., Furia, A., & Libonati, M. (1985) *Eur. J. Biochem. 152*, 275–277.
Phela, J. J., & Hirs, C. H. W. (1970) *J. Biol. Chem. 245*, 654–661.
Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176–185.
Richards, F. M., & Wyckoff, H. W. (1971) *Enzymes (3rd Ed.) 4*, 647–806.
Roth, J. S. (1956) *Biochim. Biophys. Acta 21*, 34–43.
Roth, J. S. (1958) *J. Biol. Chem. 231*, 1085–1095.
Roth, J. S. (1962) *Biochim. Biophys. Acta 61*, 903–915.
Roth, J. S. (1967) *Methods Cancer Res. 3*, 151–243.
Schneider, R., Schneider-Scherzer, E., Thurnher, M., Auer, B., & Schweiger, M. (1988) *EMBO J. 7*, 4151–4156.
Shapira, R. (1962) *Anal. Biochem. 3*, 308–320.
Shapiro, R., Fett, J. W., Strydom, D. J., & Vallee, B. L. (1986) *Biochemistry 25*, 7255–7264.
Sierakowska, H., & Shugar, D. (1977) *Prog. Nucleic Acid Res. Mol. Biol. 20*, 59–130.
Smyth, D. G., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 227–234.
Sorrentino, S., Tucker, G. K., & Glitz, D. G. (1988) *J. Biol. Chem. 31*, 16125–16131.
Spackman, D. H., Stein, W. H., & Moore, S. (1960) *J. Biol. Chem. 235*, 648–659.
Strydom, D., Fett, J. W., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry 24*, 5486–5494.
Suzuki, H., Greco, L., Parente, A., Farina, B., La Montagna, R., & Leone, E. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 2, pp 93–94, National Biomedical Research Foundation, Washington, D.C.
Watanabe, H., Katoh, H., Ishii, M., Komoda, Y., Sanda, Y., Takizawa, Y., Ohgi, K., & Irie, M. (1988) *J. Biochem. 104*, 939–945.

NMR Studies of Arc Repressor Mutants: Proton Assignments, Secondary Structure, and Long-Range Contacts for the Thermostable Proline-8 → Leucine Variant of Arc†

Michael G. Zagorski,*,‡ James U. Bowie,§ Andrew K. Vershon,§ Robert T. Sauer,§ and Dinshaw J. Patel*,‡
Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received April 18, 1989; Revised Manuscript Received June 20, 1989

ABSTRACT: Arc repressor is a 53-residue sequence-specific DNA binding protein. We report the assignment of the proton NMR spectrum and the secondary structure for the thermostable PL8 variant of Arc. This mutant, which differs from wild type by a Pro-8 → Leu substitution, was chosen for study because its enhanced stability allows spectra to be acquired at elevated temperatures where spectral resolution is higher. The first five residues of the protein play important roles in DNA binding but appear to be disordered in solution. Residues 6–14 form the remaining part of the N-terminal DNA binding region of the protein and assume an antiparallel β-conformation. This indicates that Arc is a member of a new class of DNA binding proteins. The observed interresidue nuclear Overhauser effects are consistent with a β-strand, γ-turn, β-strand structure for the residue 6–14 region, although other structures are also consistent with the data. The remaining portion of the protein is predominantly α-helical. Residues 16–26 and 35–50 form amphipathic α-helices which may pack together in a four-helix bundle in the protein dimer.

Arc is a small DNA binding protein responsible in part for the regulation of lysogeny in the *Salmonella* phage P22 (Susskind, 1980). The 53 amino acid protein (Chart I) binds to its operator as a tetramer (Brown, unpublished results) but is dimeric in solution (Vershon et al., 1985). Studies of mutant proteins show that Arc uses residues at its N-terminal end to mediate most of its interactions with operator DNA (Vershon et al., 1986; Bowie & Sauer, 1989). Moreover, a chimeric protein containing the N-terminal nine amino acids of Arc

† This research was funded by start-up funds from Columbia University and NIH Grant GM-34504 to D.J.P. and NIH Grants AI-15706 and AI-16892 to R.T.S.
‡ Columbia University.
§ Massachusetts Institute of Technology.

0006-2960/89/0428-9813$01.50/0

Chart I

```
 1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
MET LYS GLY MET SER LYS MET PRO GLN PHE ASN LEU ARG TRP PRO ARG GLU VAL
                           |
                          LEU

 19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
LEU ASP LEU VAL ARG LYS VAL ALA GLU GLU ASN GLY ARG SER VAL ASN SER GLU

 37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
ILE TYR GLN ARG VAL MET GLU SER PHE LYS LYS GLU GLY ARG ILE GLY ALA
```

repressor in place of the N-terminal six residues from a homologous repressor, Mnt, has the operator binding specificity of Arc repressor (Knight & Sauer, 1989).

Knowledge of the molecular structure of Arc is fundamental before any details of its intricate binding to DNA can be understood. The three-dimensional structure of Arc is not known, although crystals have been grown and work on the structure is currently in progress (Jordan et al., 1985). We demonstrate here the first step toward achieving this goal, namely, details of the secondary structure of PL8, a mutant of Arc having leucine substituted for proline at position 8. We also compared NMR spectra of additional Arc mutants and discuss differences with regard to their folding, stabilities, and DNA binding affinities. The results for PL8 presented in this paper were obtained from an analysis of two-dimensional (2D) NMR data. PL8 presumably folds analogously to Arc yet has a substantially higher melting temperature (T_m). This permutation allowed NMR data to be recorded at higher temperatures. Complete sequential amino acid assignments for PL8 were obtained at elevated temperatures, and analysis of NOESY data indicates that PL8 adopts a stable folded three-dimensional structure in solution. These data also indicate that PL8 is predominantley helical with two separate α-helical segments separated by a turn or extended loop structure. Residues located at the N-terminus, however, exist in a β-conformation, thus further suggesting that indeed DNA binding of PL8 and Arc do not involve a conventional helix–turn–helix or zinc finger structure.

Materials and Methods

Sample Preparation. Mutant proteins were prepared as described previously by Vershon et al. (1986). 2D NMR measurements of PL8 used a sample that was 4 mM in protein, whereas for the 1D NMR measurements of Arc and other mutant protein samples were 1 mM. All samples were prepared in either 99.96% D_2O or 90% H_2O/10% D_2O solution containing 50 mM sodium phosphate, 200 mM KCl, and 0.1 mM Na_2EDTA. For nonexchangeable proton NMR experiments, the labile NH protons were exchanged by repeated lyophilization and redissolution in D_2O. The pH of each solution was adjusted to 4.3, 5.4, 6.2, or 7.3 by the addition of microliter amounts of dilute DCl or NaOD. pH meter readings were measured at room temperature and were not corrected for isotope effects.

General Procedures for NMR Data Acquisition and Analysis. All proton NMR spectra were obtained at 500 MHz on a Bruker AM-500 spectrometer equipped with digital phase shifters and an Aspect 3000 computer. The data were transferred by magnetic tape to a micro VAX II for processing with Dennis Hare FTNMR software. Chemical shifts are referenced to H_2O and HDO, which at 50 °C, for example, resonate at 4.52 and 4.50 ppm, respectively, relative to sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$ (TSP).

With the exception of an absolute-value COSY experiment recorded in H_2O solution, all remaining 2D NMR experiments were recorded in the phase-sensitive absorption mode with quadrature in both dimensions (States et al., 1982). To achieve quadrature detection in F_1, two FIDs with the phase of the initial 90° preparation pulse differing by 90° were acquired for each t_1 increment. These two FIDs form the complex or quadrature pair in F_1. The carrier was placed in the center of the spectrum on the water signal, and a recycle delay of 1.5–2.0 s was used for all experiments. During the recycle delay and the mixing time of the NOESY experiments, the water signal (H_2O or HDO) was suppressed by continuous low-power irradiation from the proton decoupler channel. Identical low-power levels (2 mW) were used for suppressing both H_2O and HDO signals by careful shimming of the water signal. In addition, the efficiency of this suppression was further improved after Faraday shields were installed on the coil leads of our 5-mm ^1H probe (Dykstra, 1987). The spectral width in F_2 was 6024.10 Hz, and typically 512–1024 increments were collected with 2048 data points. The time domain data sets were subsequently zero filled in F_1 to 1024 points to provide 1024 × 1024 final matrices. The only exception was the phase-sensitive COSY experiment done in H_2O solution where 4096 points were used to acquire the data in F_2. These raw data were then zero filled to 8192 points in F_2 and subsequently reduced back to 4096 points to conserve disk space, while in F_1 zero filling was done to only 2048 points. This provided a final digital resolution of 0.74 Hz/point in F_2. Fourier transformation in both dimensions was performed after multiplication by a skewed sine bell window function shifted by 0–40°, depending upon the amount of resolution and sensitivity needed from the particular experiment. For all experiments care was taken to optimize the base line of the spectra by manipulating the intensity of the first points in F_1 and F_2 prior to Fourier transformation (Otting et al., 1986). Further flattening of the base line was obtained by using a fifth-order polynomial base line correction of the rows in the final 2D matrix.

Experimental Techniques. Phase-sensitive NOESY spectra (Jeener et al., 1979; Wider et al., 1984) were acquired by using the standard pulse sequence with a 16-step phase cycle to remove axial peaks and other unwanted peaks resulting from multiple quantum transitions. In addition, a 10% random variation in the mixing time was used to eliminate zero quantum coherence (Macura et al., 1981). Mixing times of 100 and 200 ms were used. For NOESY experiments conducted in H_2O and D_2O solutions, respectively, 796 and 514 increments each consisting of 160 and 64 scans were utilized for final processing.

DQF-COSY and TQF-COSY spectra in D_2O solution were recorded by using pulse sequences consisting of three noncomposite 90° pulses and, respectively, 32- and 48-step phase cycling schemes as described by Morris (1986). The longer phase cycling schemes, together with narrow-band frequency-rejection filters (Dykstra & Wand, 1987) which were installed at different points in the radio frequency (rf) receiver circuitry, were helpful in reducing many artifacts originally present in these COSY experiments. For both experiments identical acquisition and processing parameters were used,

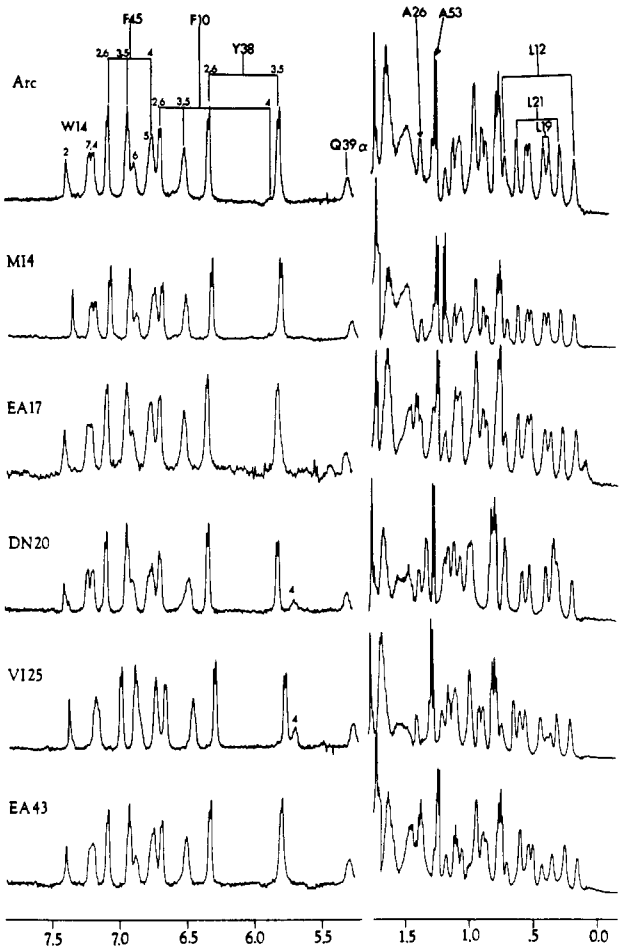


FIGURE 1: Expanded aromatic (left) and aliphatic (right) regions of ^{1}H NMR spectra of Arc and mutant proteins. Mutations are designated by one-letter codes for the wild-type and mutant amino acids and the residue position of the sequence change; i.e., MI4 changes the wild-type methionine at residue 4 of Arc to isoleucine. Conditions of sample preparations and measurements of the NMR spectra were identical for all the protein samples: 1 mM protein concentration, 25 °C, pH 7.3, in 25 mM sodium phosphate buffer containing 200 mM KCl and 0.1 mM Na_2EDTA in 99.96% D_2O solution, 96 scans acquired with 5000-Hz spectral width over 16K complex data points, then zero filled to 32K points, multiplied by a Gaussian window function containing a resolution enhancement of −2 Hz before Fourier transformation. Arc, MI4, EA17, DN20, VI25, and EA43 melt (temperature at which 50% of the protein is denatured) at 34, 33, 36, 26, 31, and 33 °C, respectively (Vershon et al., 1986). The assignments shown for Arc are based upon sequential assignments made for PL8 and 2D NOESY and HOHAHA spectra obtained for Arc in D_2O solution.

except that 2.5 times more scans per increment were used in the TQF-COSY experiment, thus compensating for the reduced signal to noise with the increased level of multiple quantum filtration. The total number of increments collected for each experiment was 700. Phase-sensitive COSY spectra without multiple quantum filtration were used for the data acquired in H_2O solution by using an eight-step phase cycle.

The HOHAHA (Bax & Davis, 1985a) or TOCSY (Braunschweiler & Ernst, 1983) experiment was run by using the MLEV-17 (Bax & Davis, 1985b) pulse sequence for the spin lock interval. Both the 90° preparation pulse and the spin lock pulses were generated by using the low-power proton transmitter (1 W) present in the Bruker AM-500 console. The 90° pulse width was 47 μs, thus providing an effective field of 5.3 kHz. The MLEV-17 cycle was preceded and followed by two purge pulses of 1.6-ms duration, and the total spin lock interval was 80 and 66 ms for the experiments performed in D_2O and H_2O. A total of 534 increments each consisting of

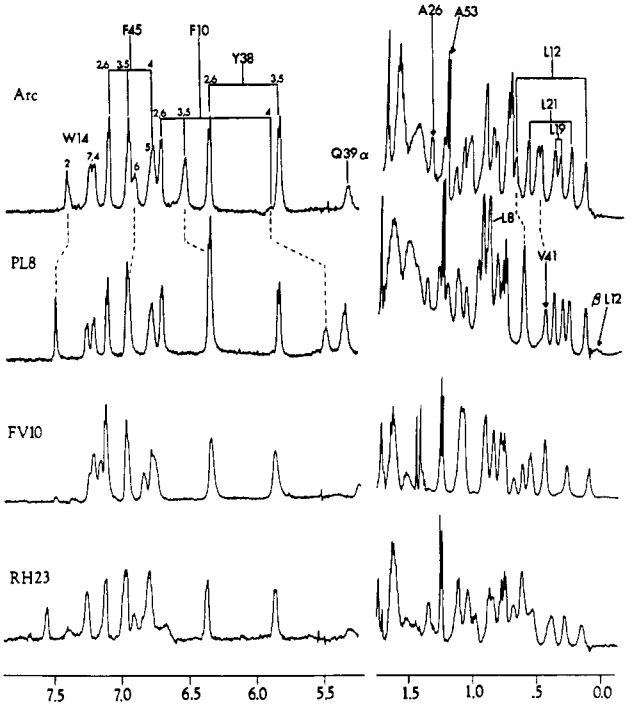


FIGURE 2: Expanded one-dimensional ^{1}H NMR spectra of Arc, PL8, FV10, and RH23. Conditions of sample preparation and NMR measurements were identical and are described in the caption to Figure 1. Those resonances of PL8 displaying different chemical shifts from Arc are connected by dashed lines. As determined previously by Vershon et al., (1986), Arc, PL8, FV10, and RH23 are 50% denatured at 34, 56, 33, and <5 °C, respectively.

96 scans were acquired for both experiments. For the experiment in H_2O solution the MLEV-17 sequence was followed by two 180° GROPE-16 pulses to eliminate the residual H_2O signal present in an homogeneous part of the field (M. G. Zagorski, manuscript in preparation). Thus, the modified pulse sequence became D_1–90°–$t1$–MLEV-17–180°–180°–$t2$, where presaturation of the H_2O signal was performed during the relaxation delay ($D1$). The two GROPE-16 pulses consisted of 270°–x–360°–x–90°–y–270°–y–360°–y–90°–x pulse trains.

RESULTS

Previously we determined the stabilities of Arc and several mutants to thermal unfolding by monitoring changes in their molar ellipticity by circular dichroism (CD) (Vershon et al., 1986). These studies were done at relatively dilute concentrations (50 μg/mL) and at low pH where denaturation is fully reversible. From these studies it was demonstrated that a mutant of Arc, PL8, which contained leucine for proline at position 8 had a 22 °C increase in melting temperature relative to wild-type Arc. For NMR studies of proteins it is often desirable to work at elevated temperatures as this reduces line widths and generally improves the quality of the data. Consequently, we chose to first determine the structure of the mutant PL8 using 2D NMR techniques. Furthermore, once the NMR assignments and structure of PL8 are complete, one can next move on to study Arc and other mutants as needed to achieve the eventual goal of studying the Arc–DNA complex in solution.

One-Dimensional Spectra of Arc Mutants. Before embarking on 2D NMR work of PL8, we first compared 1D spectra of Arc, PL8, and other mutant proteins. Expanded plots of the aromatic and upfield aliphatic (methyl) regions of these spectra are shown in Figures 1 and 2. The conditions for recording of the spectra were identical: pH 7.3, 25 °C, and 1 mM protein concentration. The missense mutations are

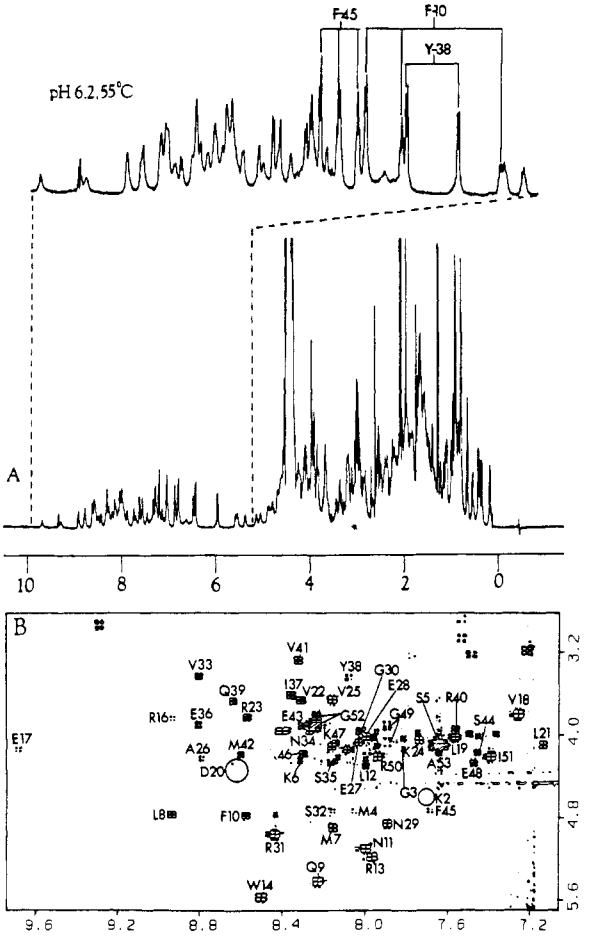


FIGURE 3: (A) One-dimensional ^1H NMR spectra of PL8 in H_2O solution at pH 6.2 and 55 °C. A Gaussian window function with a resolution enhancement parameter of −2.5 Hz was applied before Fourier transformation. The H_2O signal was suppressed by continous low-power (2-mW) irradiation before data acquisition. The concentration of PL8 was 4 mM. The upper trace is an expanded region showing the amide and aromatic signals with the assignments given for the ring protons of F10, Y38, and F45. (B) the αH–NH or "fingerprints" region of a phase-sensitive COSY spectrum acquired at 50 °C and pH 5.4. Cross peaks for αH–NH couplings are sequence specifically labeled and most of the remaining unlabeled cross peaks are side-chain lysine and arginine ε- and ζCH_2–NH couplings. Some extra peaks due to sample degradation are also present. Fortunately, none of these peaks interfered with making sequential assignments since they were not present in another earlier recorded absolute-value COSY experiment.

designated by the one-letter codes for the wild-type and mutant amino acids and the residue position of the sequence change. For example, the MI4 mutation changes the wild-type methionine at residue 4 of Arc to isoleucine. The assignments of the aromatic ring protons and methyl aliphatic signals in Arc are based upon the sequential assignments made for PL8 discussed later in this paper and upon 2D NOESY and HOHAHA data recorded in D_2O solution for both Arc and FV10. Typical HOHAHA contour plots of the aromatic region of Arc and FV10 in D_2O at 25 °C are plotted in panels A and B, respectively, of Figure 3.

For all mutants the aromatic and methyl signals are well dispersed, indicating that all proteins fold into ordered, nonrandom-coiled structures. MI4, EA17, DN20, VI25, EA43, and Arc all have nearly identical NMR spectra (Figure 1). This suggests that the tertiary, hydrophobic interactions in these proteins are similar. There is a slight change in the chemical shift for the 4H (para proton) of Phe-10 in DN20 and VI25 (Figure 1, left). The slight *upfield* shift in DN20 and VI25 may be the result of a local ring current effect, perhaps from the nearby Trp-14. In fact, for all the mutants studied (Figures 1 and 2, left) the 4H signal of Phe-10 is broad and shifted 0.7–1.0 ppm *upfield* relative to the 3,5H signals of Phe-10. Most of the upfield methyl signals have similar chemical shifts for mutants MI4, EA17, VI25, EA43, and Arc (Figure 1, right), although the methyls of Leu-21 change slightly for DN20. The differences in the environments of both Phe-10 and Leu-21 in DN20 may be related to its somewhat lower stability compared to MI4, EA17, VI25, EA43, and wild-type Arc (Vershon et al., 1986).

In Figure 2 the expanded aromatic and aliphatic regions are shown for Arc, PL8, FV10, and RH23. RH23 is the least stable of the mutants (Vershon et al., 1986) and is presumably unfolded at the temperature used for the NMR measurements (25 °C). The broader peaks in the RH23 spectrum demonstrate that it is indeed partially unfolded at room temperature. For FV10 there are a few differences in the shifts of Trp-14 such as the *upfield* shift of 2H and, of course, the disappearance of the Phe-10 resonances. In PL8, however, the 3,5H and 4H ring protons of Phe-10 shift *upfield* 0.2 and 0.4 ppm, respectively, whereas the 2H of Trp-14 shifts *downfield* 0.1 ppm. Other changes observed in the aliphatic region of PL8 include several *upfield* shifts of Leu-12 and Val-41 as shown in Figure 2. The large differences in shifts of the 4H and 3,5H of Phe-10 in PL8 compared to their shifts in Arc and the other mutants may be connected to its remarkable increase in thermal stability. These data will have to be borne in mind later for final three-dimensional structure determinations of PL8 and Arc.

Preliminary Analysis of PL8. In Figure 3A is shown the one-dimensional proton NMR spectrum of PL8 at 4 mM concentration in H_2O solution at pH 6.2 and 55 °C with the amide and aromatic region shown in an expanded plot. As the temperature was varied from 5 to 55 °C, no significant changes in the chemical shifts of the four aromatic residues or upfield aliphatic residues were observed, suggesting that there are no major changes in the overall folding of the molecule within this temperature range. Considerable broadening of the amide (NH) resonances occurred at lower temperatures, and many cross peaks in 2D NOESY data recorded at 45 °C became too broad to be distinguished from base-line noise. Since presumably Arc and PL8 are dimers in solution (Vershon et al., 1986), the loss of NOESY cross peaks at 45 °C could be due to the presence of a higher ordered oligomeric form (such as tetramer). Therefore, for the sample prepared at 4 mM concentration, we decided to first study by NMR its structure at temperatures of 50 and 55 °C. Degeneracies due to overlap of NH resonances were removed by varying the pH at 50 and 55 °C, as the NH chemical shifts are very sensitive to both temperature and pH. NOESY and COSY data sets were recorded in H_2O solution under three different conditions: (1) pH 4.3, 50 °C, (2) pH 5.4, 50 °C, and (3) pH 6.2, 55 °C. For these conditions, no major differences in the chemical shifts of the aromatic ring protons were observed, thus suggesting that there are no major changes within the hydrophobic interior of the molecule. All the two-dimensional NMR data presented in this paper will deal with the pH 5.4, 50 °C condition, and the remaining COSY and NOESY data recorded under the other two conditions are available on request.

Spin System Identification. The complete amino acid sequences of Arc and PL8 are shown in Chart I. PL8 is especially rich in several amino acids. For example, six glutamic acids, six arginines, five lysines, five valines, and four leucines (three leucines for Arc) account for half of the protein. Since

contacts between the nonpolar side chains will to a large extent determine the tertiary folding of the molecule, while the charged and more mobile lysine and arginine residues presumably maintain solubility and DNA binding affinities, we decided to undertake a complete and careful assignment of all amino acid residues within PL8. By use of the sequential assignment procedure as devised by Wüthrich and colleagues (Wüthrich et al., 1982; Billeter et al., 1982), the majority of the spin systems in PL8 were first identified from a combined analysis of the following two-dimensional data sets recorded in D_2O solution: double- and triple-quantum-filtered (DQF and TQF) COSY, homonuclear Hartmann–Hahn (HOHAHA), and NOESY. In general, data from the first three experiments which provide through-bond correlations were first utilized, and NOESY data which provides through-space information were used to corroborate these assignments.

As a first step in the analysis, the four aromatic residues (Phe-10, Trp-14, Tyr-38, Phe-45) were assigned. The ring protons of Tyr-38 and Phe-10 were fairly apparent, appearing as two isolated but slightly overlapping AA′XX′ and AA′XX′M spin systems shifted upfield (5.51–6.76 ppm) relative to the signals of Trp-14 and Phe-45 (6.84–7.31 ppm) in DQF-COSY and HOHAHA obtained in D_2O solution. The degenerate 3.5H and 4H of Phe-45 and 6H and 5H or Trp-14 were differentiated on the basis of couplings and NOEs from the labile 1-NH ring proton of Trp-14, seen in NOESY and COSY data recorded in H_2O solution. The aliphatic protons of the aromatic amino acids, all presumably AMX spin systems, were first tentatively located in NOESY data by searching for NOEs from the 2 and 6 ring protons of both phenylalanines and Tyr-38 to their corresponding βCH_2 and αH signals. Next, the β-protons were located by eliminating all the interresidue NOE cross peaks (all peaks upfield of 2.3 ppm). In most casss an unambiguous assignment of these protons was later made by using H_2O data.

The two alanine residues were easily located in DQF-COSY data by the presence of two intense cross peaks in the αH region (3.27–5.53 ppm) manifesting the βCH_3 to αH vicinal couplings. The *gem*-dimethyl moieties [$(CH_3)_2CH$] of the five valines were initially distinguished from those of the four leucines on the basis of the lower field position of the methine proton (1.9–2.4 ppm for valines and 1.2–1.9 ppm for leucines are typical). The shifts of the αHs were next found in HOHAHA spectra since all five valines displayed distinct four-bond relay patterns from both the γ-methyls and the β-methines to each αH (Figure 4A).

As discussed before by others (Muller et al., 1986; Rance & Wright, 1986), simplification of correlation data can be achieved by recording COSY experiments with multiple quantum filters. This technique takes advantage of the characteristic phase properties of multiple quantum coherence such that specific coupling patterns can be selectively suppressed or enhanced. For example, in Figure 4B, bottom right, is shown how the A_3MP spin system of Ile-51 is easily located in a TQF-COSY experiment. Both of the γ-methylene protons in isoleucine are nonequivalent due to the adjacent β-carbon which is asymmetric. Consequently, not only do the two diastereotopic γCH_2s couple with each other, but also each couples differently with the adjacent δCH_3. This three-spin system (A_3MP) is seen in a TQF-COSY whereas leucines and valines which possess only two-spin systems for methyl couplings cannot be seen. The remaining isoleucine (Ile-37) was not present in the TQF-COSY data since presumably its lines are too broad to be detected due to its being buried within the interior of the protein. However, some cross peaks for Ile-37

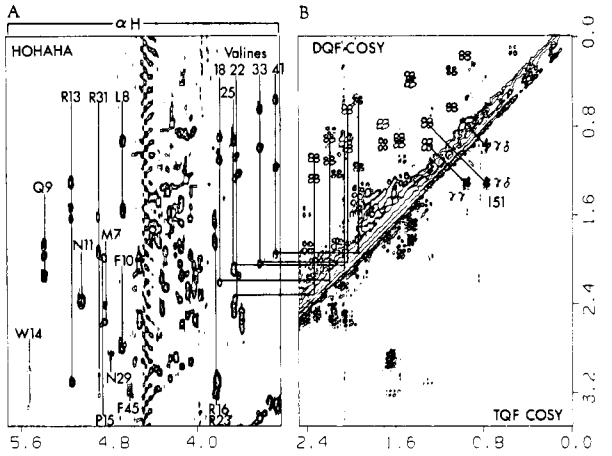


FIGURE 4: Two-dimensional data for PL8 recorded in D_2O solution at pH 5.4 and 50 °C. Initial spin system assignments for Ile-51 were made on the basis of the apparent A_3MP spin system seen in TQF-COSY data. The five valine spin system networks were discerned by use of DQF-COSY and relays present in HOHAHA (spin lock time 80 ms). Other assignments are as indicated.

were seen in the HOHAHA data, reflecting the greater efficiency of coherence transfer within the isotropic mixing period of this experiment. The rest of the CH_3 cross peaks in the DQF-COSY (Figure 4B, top left) had patterns typical for leucines; that is, as for the valines, two CH_3s displayed couplings to an identical methine proton. Another useful relay seen in the HOHAHA data was the methyl to methyl relays of the valines and leucines which were not seen for either of the isoleucines. The other remaining βCH_2s and αHs of the leucines were identified on the basis of relays seen in HOHAHA data. In addition, the assignment of Leu-8 was confirmed by direct comparison of 1D spectra with Arc, since in Arc, Leu-8 is missing.

Most of the arginines and lysines displayed, respectively, five- and six-bond relay peaks in HOHAHA data from the δ- and ϵCH_2s (2.9–3.2 ppm) to their corresponding αHs (3.8–5.5 ppm) (Figure 4A). This feature made them (together with Pro-15) readily discernible from the majority of the other amino acids. No NOE between the NH of Trp-14 and the δCH_2 of Pro-15 was detected, which may indicate that the 14–15 peptide bond is cis in configuration (Wüthrich et al., 1984). However, no NOE was detected between the αH of Trp-14 and the αH of Pro-15, and this implies a trans configuration about the 14–15 peptide bond.

Three of the four serines were identified with moderate certainty on the basis of their chemical shifts all being greater than 3.9 ppm and their distinct AMX patterns seem in DQF- and TQF-COSY spectra. The remaining intense cross peaks in DQF-COSY data in the 3.9–4.5 ppm region were assigned as glycines. Two of the four glycines, Gly-49 and Gly-52, were assigned from D_2O data and confirmed with H_2O COSY data since for each glycine both methylene protons coupled to an identical amide proton. The remaining glycines, Gly-3 and Gly-30, had to be found with sequential assignments in H_2O since both methylene protons were degenerate.

The most difficult residues to assign were the six Glu, two Gln, and four Met residues. Unfortunately, all of these residues overlapped considerably in DQF-COSY and HOHAHA data in D_2O solution. Two exceptions were Gln-9 and Met-7 in which the αH signals were shifted downfield (Figure 4A) since, as it turns out, both are part of extended β-structure. The remaining Met and Glx residues were assigned in D_2O data with provisions, since their assignments were made on the basis of elimination.

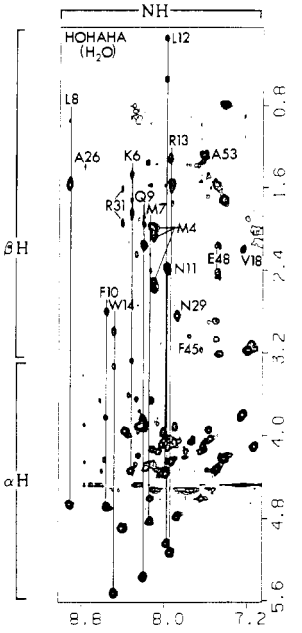


FIGURE 5: HOHAHA spectrum recorded in H_2O solution, spin lock time 66 ms. Relayed connectivities representing NH–αH–βH(s) through-bond interactions are sequence specifically identified. Those assignments for residues 6–14 are indicated by vertical lines.

Sequential Assignments. The αH–NH region or "fingerprint" region of the phase-sensitive COSY spectrum collected in H_2O solution at 50 °C and pH 5.4 is shown in Figure 3B. Each cross peak in sequence-specifically labeled. These assignments were made on the basis of interresidue patterns seen in NOESY spectra. As shown, cross peaks for Lys-2 and Asp-20 are absent, due to saturation of the water line by the proton decoupler. The cross peak for Asp-20 was more visible at lower contour levels and in other COSY data recorded at 55 °C and pH 6.2. In addition, a weak cross peak for Lys-2 was visible in absolute-value COSY data recorded in H_2O, and its assignment rested upon intraresidue NOEs from the NH at 7.57 ppm to β-, γ-, and δ-methylene protons.

Attempts at estimating $^3J_{NH-\alpha H}$ coupling constants from phase-sensitive COSY data gave values in the range 6.0–11.1 Hz. The COSY data set was obtained under conditions that favor narrowing of the NH resonances: elevated temperature (50 °C) and relatively low pH (5.4). Surprisingly, the NH peaks become broader at pH values less than 5.4; therefore, no further attempts were made to obtain additional data at lower pH values. In addition, the data were acquired and processed with high digital resolution (0.74 Hz/point). Normally, α-helical protein segments have $^3J_{NH-\alpha H}$ values in the range of 3.5–5.0 Hz whereas β-sheets have larger values of 8.9–9.7 Hz (Wüthrich, 1986). As will be shown later, NOE patterns indicate that PL8 is predominantly helical as demonstrated by numerous stretches of NN(i,i+1) NOEs. Thus, despite the optimal conditions of the COSY measurements, all the measured $^3J_{NH-\alpha H}$ values for the helical regions provide coupling constants that are too large. This apparent discrepancy must therefore be due to the broad line widths of the NH peaks (>10 Hz). Since the minimum separation between the antiphase components of a COSY cross peak is equal to approximately 0.58 times the line width at half-height (Neuhaus et al., 1985) the lower limiting value that can be measured for the line separation is in the range of 6.0–6.5 Hz. We consequently simulated each COSY cross peak in the fingerprint region by removing the line-width contributions to the measured antiphase peak separation. This was done for each cross peak by measuring both the line widths and

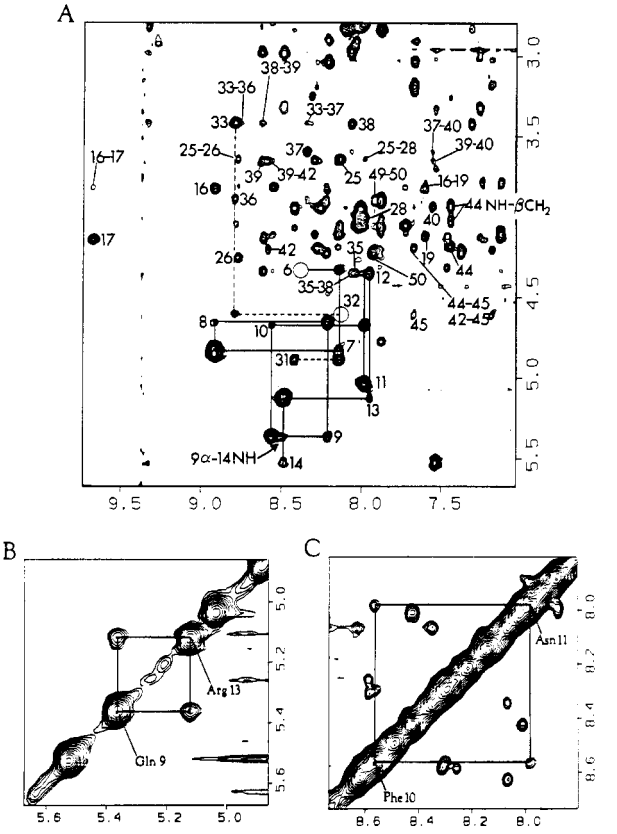


FIGURE 6: (A) Sequential αH–NH connectivities shown in an expanded NOESY (100-ms) plot of the fingerprint region. Intraresidue NOEs are labeled with numbers directly next to the cross peak. Residues 6–14 and 31–33 which contain stretches of αN(i,i+1) NOEs are connected by lines. Other αH–NH NOEs between nonadjacent residues are also shown. (B and C) Expanded regions of NOESY (H_2O solution, 100-ms mixing time) showing (B) the αH–αH connection between Gln-9 and Arg-13 and (C) NH–NH connection between Phe-10 and Asn-11.

antiphase peak separations. The resulting $^3J_{NH-\alpha H}$ values were then placed into two categories: "large" being >7.0 Hz and "small" being <7.0 Hz.

A useful starting point for making sequential assignments was Leu-8 as this assignment is obviously known with great accuracy (Arc has a Pro instead of Leu at position 8). The NH assignment of Leu-8 was confirmed with relays from the NH to the βCH_2s and to one δCH_3 in HOHAHA spectra recorded in H_2O solution (Figure 5). Further, intraresidue NOEs from the NH to the other protons of Leu-8 were present in NOESY data. This secured the location of the NH of Leu-8 at 8.93 ppm. The αH of Leu-8 (4.66 ppm) has NOEs to two labile NH protons: one intraresidue NOE to its own NH and another stronger interresidue NOE to the Gln-9 NH (Figure 6A). This latter NOE is an αN(i,i+1) interaction between αH of Leu-8 and the NH of Gln-9. As for the NH assignment of Leu-8, the assignment of the NH of Gln-9 was confirmed by using HOHAHA and NOESY data, i.e., relays from the NH to the βCH_2s and γCH_2s (Figure 5). Following this assignment procedure, a path or "walkway" (Wüthrich, 1986) of αH(i) to NH(i+1) NOEs could be traced from Lys-6 to Trp-14 (Figure 6A). All residues within the Lys-6 to Trp-14 segment displayed NH to βCH_2 relays in the HOHAHA data (Figure 5) and $^3J_{NH-\alpha H}$ values greater than 7.0 Hz.

In the αH–αH region there was one observable intense NOE between Gln-9 and Arg-13 (Figure 6B). Further, amide to amide NOEs were detected: Phe-10 to Asn-11 (Figure 6C) and Trp-14 to Arg-16 (Figure 7A). The latter interaction is an NN(i,i+2) NOE across Pro-15 and was present in

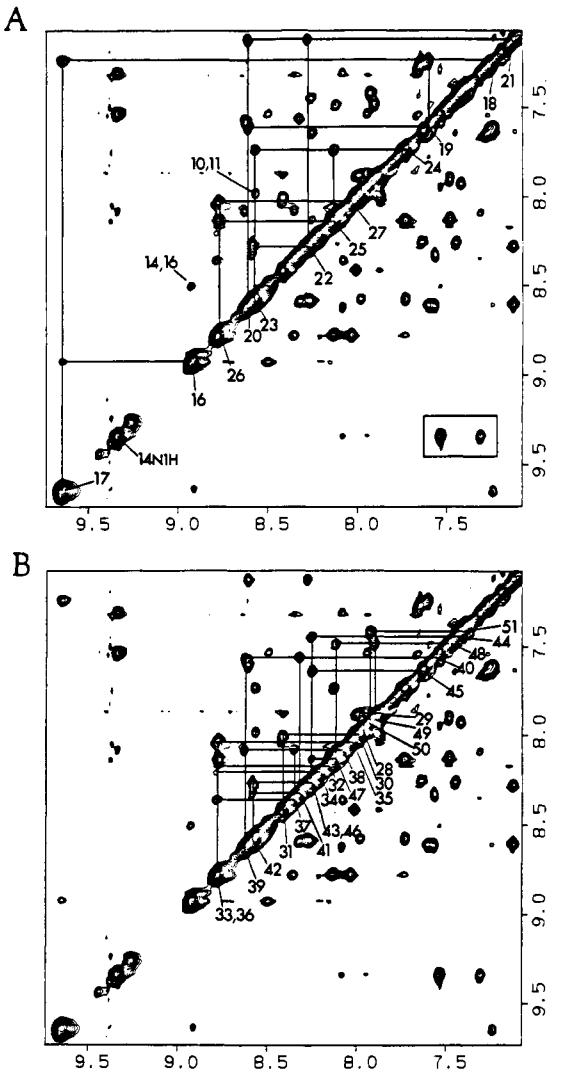


FIGURE 7: (A) Expanded NOESY (200-ms) plot of the NH–NH region. NN(i,i+1) NOEs among residues 10 and 11 and 16–27 are shown with NH peaks for the helical segment labeled along the diagonal. Also, an NN(i,i+2) NOE from residue 14 to 16 is shown. (B) NOESY (200-ms) plot of NH–NH region showing many sequential NN(i,i+1) NOEs among residues 28–51. NH peaks are labeled along the diagonal.

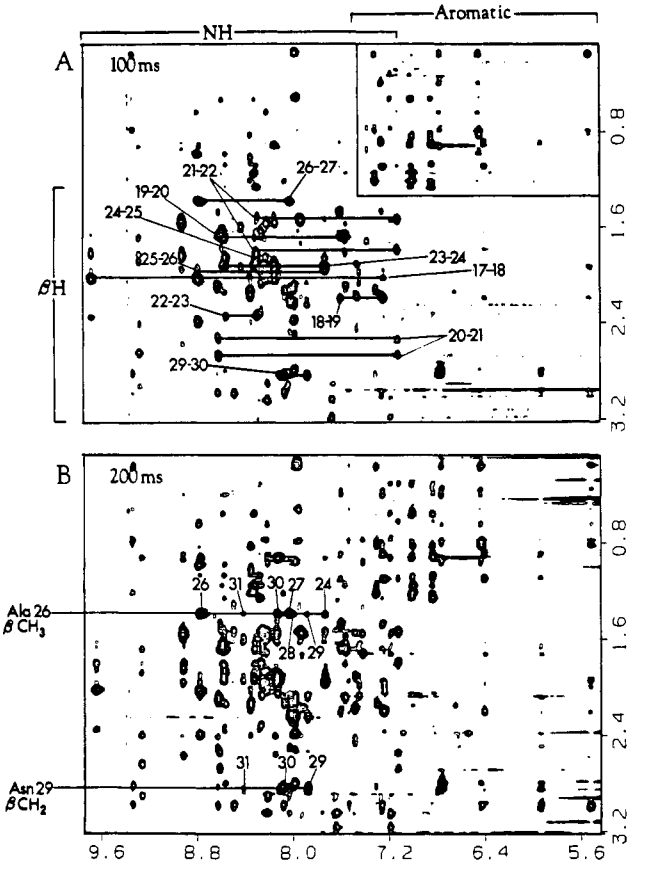


FIGURE 8: (A) Sequential βN(i,i+1) connectivities covering residues 17–27 in the shorter mixing time (100-ms) NOESY spectra. (B) In the longer mixing time NOESY (200-ms) spectra, several additional long-range (between nonadjacent residues) βN NOEs are observed. Those from Ala-26 and Asn-29 are shown in the lower plot. These latter NOEs were useful for confirming sequential assignments made with the other backbone protons. Identical acquisition and processing parameters were used for both NOESY experiments.

NOESY spectra run with both short (100-ms) and long (200-ms) mixing times. No NOEs were seen from any of the exchangeable protons to Pro-15.

The NH of Arg-16 was assigned on the basis of an NN(i,i+2) NOE from the NH of Trp-14. Continuing on, a stretch of sequential NN(i,i+1) connectivities could be traced from Arg-16 to Glu-27 (Figure 7A). This type of interresidue NOE pattern is typical for α-helices. In general, helices display not only NN(i,i+1) but also βN(i,i+1) NOE patterns (Englander & Wand, 1987; DiStefano & Wand, 1987). This allows one to make checks on a NN(i,i+1) path, as simultaneously a βN(i,i+1) path can be followed. In Figure 8A the latter tracing was followed from Glu-17 to Glu-27 by using shorter mixing time (100-ms) NOESY data.

The region within residues 28–35 was by far the most difficult for making sequential amino acid assignments, since many NH and αH cross peaks were weak and overlapped. Since the NN(i,i+1) path terminated with Glu-27, the NH of the next residue (Glu-28) was found by searching for other NOEs among the backbone protons. For example, in the fingerprint region of the short mixing time (100-ms) NOESY (Figure 6A), the αH of Val-25 has NOEs to three NH protons: an intraresidue NOE to its own NH, an αN(i,i+1) NOE to the NH of Ala-26, and an αN(i,i+3) or an αN(i,i+4) NOE to the NH of Glu-28 or Asn-29, respectively. This latter NOE was reasoned to be an αN(i,i+3) interaction since for the αH that coupled to this NH (Figure 3B), an Asn spin system did not fit the cross peak pattern seen in DQF-COSY or HOHAHA D_2O data (Figure 4). Next, NN(i,i+1) connectivities allowed a path from Glu-28 to Arg-31 to be delineated (Figure 7B). Most of the NH(i) to NH(i+1) NOEs were weak except between Gly-30 and Arg-31.

The αH chemical shifts for Asn-29 and Arg-31 appear downfield (4.77 and 4.89 ppm, respectively) relative to the remaining αHs (3.41–4.07 ppm) within the 28–35 amino acid region of PL8. Also, $^3J_{NH-\alpha H}$ values were greater than 8 Hz for these residues, and only Asn-29 and Arg-31 displayed relays in HOHAHA data from their NHs to all side-chain protons (Figure 5). None of the remaining residues in this region showed relays in the HOHAHA data, which may be due to the weaker $^3J_{NH-\alpha H}$ values. Additionally, the βCH_2s of Asn-29 showed sequential NOEs to the NH of Gly-30 (100-ms mixing time NOESY data; Figure 8A) and to the NH of Arg-31 (200-ms mixing time NOESY data; Figure 8B). These latter NOEs secured the assignment of residues 29–31.

Other NOEs among the βHs and NHs were used to confirm sequential assignments made with only the αH and NH protons. Ala-26 was especially useful since the βCH_3 had NOEs to the NHs of many neighboring residues, namely, 24, 26, 27, 28, 29, 30, and 31 (Figure 8B). Many of these interactions

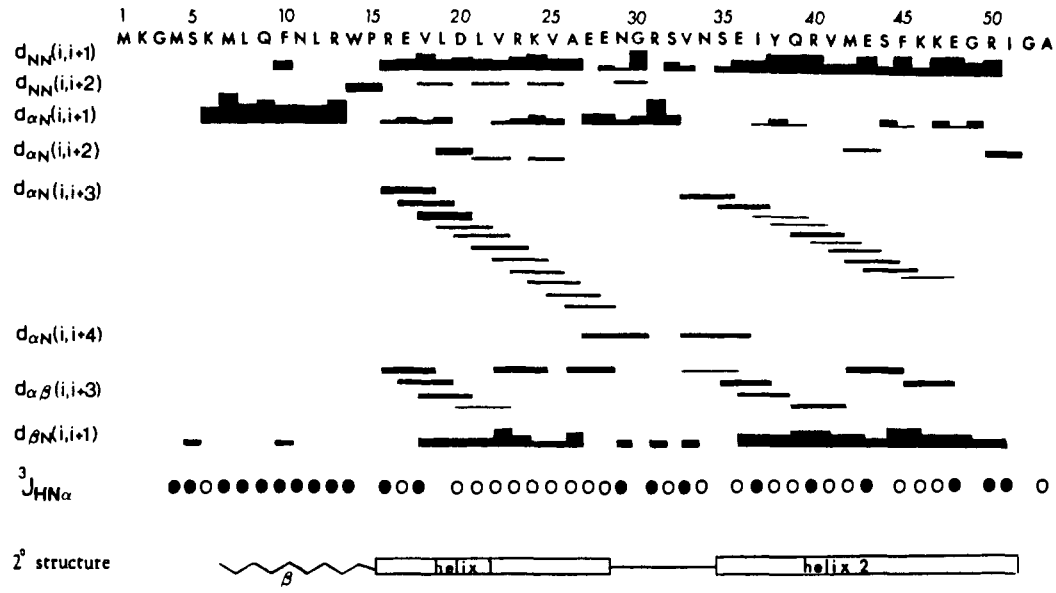


FIGURE 9: A summary of the sequential NOEs observed in PL8. The intensities of the NOEs are classified according to the thickness of the bar heights. The NOEs shown are a combination of 100-ms mixing time data acquired at pH 5.4 and 50 °C. $^3J_{NH-\alpha H}$ values classified as large (>7.0 Hz) or small (<7.0 Hz) are indicated by closed and open circles, respectively.

could be due to spin diffusion, being present in 200-ms (Figure 8B) but not 100-ms (Figure 8A) mixing time NOESY data. Nevertheless, the spin diffusion of Ala-26 was useful for reinforcing assignments made previously with other backbone protons.

Ser-32 was located by using an intense $\alpha N(i,i+1)$ NOE from the αH of Arg-31 to the NH of Ser-32 (Figure 8A). Unfortunately, the αH–NH COSY cross peak (Figure 3B) of Ser-32 was weak and close to the H_2O signal, despite changes made in pH or temperature. This initially made location of Ser-32 difficult. No NOE was seen from the NH of Arg-31 to the NH of Ser-32, although a weak $NN(i,i+2)$ NOE was seen from Gly-30 in 200-ms NOESY data (Figure 7B). The adjacent Val-33 was located on the basis of two NOEs from Ser-32: a medium-sized $\alpha N(i,i+1)$ and a weak $NN(i,i+1)$, shown in Figures 6A and 7B, respectively. Continuing on the sequence toward the C-terminus, the αH of Val-33 had NOEs to several NHs (Figure 6A): an intraresidue NOE to its own NH, an $\alpha N(i,i+3)$ NOE to Glu-36 which was seen more clearly in the pH 4.3, 50 °C data, and a weak $\alpha N(i,i+4)$ NOE to Ile-37. Also, an extremely weak $NN(i,i+1)$ NOE between Val-33 and Asn-34 was seen in the NH–NH region of the 200-ms NOESY spectrum (Figure 7B). However, this latter weak NOE and a medium-sized $\beta N(i,i+1)$ NOE from Val-33 were the only links to the NH of Asn-34. Further evidence suggesting that the NH of Asn-34 resided here was the presence of two intense NOEs at 2.90 and 3.04 ppm which are NH to βCH_2 intraresidue NOEs characteristic of asparagine residues; i.e., these same strong intraresidue NOEs were also seen for Asn-11 and Asn-29. Unfortunately, no interresidue NOEs from any of the backbone protons of Asn-34 were detected. This made locating the next residue, Ser-35, difficult. Location of the NH of Ser-35 relied upon an NOE observed at 55 °C and pH 6.2: a weak $\alpha N(i,i+4)$ NOE from the αH of Arg-31 to the NH of Ser-35. However, the sequential assignment of Ser-35 was ratified by two additional NOEs from the αH of Ser-35 to the NH and βCH_2s of Tyr-38 $\alpha N(i,i+3)$ (Figure 6A) and $\alpha\beta(i,i+3)$ interactions, respectively.

The peptide segment within residues 36–50 was characterized by stretches of the following contiguous NOEs: medium-sized $NN(i,i+1)$ (Figure 7B), weak $\alpha N(i,i+1)$, and medium-sized $\alpha N(i,i+3)$ and $\alpha\beta(i,i+3)$. These data indicate that PL8 is predominately helical between residues 36 and 50. In addition, assignments of residues 36–50 were simplified by elimination as in most cases only one amino acid type remained for each residue. For example, there are five valines in PL8 and all but Val-41 have been sequentially assigned. Since the αH shifts of the five valines are known from DQF-COSY (Figure 4B) and HOHAHA (Figure 4A) data obtained in D_2O solution, the remaining Val-41 was easily located in the αH–NH fingerprint region (Figure 3C).

The last three residues (Ile-51, Gly-52, and Ala-53) showed no interresidue NOEs among any exchangeable NH protons. This may be due to fraying at the C-terminus of the protein. Nonetheless, the sequence-specific assignments of Ile-51, Gly-52, and Ala-53 were fairly straightforward since at this stage of the assignment procedure most of the other amino acids were already assigned; consequently cross peaks for these last three residues were known by elimination. In addition, the line widths of the NHs obtained from cross sections in the αH–NH COSY data were considerably narrower for these last residues, further suggesting their location at the C-terminus.

Table I summarizes the complete proton NMR assignments for PL8, and in Figure 9 the observed interresidue NOE connectivities are presented.

Long-Range Contacts. In order to obtain a more complete picture of the residues involved in folding, a complete and rigorous assignment of all aliphatic and aromatic residues was undertaken, since identification of all interproton distances within the hydrophobic core is needed for an accurate three-dimensional structure determination. In Figure 10 is shown an expanded region of the 100-ms NOESY spectrum (boxed-in area of Figure 8A) of PL8 obtained in H_2O solution, which establishes distance connectivities between aromatic ring protons and methyl aliphatic protons. The NOESY data shown in Figure 10, obtained in H_2O solution, were virtually identical with NOESY data obtained in D_2O solution, thus demonstrating the high quality of the H_2O suppression. Presumably, PL8 has a well-defined hydrophobic core since it is quite thermally stable. Interresidue NOEs from aromatic ring protons are summarized in Table II. Clearly, many nonpolar contacts between amino acids far apart in the primary

Table I: Proton NMR Resonance Assignments for PL8 at 50 °C and pH 5.4[a]

residue	NH	αH	βH	other		residue	NH	αH	βH	other	
M1		4.13	2.14, 2.14	γCH_2 2.57, 2.53	ϵCH_3 2.13	L21	7.12	4.10	1.82, 1.54	γCH 1.09 δCH_3 0.31	δCH_3 0.63
K2	7.57	4.42	1.78, 1.69	γCH_2 1.69, 1.69	δCH_2 1.47, 1.44	V22	8.27	3.64	2.34	γCH_3 1.26, 1.08	
G3	8.39	4.14, 4.05		ϵCH_2 3.01, 3.01	ζNH_2 7.49	R23	8.56	3.82	1.92, 1.83	γCH_2 1.68, 1.68 ϵNH 7.19	δCH_2 3.10, 3.09
M4	8.11	4.52	2.09, 2.00	γCH_2 2.58, 2.50	ϵCH_3 2.12	K24	7.73	4.05	1.94, 1.93	γCH_2 1.40, 1.34 ϵCH_2 2.93, 2.93	δCH_2 1.63, 1.52 ζNH_2 7.49
S5	7.63	4.03	3.91, 3.89			V25	8.13	3.65	2.06	γCH_3 0.94, 0.94	
K6	8.33	4.32	1.87, 1.83	γCH_2 1.73, 1.73	δCH_2 1.48, 1.49	A26	8.78	4.24	1.39		
				ϵCH_2 3.03, 3.01	ζNH_2 7.49	E27	8.02	4.07	2.24, 2.13	γCH_2 2.50, 2.33	
M7	8.14	4.82	1.98, 1.57	γCH_2 2.56, 2.41		E28	7.98	4.00	2.18, 2.15	γCH_2 2.55, 2.27	
L8	8.93	4.66	1.61, 1.53	γCH 1.58 δCH_3 0.91	δCH_3 0.94	N29	7.88	4.77	2.89, 2.84	δNH_2 8.11, 6.78	
Q9	8.22	5.36	1.97, 1.87	γCH_2 2.16, 2.15	ϵNH_2 7.27, 6.65	G30	8.00	3.95, 3.95			
F10	8.57	4.67	2.80, 2.80	2,6H 6.76 4H 5.51	3,5H 6.43	R31	8.42	4.89	1.98, 1.94	γCH_2 1.62, 1.61 ϵNH 7.75	δCH_2 3.23, 3.04
N11	7.98	5.04	2.37, 2.37	δNH_2 7.09, 6.54		S32	8.14	4.60	4.09, 4.09		
L12	7.97	4.36	0.52, 0.12	γCH 0.97 δCH_3 0.17	δCH_3 0.79	V33	8.79	3.41	2.04	γCH_3 0.99, 0.65	
R13	7.95	5.13	1.63, 1.53	γCH_2 1.30, 1.30	δCH_2 3.11, 3.11	N34	8.21	3.91	3.04, 2.90	δNH_2 6.75	
				ϵNH 7.16		S35	8.04	4.36	4.05, 3.98		
W14	8.50	5.53	3.33, 2.99	1NH 9.33	2H 7.54	E36	8.79	3.89	2.02, 1.90	γCH_2 2.41, 2.41	
				4H 7.26	5H 6.84	I37	8.35	3.60	2.14	γCH_3 1.22 δCH_3 0.83	γCH_2 2.18, 2.55
				6H 7.00	7H 7.31	Y38	8.08	3.43	2.98, 2.81	2,6H 6.41	3,5H 5.93
P15		4.85	2.56, 1.97	γCH_2 2.19, 1.94	δCH_2 3.72, 3.36	Q39	8.63	3.66	2.10, 2.08	γCH_2 2.50, 2.42	ϵNH_2[b]
R16	8.92	3.83	1.85, 1.83	γCH_2 1.68, 1.65	δCH_2 3.20, 3.20	R40	7.57	3.93	1.80, 1.70	γCH_2 1.62, 1.58 NH[b]	δCH_2[b]
				ϵNH 7.46		V41	8.32	3.27	1.94	γCH_3 1.16, 0.56	
E17	9.64	4.14	2.03, 1.83	γCH_2 2.36, 2.36		M42	8.58	4.22	1.64, 1.74	γCH_2 2.41, 2.14	ϵCH_3 2.14
V18	7.24	3.79	2.20	γCH_3 1.10, 0.90		E43	8.27	3.97	2.09, 2.00	γCH_2 2.32, 2.32	
L19	7.60	4.12	1.68, 1.47	γCH 1.47 δCH_3 0.35	δCH_3 0.41	S44	7.45	4.18	4.02, 3.91		
D20	8.61	4.34	2.68, 2.54			F45	7.63	4.60	3.17, 3.04	2,6H 7.19 4H 6.83	3,5H 7.01
						K46	8.26	4.18	1.97, 1.87	γCH_2 1.39, 1.35 ϵCH_2 2.91, 2.89	δCH_2 1.63, 1.52 ζNH_2 9.26
						K47	8.12	4.12	1.91, 1.84	γCH_2 1.44, 1.41 ϵCH_2 2.96, 2.96	δCH_2 1.67, 1.64 ζNH_2[b]
						E48	7.48	4.31	2.17, 2.17	γCH_2 2.46, 2.41	
						G49	7.90	4.05, 3.89			
						R50	7.93	4.23	1.76, 1.57	γCH_2 1.55, 1.53 ϵNH 7.51	δCH_2 3.07, 2.90
						I51	7.42	4.21	1.72	γCH_3 0.77 δCH_3 0.76	γCH_2 1.32, 0.96
						G52	8.22	3.91, 3.82			
						A53	7.62	4.09	1.27		

[a] Chemical shifts are reported in ppm relative to the H_2O or HDO signal at 4.52 or 4.50 ppm. [b] These peaks were unassignable due to overlap or exchange with solvent.

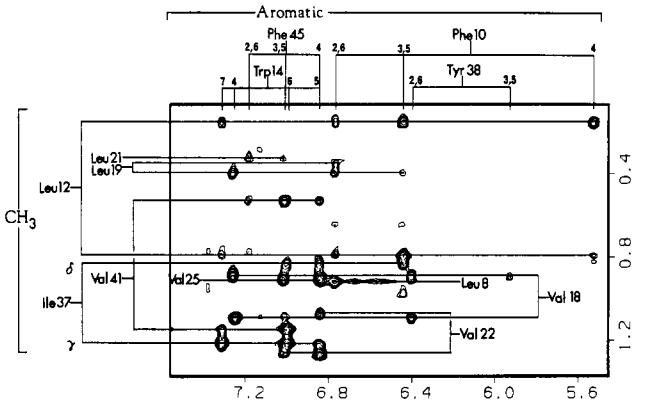
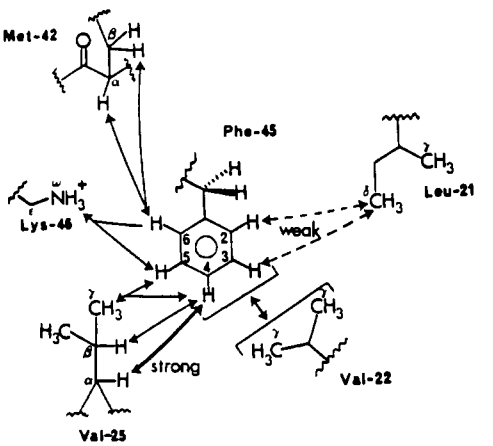


FIGURE 10: Assignment of NOE contacts between aliphatic and aromatic residues. This figure shows an expanded plot (boxed in section of Figure 8A) of the 100-ms NOESY spectrum demonstrating interproton interactions between the aromatic ring protons and the methyl aliphatic protons.

Chart II



structure are apparent. These interactions are outlined in the next four sections as to their origin from the particular aromatic ring system.

Phe-10. The methyls of three leucine residues account for all the interresidue NOEs to the ring protons of Phe-10. These include Leu-8, Leu-12, and Leu-19. Only one methyl of Leu-8 shows interactions with the 2,6H but not the 3,5H or 4H of Phe-10, and only one methyl group of Leu-12 interacts with the 4H of Phe-10. No other NOEs to the 4H of Phe-10 were detected. It should be kept in mind that the 4H of Phe-10 in PL8 shows the largest chemical shift variance (0.4 ppm)

Table II: Summary of Additional Long-Range NOEs from Aromatic Ring Protons[a]

aromatic		other	
residue	proton(s)	residues	proton(s)[b]
F10	2,6	L8	δCH_3
		L12	δCH_3, $\delta' CH_3$
		L19	δCH_3, $\delta' CH_3$
		N11	NH
	3,5	L12	δCH_3
		L19	δCH_3, $\delta' CH_3$
	4	L12	δCH_3
W14	1NH	L12	δCH_3, $\delta' CH_3$
		Y38	3,5H, βCH_2, αCH
	2	L12	αCH
	4	L19	δCH_3, γCH, βCH_2, αCH
		V18	γCH_3, $\gamma' CH_3$
	5	V41	γCH_3
		I37	δCH_3, γCH_3
		V18	γCH_3, $\gamma' CH_3$
	6	V41	γCH_3, $\gamma' CH_3$
		I37	δCH_3, γCH_3
		V18	γCH_3
	7	V41	$\gamma CH_3'$
		I37	γCH_3
		L12	δCH_3, $\delta' CH_3$
		Y38	3,5H, βCH_2, αH
Y38	2,6	V18	γCH_3, $\gamma' CH_3$
		W14	βCH_2, αCH, NH
	3,5	S35	αCH
		Q39	αCH, NH
		V18	γCH_3, $\gamma' CH_3$
F45	2,6	L21	δCH_3
		M42	βCH_2, αCH
		K46	ωNH_3
	3,5	L21	δCH_3
		K46	ωNH_3
		V25	γCH_3
		V22	γCH_3, $\gamma' CH_3$
	4	V25	γCH_3, βCH, αCH
		V22	γCH_3, $\gamma' CH_3$

[a] Interproton NOEs obtained from NOESY (100-ms) data recorded in H_2O solution at 50 °C and pH 5.4 (see Figure 10). [b] Methyls of leucines and valines are distinguished with primes (i.e., δCH_3 and $\delta' CH_3$).

when compared to Arc (Figure 2).

Phe-45. Several long-range contacts to the ring protons of Phe-45 were identified and are summarized in Chart II, where arrows are representative for the NOEs. It should be kept in mind that the arrows, as drawn for NOE contacts, are not stereospecific for the 2,6H and 3,5H ring protons as these signals are chemically shift equivalent under the conditions used for NMR measurements. Interactions from the αH and one βCH_2 proton of Met-42 to the 2,6H of Phe-45 are probably not long-range contacts but instead may be interactions between residue i and $i + 3$ within a turn of an α-helix. Nevertheless, this latter NOE enabled us to make stereospecific assignments, since only one βCH_2 of Met-42 showed an NOE to Phe-45. Thus, the β-methylene proton (1.74 ppm) of Met-42 which interacted with Phe-45 also displayed an intraresidue NOE to its own αH, while the other β-methylene proton (1.64 ppm) did not show an NOE to its own αH. A medium-intensity NOE from the αH of Met-42 to one (not both) βCH_2 of Phe-45 was also observed. This β-methylene proton (3.04 ppm) of Phe-45 which interacts with Met-42 also showed intraresidue NOEs to the αH and NH of Phe-45 whereas the other β-methylene proton (3.17 ppm) of Phe-45 showed an NOE to its own NH but not the vicinal αH. Thus, the βCH_2s of both Met-42 and Phe-45 are now stereospecifically assigned. Other contacts from Leu-21, Val-22, and Val-25 to Phe-45 are diagnostic for long-range NOEs.

Trp-14 and Tyr-38. Trp-14 seems to play a key role in the folding of PL8 in solution. NOEs observed to Trp-14 and Tyr-38 are summarized in Charts III and IV. The αHs of both Ser-35 and Gln-39 show NOEs to the 3,5H of Tyr-38 (Table II). Ser-35, Tyr-38, and Gln-39 are all polar residues and are close in primary structure. The NOEs among these residues are omitted in Chart IV. However, the interactions between Ser-35 and Tyr-38 permitted stereospecific assignments to be made for Tyr-38. Thus, the αH of Ser-35 which interacted with the 3,5H of Tyr-38 also interacted with one βCH_2 (2.81 ppm) of Tyr-38. This β-methylene proton of Tyr-38 exhibited an NOE to the αH of Tyr-38, but the other β-methylene proton (2.98 ppm) of Tyr-38 did not display a similar intraresidue NOE to the αH. Trp-14 seems to be a focal point for the hydrophobic core as many nonpolar residues display numerous NOE contacts to the Trp aromatic ring. These include Leu-12, Leu-19, Ile-37, and Val-41 (Chart III). In addition, Trp-14 and Tyr-38 display many contacts to each other: 2,6H of Tyr-38 to αH and βCH_2 of Trp-14, and 1NH and 7H of Trp-14 to αH, βCH_2, and 3,5H of Tyr-38 (Chart IV). In addition, both Trp-14 and Tyr-38 have numerous interactions to Pro-15 and Val-18 (Chart IV).

Chart III

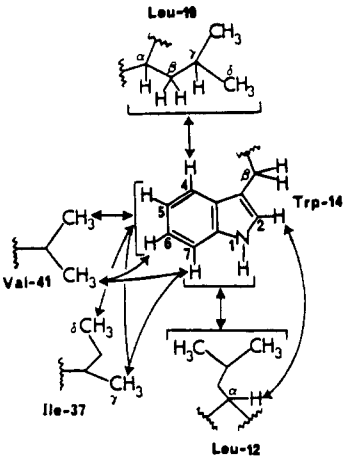
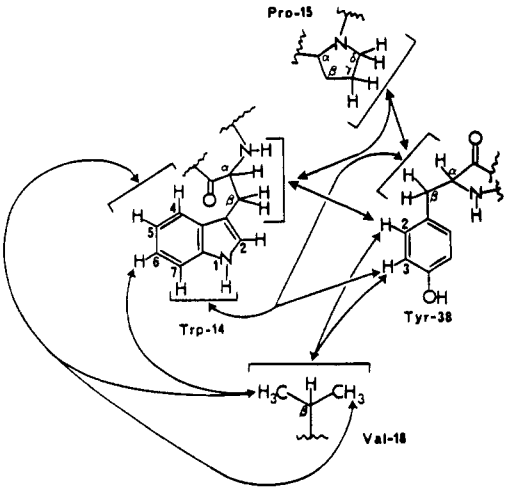


Chart IV



Discussion

Wild-Type Arc and Mutants. In a previous report, Vershon et al. (1986) determined the stabilities of Arc and its mutants toward both guanidine hydrochloride induced denaturation and temperature. Both of these experiments monitor the equilibrium between folded and unfolded structures of the mutant

proteins but do not provide specific details about which amino acid residues may be responsible for the changes in stability. A useful indicator of local perturbations are proton NMR chemical shifts which complement the spectrophotometric data.

The one-dimensional proton NMR spectra of Arc and mutant proteins MI4, EA17, and EA43 suggest that all have identical hydrophobic cores since there are no significant differences in the chemical shifts of the aromatic ring or aliphatic methyl protons (Figure 1). However, mutant proteins DN20 and VI25 show slight differences in the chemical shift of the 3,5H and 4H of Phe-10 (Figure 1). As it turns out, both DN20 and VI25 are somewhat less stable relative to MI4, EA17, EA43, and wild-type Arc (Vershon et al., 1986). The amino acid change in DN20 which substitutes Asp-20 for Asn-20 is interesting since no long-range NOE contacts ar seen to Asp-20 in PL8. However, numerous interactions are found between Leu-19 and Phe-10. One possibility is that Asn-20 weakens the hydrophobic contacts between Leu-19 and Phe-10. This may account for the lower stability of DN20 relative to wild-type Arc.

PL8 has the largest variations in chemical shifts of the aromatic resonances relative to Arc (Figure 2). PL8 also has the largest divergence in thermal stability and guanidine hydrochloride induced denaturation (Vershon et al., 1986). The differences seen by NMR are most pronounced for Phe-10 and Trp-14. For example, the 4H and 3,5H of Phe-10 shift *upfield* while the 2H of Trp-14 shifts slightly *downfield*. There are also variatons in the chemical shifts of several aliphatic resonances, notably, *upfield* shifts of both methyls and one β-methylene proton of Leu-12 (Figure 2). These results establish that the substitution Pro-8 $\rightarrow$ Leu effects the conformation of the adjacent Phe-10–Asn-11–Leu-12–Arg-13–Trp-14 segment which may in part be responsible for the increased thermal stability of PL8 relative to Arc and the other mutant proteins.

In all the mutants examined by NMR, no significant differences in shifts were observed for the Phe-45 or Tyr-38 ring protons. As discussed later, both Phe-45 and Tyr-38 are residues within the same helix.

Secondary Structure Determination. Regular secondary structure elements of proteins can be easily determined from a qualitative interpretation of specific short-range NOE patterns found among the backbone protons (Wüthrich, 1986; Clore & Gronenborn, 1987). In general, helices are characterized by stretches of strong or medium NN(i,i+1), αN(i,i+3), and $\alpha\beta$(i,i+3) NOEs and weak αN(i,i+1), αN(i,i+4), and NN(i,i+2) NOEs. On the other hand, a β-structure is characterized by strong αN(i,i+1) NOEs and the absence of other short-range NOEs from the backbone protons. β-Sheets can be aligned from interstrand NOEs involving αH or NH protons; i.e., antiparallel β-sheets are characterized by NOEs between nonadjacent αHs.

Interpretation of the long-range distance connectivities for Arc and its mutants is complicated by the proteins being symmetrical dimers in solution. It was previously established by gel filtration chromatographic techniques that wild-type Arc and its mutants (including PL8) elute at positions expected for dimers (Vershon et al., 1986). These chromatographic studies were done at dilute concentrations (0.007 mM) relative to the present two-dimensional NMR study of PL8 (4 mM). Since the overall equilibrium between monomer and dimer species is a function of protein concentration with the dimer being favored at higher concentrations, PL8 is almost certainly a dimer for the solution NMR work described herein. Further, the broad NH line widths (>10 Hz) seen in cross sections of phase-sensitive COSY data (Figure 3B) suggest that the protein is dimeric but not monomeric, since typically for a (monomer) protein of 53 residues the NH line widths are generally much narrower (Wüthrich, 1986). The existence of a higher ordered oligomer (such as a tetramer) was ruled out since, for such species, due to reductions in correlation times, the NMR line widths would have been too broad to permit detection. Thus, for each dimer composed of two *symmetrical* monomer units, two types of interproton NOEs can be expected: between protons within the same monomer or between protons of different monomers. Consequently, many of the NOEs observed may be dimer contacts. We refer to NOEs between protons within the same monomer as "monomer NOEs" and those arising between protons of different monomers as "dimer NOEs". Work is currently in progress to discriminate between these two type of interactions.

Despite the problem of distinguishing between monomer and dimer NOEs, we will assume that most backbone proton–proton NOEs (among β, α, and NHs) are "monomer NOEs". This assumption permits us to obtain a secondary structure of PL8 in solution. Thus, the secondary structure of PL8 as derived from the sequential backbone NOEs shown in Figure 9 is composed of a β-structure (residues 6–14) and two α-helices (residues 16–26 and 35–50). The secondary structure for PL8 determined from the NMR data described herein was virtually identical with recent predictions for Arc from analysis of structurally allowed substitution patterns (Bowie & Sauer, 1989). The different secondary structures are discussed.

N- and C-Terminal Ends. The first five residues at the N-terminus and the last three residues at the C-terminus are flexible in solution and do not assume an ordered structure as no interresidue NOEs were observed for any of these residues. This may be due to fraying at both ends of the protein. Nevertheless, the first five residues are important for operator recognition and DNA binding (Vershon et al., 1986). For example, Arc mutants KT2, GR3, MI4, and SC5 show decreased DNA binding affinities relative to that of wild-type Arc. The lack of a defined structure for these terminal residues may be due to the elevated temperatures used for NMR measurements. These residues, however, may assume a more ordered structure only when bound to DNA. Additional work at lower temperaturs is planned to help distinguish among these possibilities.

β-Structure (Residues 6–14). As mentioned before, residues near the N-terminus are intimately involved in operator binding (Vershon et al., 1986). In fact, more recent studies (Knight & Sauer, 1989) have demonstrated that the binding specificity of the related protein Mnt can be switched to that of Arc by replacing six residues at the N-terminus of Mnt with the corresponding nine N-terminal residues from Arc. In PL8, interresidue NOEs within residues 1–5 are too weak to be observed at 50 °C, probably due to the mobility of the N-terminus. However, residues 6–14 form part of well-defined β-structure in solution. Strong sequential αN(i,i+1) NOEs from Lys-6 to Arg-13 with the absence of other short-range NOEs among the backbone protons suggest that this region exists in a β-structure. For the majority of these residues this conclusion is supported by the lower field position of the αH resonances; all except Lys-6 and Leu-12 resonate *downfield* of the H_2O signal between 4.5 and 5.5 ppm. In general, low-field-shifted αH resonances (>4.8 ppm) tend to be part of extended, nonhelical regions of proteins, as observed for Thr-5 to Val-15 in the *lac* repressor headpiece (Zuiderweg et al., 1983). Also, all residues within the 6–13 span have $^3J_{NH-\alpha H}$ values greater than 8 Hz. This corresponds to a ϕ torsion angle

of −110° to −140°, which is more consistent with β-structures than helices (Pardi et al., 1984).

No NOEs were seen to the αH of Pro-15 although numerous NOEs were apparent among the remaining ring protons to the αH and 2H of Trp-14. A medium-sized NOE was seen between the NH of Trp-14 and the NH of Arg-16 in both the short (100-ms) and long (200-ms) mixing time NOESYs (Figure 7A), indicating that Trp-14 and Arg-16 may bend toward each other around Pro-15. This is not unusual since a helix starts with the next residue, Arg-16, and prolines are rigid and capable of adopting fixed dihedral angles required for helix initiation (Schulz & Schirmer, 1979).

α-Helix (Residues 16–26). In contrast to residues 6–14, the αH resonances of residues 16–26 all appear *upfield* of H_2O and give backbone NOE patterns typical of helices. These include stretches of repeated NN(i,i+1) and αN(i,i+3) and some αβ(i,i+3) NOEs of medium intensity and weaker αN(i,i+1) NOEs (Figure 9). There are a few irregularities, especially at the ends of the helix. Thus, an αN(i,i+3) for the αH of Ala-26 to the NH of Asn-29 implies that the helix continues on for one more turn, although for the next residue Glu-27, the NN(i,i+1) connection ends. This region may be the beginning of a turn or loop since both Glu-27 and Glu-28 have strong αN(i,i+1) NOEs. In practice, the beginning and end of helices are hard to define from NMR data alone, particularly as the pattern of NOEs for turns or loops is similar to those for helices (Clore & Gronenborn, 1987).

Additionally, some αN(i,i+2) NOEs were observed within the 16–26 segment (Figure 9). This type of NOE is usually diagnostic for the less common 3_{10} helix. Normally, α-helices are distinguished from 3_{10} helices by the presence of αN(i,i+4) NOEs in the former and αN(i,i+2) NOEs in the latter (Wüthrich, 1986). The length of the first helix is 10 or 11 residues, and after one full turn of helix, an αN(i,i+2) NOE between the αH of Leu-19 and NH of Leu-21 is present. The next two (much weaker) αN(i,i+2) NOEs are between residues 21–23 and 24–26. These through-space NOE contacts support a 3_{10} helix within the 16–26 span. However, the strong αβ(i,i+3) NOEs present throughout most of the helix indicate that the entire residue 16–26 stretch is an α-helix. Perhaps the medium-sized αN(i,i+2) NOE between Leu-19 and Leu-21 is the result of some hydrophobic stacking within the helix. The smaller $^3J_{NH-\alpha H}$ data (Figure 9) suggest that no distortions exist for the entire length of the helix since all coupling constants are relatively the same in magnitude. Additional data from hydrogen exchange studies would be most useful. However, at the temperatures used in the present study (45–55 °C), the exchange rates were too fast to permit any measurements to be obtained, and at temperatures below 45 °C the NH peaks were too broad to allow satisfactory NOESY or COSY data to be recorded.

In general, this helical segment (residues 16–26) is amphiphilic with hydrophilic (Arg-16–Glu-17, Asp-20, and Arg-23–Lys-24) and hydrophobic (Val-18–Leu-19, Leu-21–Val-22, and Val-25–Ala-26) surfaces. This amphiphilic character may allow the 16–26 helical segment to pack with the nonpolar residues positioned to the interior and polar residues exposed to the surface near water.

Looped Structure (Residues 27–34). The pattern of NOEs observed in this region is quite variable. Except for Val-33 and Asn-34, strong αN(i,i+1) NOEs are present throughout the segment. The $^3J_{NH-\alpha H}$ for Asn-29 and Arg-31 are >8 Hz, and their αH chemical shifts are *downfield* of the H_2O signal. The pattern of NOEs in this stretch is not indicative of a turn (Wüthrich, 1986), and therefore we propose that this region exists as a looped structure in solution. Similar looped structures are found in the EF hands of certain calcium binding proteins such as calmodulin (Seamon, 1982). The only hydrophobic residue in this region is Val-33. Not surprisingly, as all the residues except for Val-33 are polar and presumably solvent exposed, Val-33 does not display any long-range NOE contacts to any aromatic residues. A strong NOE between the αH of Val-33 and the βCH_3 and the αH of Ala-26 was seen in NOESY data obtained in D_2O solution. This latter NOE may be either a "monomer" or a "dimer" contact, and a final choice between the two possibilities may have to wait until additional data are obtained.

α-Helix (Residues 35–50). This segment of PL8 is characterized as an α-helix since it contains an uninterrupted stretch of NN(i,i+1) NOEs. There are also αN(i,i+3) and αβ(i,i+3) NOEs throughout the sequence, except from Lys-46 to Arg-50. The entire helix is amphiphilic except for the last five residues, which are all hydrophilic. Perhaps the latter αN(i,i+3) and αβ(i,i+3) NOEs are missing because of the increased polarity at the C-terminal end of the helix, causing it to be more solvent exposed.

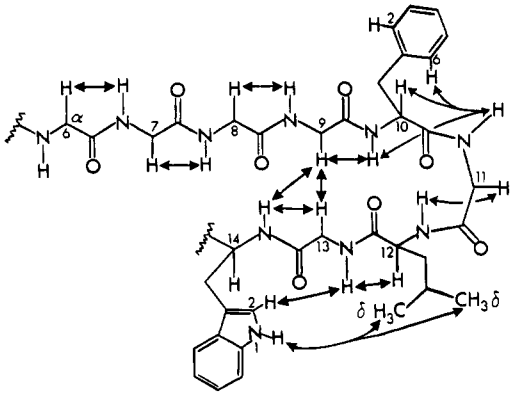
The amphiphilic character of the second helix (residues 35–50) is reflected in its distribution of distinct hydrophobic surface residues: Ile-37 (and possibly Tyr-38), Val-41, Met-42, and Phe-45 (and possibly Lys-46). Although Tyr-38 and Lys-46 are not hydrophobic residues, the aromatic ring of Tyr-38 and long side of Lys-46 allow them to substitute for hydrophobic residues under some circumstances.

Tertiary Structure and Globular Fold in Solution. An essential prerequisite for determining the tertiary structure of a protein is to identify the long-range (tertiary) NOEs as these are needed for determining the polypeptide fold of the protein. These long-range NOEs are summarized in Table II. At the present time, our interpretation of these interactions is limited, due to our inability to distinguish between monomer and dimer interactions.

Several long-range NOEs are found among residues 6–14 which form part of the DNA binding motif of the Arc repressor. These include a strong αH–αH NOE between Gln-9 and Arg-13 (Figure 6B), a medium αN(i,i+5) NOE between Gln-9 and Trp-14 (Figure 6A), and a NN(i,i+1) NOE between Phe-10 and Asn-11 (Figure 6C). In addition, the side-chain protons of Phe-10 exhibit NOEs to the side-chain protons of Leu-8 and Leu-12 (Table II). Two types of structural motifs can accomodate the NOEs seen in the 6–14 region of the protein. One possible structure is a β-strand which folds back upon itself, forming a turn at Asn-11, and another possibility would be two β-strands from different monomers aligned antiparallel with respect to each other. In the first structure the three long-range NOEs discussed above are "monomer NOEs", and in the second structure the NOEs are "dimer NOEs". An unambiguous choice between the two types of interactions will have to wait until additional NMR data are obtained on specifically labeled PL8 dimers. A β-strand, γ-turn, β-strand structure for the residue 6–14 region is shown in Chart V. The structure contains a γ-turn (Milner-White et al., 1988) centered at Asn-11 with the chain reversal stabilized by a hydrogen bond between the NH of Phe-10 and the carbonyl of Leu-12. This structural motif accommodates the observed NOEs and allows the polar and nonpolar residues to occupy opposite faces of the β-strand.

The relative alignment of the first and second helices is constrained by the NOEs observed between the hydrophobic residues on the two helices. The nonpolar and proximal Leu-21, Val-22, and Val-25 residues form a hydrophobic patch

Chart V



on the first amphiphilic α-helix which interacts with Phe-45 on the hydrophobic face of the second amphiphilic α-helix. This could account for the strong NOE between Val-25 and Phe-45 and the additional weaker NOE between Leu-21 and Phe-45 (Table II). However, these contacts can originate within each monomer or across the dimer, and additional NMR measurements on a selectively labeled PL8 dimer will be needed to discriminate between the two types of interactions. The structural constraints of the helix–loop–helix motif require that the two α-helices are antiparallel with respect to each other in the monomer, and for the dimer it is possible that the four helices could be arranged in a four-helix bundle.

Several long-range NOEs are observed between the β-structure (residues 6–14) and the two helices (residues 16–26 and 35–50). Thus, Phe-10 of the β-structure exhibits NOEs to Leu-19 in the first helix as does Trp-14 of the β-structure to Val-18 and Leu-19 on the same helix (Table II). Further, Trp-14 exhibits a set of NOEs to Ile-37, Tyr-38, and Val-41 on the second helix (Table II). Thus, it appears that Trp-14 is a focal point within a hydrophobic pocket and makes contacts with both residues 18 and 19 toward the N-terminus of the first helix (residues 16–26) and residues 37, 38, and 41 toward the N-terminus of the second helix (residues 35–50). These contacts could be either within the PL8 monomer or across the PL8 dimer.

The above discussion has been kept brief since the establishment of the tertiary structure of the PL8 dimer requires unambiguous discrimination between monomer and dimer contacts for the observed long-range NOEs. This discrimination should be possible following NMR studies on selectively deuterated PL8 analogues which are mixed to generate heterodimers that differ in their deuterium substitution positions.

Acknowledgments

We acknowledge extensive discussions with Dr. David Norman on potential γ-turns and helix–helix interactions in PL8 observed in his molecular dynamic calculations using NMR-based distance constraints. We thank Dr. Michael Weiss for helpful discussions and continued encouragement. Prof. Robert Kaptein kindly provided us with a preprint of his paper on NMR studies of Arc and agreed to a back-to-back publication of our respective contributions. The NMR spectrometers were purchased from funds donated by the Robert Woods Johnson, Jr., Trust and the Matheson Trust toward setting up an NMR Center in the Basic Medical Sciences at Columbia University.

References

Bax, A., & Davis, D. G. (1985a) *J. Am. Chem. Soc. 107*, 2820–2821.

Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 65*, 355–360.

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

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*, 521–528.

Clore, G. M., & Gronenborn, A. M. (1987) *Protein Eng. 1*, 275–288.

DiStefano, D. L., & Wand, A. J. (1987) *Biochemistry 26*, 7272–7281.

Dykstra, R. (1987) *J. Magn. Reson. 72*, 162–167.

Dykstra, R., & Wand, A. J. (1987) *J. Magn. Reson. 75*, 158–161.

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

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

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.

Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.

Milner-White, E. J., Ross, B. M., Ismail, R., Belhadj-Mostefa, K., & Poet, R. (1988) *J. Mol. Biol. 204*, 777–782.

Morris, G. A. (1986) *Magn. Reson. Chem. 24*, 371–403.

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

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

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.

Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.

Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.

Schulz, G. E., & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York.

Seamon, K. B. (1982) *Biochemical Structure Determination by NMR* (Bothner-By, A. A., Glickson, J. D., & Sykes, B. D., Eds.) pp 189–221, Marcel Dekker, New York.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson 48*, 286–292.

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

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

Vershon, A. K., Bowie, J. U., Karplus, T. M., & Sauer, R. T. (1986) *Proteins 1*, 302–311.

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

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

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

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