

^1H , ^{15}N and ^{13}C NMR assignments of the 434 repressor fragments 1–63 and 44–63 unfolded in 7 M urea

D. Neri, G. Wider and K. Wüthrich

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

Received 7 April 1992

An *E. coli* overexpression system for the N-terminal domain of the 434 repressor with residues 1–63 (434 repressor(1–63)) was constructed and used to produce this polypeptide with uniform ^{15}N -labeling, and with ^{13}C -labeling of the methyl groups of valine and leucine. Using these protein preparations almost complete sequence-specific resonance assignments were obtained for the urea-unfolded form of the 434 repressor(1–63). In addition, the isotope-labeled tryptic peptide, 44–63, was produced by enzymatic cleavage of the recombinant 434 repressor(1–63), and its NMR spectrum was assigned. Corresponding residues in 434 repressor(1–63) and 434 repressor(44–63) in 7 M urea were found to have nearly identical chemical shifts, and in both species similar deviations from ^1H random coil shifts were found as previously in 434 repressor(1–69). These indicate the presence of residual non-random structure in the polypeptide segment 50–60. The present NMR assignments, which include stereospecific assignments for the diastereotopic methyl groups of Val and Leu, are the basis for detailed studies of this residual structure in the urea-unfolded form of the 434 repressor.

Sequence-specific NMR assignment; Protein folding; Isotope labeling; Urea denaturation

1. INTRODUCTION

Important insights into the laws that govern the three-dimensional folding of proteins [1] will probably result from structural characterization of partially unfolded proteins [2], but up to now reliable structural information on folding intermediates is still scarce [3–10]. The NMR method that has been used for numerous structure determinations of folded proteins [11,12] is, in principle, also applicable for partially unfolded proteins, provided that sequence-specific resonance assignments can be obtained. This paper describes such resonance assignments for two polypeptide fragments of 434 repressor that were unfolded in 7 M urea.

The N-terminal DNA-binding domain of the phage 434 repressor with residues 1–69 (434 repressor(1–69)) is an α -helix protein for which an X-ray crystal structure [13] and a NMR solution structure [14] are available. For the intact 434 repressor an over-expression system is available from which 434 repressor(1–69) can be obtained by proteolytic cleavage [15,16]. For systematic folding studies it was desirable, however, to have an over-expression system which would directly produce the N-terminal domain. This would simplify the protein purification procedure [17] and allow more efficient incorporation of isotope labels. In this paper we describe

the construction of a new expression system for 434 repressor(1–63).

The detailed plan for the work reported here was based on earlier work with 434 repressor(1–69). Because the six C-terminal residues, 64–69, in this protein are conformationally disordered [13,14] we decided to produce an over-expression system for 434 repressor(1–63). Major difficulties in previous studies of the urea-unfolded form of 434 repressor(1–69) arose because the ^1H chemical shift dispersion [11] is greatly reduced when compared to the folded protein [18], showing that labeling with ^{15}N and possibly ^{13}C was needed for more detailed investigations. Finally, the nearly complete ^{15}N and ^1H NMR assignments for the 434 repressor(1–69) in 7 M urea indicated that residual, non-random structure was most probably located in the region of residues 50–60 [18], as evidenced by unusual chemical shifts and slowed amide proton exchange. The isotope-labeled polypeptide 434 repressor(44–63) was therefore prepared by proteolytic cleavage of recombinant 434 repressor(1–63) to investigate in more detail the nature of this residual structure, and to check whether its stability is a local property of the amino acid sequence or depends on longer-range interactions with other parts of the intact polypeptide chain.

2. MATERIALS AND METHODS

2.1. Cloning of the 434 repressor(1–63)

The plasmid pRW190 [15], bearing tetracycline resistance, the p_{lac} promoter and the 434 repressor gene was used to derive an over-expression system for the 434 repressor(1–63). This plasmid contains

Correspondence address: D. Neri, Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland.

a unique *EcoRI* restriction site in the segment of the 434 repressor gene coding for the N-terminal domain, and a unique *HindIII* restriction site in the segment coding for the C-terminal domain. Two oligonucleotides with the sequences AAAGCAAAGAATTCAGCTTGG and AACATTCGAAGCTTAGGTGCCATTGAGCAG were synthesized (Applied Biosystems), the first of which overlapped the *EcoRI* restriction site and the gene fragment coding for residues 8–16. The second oligonucleotide was complementary to the segment of the 434 repressor gene coding for residues 58–68, with three mismatches in order to introduce a stop codon after Thr-63 and a *HindIII* restriction site. The two oligonucleotides were used to amplify, by the polymerase chain reaction (PCR), a 180 bp gene segment, which was then digested with *HindIII* and *EcoRI*. The plasmid pRW190 was also digested with *EcoRI/HindIII* to yield two DNA fragments of 5 kb and 260 bp, which were then treated with calf intestinal alkaline phosphatase. The 5 kb DNA fragment was purified by low-melting agarose gel electrophoresis and ligated to the PCR-derived 180 bp fragment to yield the plasmid pDN01. This plasmid, which bears the tetracycline resistance, the *p_{lac}* promoter and the 434 repressor(1–63) gene, was finally transfected to NM522 *E. coli* competent cells (Stratagene).

2.2. Production of ¹⁵N-labeled 434 repressor(1–63)

The NM522/pDN01 over-expression system was grown at 35°C in a minimal medium containing 1 g/l of ¹⁵N-labeled ammonium sulfate (ISOTEC), 1 mg/l biotin, 5 mg/l thiamine, 10 g/l glucose, 0.5 g/l sodium acetate, succinic acid, citrate, 10 g/l K₂HPO₄, 5 g/l KH₂PO₄, 15 mg/l CaCl₂, 300 mg/l MgCl₂, 0.5 g/l NaCl, 5 mg/l Fe₂SO₄ and 8 mg/l tetracycline. When the *A₆₀₀* of the cell suspension reached 1.6 the over-expression system was induced with 0.1 mM isopropyl-β-D-l-thiogalactopyranoside (IPTG) and growth was continued for 3 h. Cells were then harvested by centrifugation for 10 min at 5,000 rpm, resuspended in lysis buffer (50 mM citrate, pH 3.4, 10% glycerol, 50 mM KCl, 1 mM EDTA, 1.4 mM 2-mercaptoethanol, PMSF as a protease inhibitor) and lysed by sonication. The resulting suspension was centrifuged for 30 min at 17,000 g. The cell debris formed a solid pellet which was discarded, and the clear supernatant containing the 434 repressor(1–63) was exchanged against buffer A (50 mM acetate, pH 4.6, 1 mM EDTA, 1.4 mM 2-mercaptoethanol, 30 mM NaCl) with an Amicon ultrafiltration apparatus (using a membrane cut-off of molecular weight 5,000). The resulting solution was centrifuged to remove material that had precipitated during ultrafiltration (this material did not contain 434 repressor(1–63) as judged by gel electrophoresis), applied to a Mono S FPLC column (Pharmacia), and eluted with a linear gradient of buffer A + 1 M NaCl. The 434 repressor(1–63) eluted as a single, well-separated peak. The collected fractions contained >95% pure protein, as judged by HPLC and by gel electrophoresis. This material was concentrated and purified by reverse-phase HPLC (starting buffer, water + 0.1% TFA; eluent, acetonitrile + 0.1% TFA). This second purification step removed a protease contaminant that was active at pH's higher than 4.5. The eluted fractions containing the 434 repressor(1–63) were lyophilized before preparation of the NMR samples. The final yield was 11 mg of ¹⁵N-labeled 434 repressor(1–63) per liter of minimal medium.

2.3. Production of ([4-¹³C]Val,[5-¹³C]Leu)-434 repressor(1–63)

This protein sample was prepared by growing the NM522/pDN01 *E. coli* over-expression system at 35°C in the same minimal medium as described in the preceding section except that unlabeled ammonium sulfate was used. To this medium the following amino acids were added: 0.4 g/l of Ala, Glu, Gln, Arg and Gly; 0.7 g/l of Met; 0.25 g/l of Asp; 0.1 g/l of His, Ile, Lys, Pro, Thr, Tyr and Asn; 1.6 g/l of Ser; 0.05 g/l of Phe and Trp. 20 min before induction 0.05 g/l each of [4-¹³C]-L-Val and [5-¹³C]-L-Leu were added. The cells were then grown and the protein sample isolated as described in the preceding section. The final yield was 12 mg of labeled 434 repressor(1–63) per liter of minimal medium.

2.4. Production of isotope-labeled 434 repressor(44–63)

434 repressor(44–63) labeled with ¹⁵N or with [4-¹³C]Val and [5-

¹³C]Leu was obtained by tryptic cleavage from the protein preparations described in the preceding two sections. For both preparations 10 mg of 434 repressor(1–63) were dissolved in water containing 4 M urea at pH 7.0, and digested for 7 h at room temperature with trypsin (trypsin/protein ratio 1:200). The peptide, 44–63, was then purified by reverse-phase chromatography on a Pharmacia FPLC chromatograph, with a PepRPC HR 5/5 column (starting solvent, water + 0.1% TFA; eluent, acetonitrile:isopropanol:methanol 2:1:1 + 0.1% TFA). The eluted fractions containing the 434 repressor(44–63) were lyophilized for the preparation of the NMR samples.

2.5. NMR measurements

For the NMR measurements the lyophilized peptides were dissolved either in a mixed solvent of 85% H₂O/15% D₂O or in D₂O. Both solvents contained 20 mM NaClO₄ and 7 M urea (Sigma) at pH 4.8 or pD 4.8 (uncorrected), respectively. The concentration of the ¹⁵N-

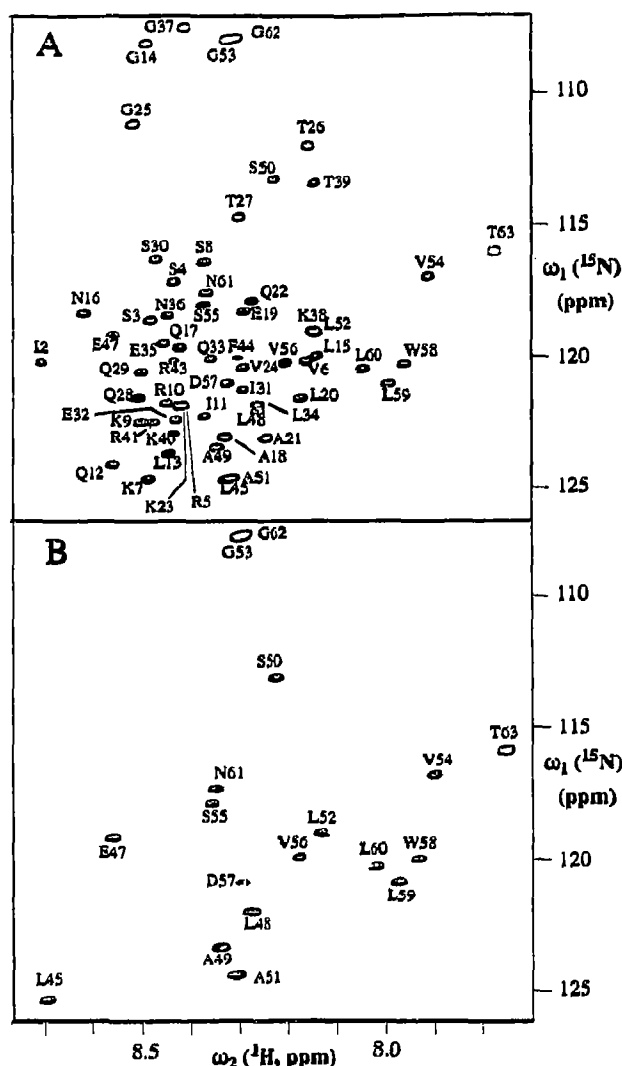


Fig. 1. (A) Spectral region containing the backbone ¹⁵N-¹H cross-peaks in a 2D TOCSY-related [¹⁵N,¹H] COSY spectrum of urea-unfolded uniformly ¹⁵N-labeled 434 repressor(1–63) (3 mM solution in H₂O containing 7 M urea and 20 mM sodium perchlorate, pH 4.8, T = 18°C; proton frequency 600 MHz, mixing time = 100 ms). The cross-peak assignments are given by the one-letter amino acid code and the sequence position. (B) Same as (A) for uniformly ¹⁵N-labeled 434 repressor(44–63) (2 mM solution in H₂O containing 7 M urea and 20 mM sodium perchlorate, pH 4.8, T = 18°C, proton frequency 500 MHz, mixing time = 100 ms).

labeled 434 repressor(1–63) in H_2O was 3 mM, the one of the ^{15}N -labeled 434 repressor(44–63) was 2 mM. For measurements with urea-unfolded ($[4-^{13}C]Val, [5-^{13}C]Leu$)-434 repressor(1–63) and ($[4-^{13}C]Val, [5-^{13}C]Leu$)-434 repressor(44–63) in D_2O solution 6 and 3 mM samples were used, respectively.

All NMR spectra were recorded on a Bruker AMX500 or AMX600 spectrometer at 18°C. In H_2O a TOCSY-relayed $[^{15}N, ^1H]$ COSY spectrum [19] with $\tau_{mix} = 100$ ms was measured at 600 MHz for the 434 repressor(1–63) (180 t_1 values, 2,048 t_2 values, $t_{1max} = 65$ ms, $t_{2max} = 131$ ms) and at 500 MHz for the 434 repressor(44–63) (200 t_1 values, 2,048 t_2 values; $t_{1max} = 84$ ms, $t_{2max} = 155$ ms). $[^{13}C, ^1H]$ COSY spectra in D_2O were recorded with the pulse sequence of Otting and Wüthrich [19] at 600 MHz for both the selectively ^{13}C -labeled 434 repressor(1–63) (420 t_1 values, 4,096 t_2 values; $t_{1max} = 151$ ms, $t_{2max} = 254$ ms) and the selectively ^{13}C -labeled 434 repressor(44–63) (320 t_1 values, 4,096 t_2 values; $t_{1max} = 115$ ms, $t_{2max} = 254$ ms). An exchange-relayed $[^{13}C, ^1H]$ COSY spectrum [20] was used to transfer the stereospecific assignments of the diastereotopic methyl groups of Val and Leu from the native form to the urea-unfolded form of the 434 repressor(1–63). It was recorded at 500 MHz with a 3 mM ($[4-^{13}C]Val, [5-^{13}C]Leu$)-la-

beled sample in a D_2O solution at pD 4.8, containing 20 mM NaClO₄ and 4.2 M deuterated urea (Sigma) ($\tau_{mix} = 100$ ms, 200 t_1 values, 2,048 t_2 values; $t_{1max} = 84$ ms, $t_{2max} = 115$ ms).

3. RESULTS AND DISCUSSION

An aqueous solution of 434 repressor(1–63) displays the same phenomena upon addition of urea as was previously described for 434 repressor(1–69) [18]. At pH 4.8 and 18°C the protein is in the native folded form in the absence of urea. Upon addition of 4.3 M urea, the native form and a urea-unfolded form are equally populated. Chemical exchange with a lifetime of approximately one second can be observed between the well-separated resonances of the two forms. At 7 M urea the urea-unfolded form is the only species present in solution. With the exceptions that the resonances of residues 64–69 are absent and that the C-terminal residue Thr-63 has different chemical shifts the 1H NMR spectrum of the native form of 434 repressor(1–63) is very similar to that of native 434 repressor(1–69) [14]. It was completely assigned [21] using standard procedures [11]. The amide 1H and ^{15}N resonance assignments were then transferred to the urea-unfolded form by means of an exchange-relayed $[^{15}N, ^1H]$ COSY difference spectrum [20], and assignments for the non-labile protons were obtained by means of a TOCSY-relayed $[^{15}N, ^1H]$ COSY spectrum [19] (Figs. 1A and 2A). In the fingerprint region [11] of the $[^{15}N, ^1H]$ COSY spectrum of the unfolded protein in 7 M urea there are only three residue pairs with overlapped cross-peaks, i.e. Arg-5 and Lys-23, Lys-38 and Leu-52, and Leu-45 and Ala-51 (Fig. 1A). Except for Thr-63 all ^{15}N chemical shifts are within 0.3 ppm and all amide proton shifts within 0.03 ppm of those in 434 repressor(1–69) [14]. Fig. 2A illustrates that the dispersion of the ^{15}N chemical shifts in the TOCSY-relayed $[^{15}N, ^1H]$ COSY spectrum provided sufficient peak separation for complete resonance assignments, although the 1H shifts for identical residue types are nearly identical (Table I).

The 434 repressor(44–63) was studied only in the unfolded form in 7 M urea, since no second form could be observed in $[^{15}N, ^1H]$ COSY spectra recorded at lower urea concentrations. (Although precipitation was observed at lower urea concentrations the unchanged NMR spectrum of the species present at 7 M urea was observed even at 1.5 M urea.) Figs. 1B and 2B show that with the exception of the N-terminal dipeptide of residues 44 and 45 in the shorter polypeptide identical NMR spectra were observed for corresponding residues in the fragments 1–63 and 44–63. The resonance assignments for 434 repressor(44–63), given in Figs. 1B and 2B and in Table I, were in most instances confirmed by the conventional sequential assignment approach [11]. The results of Figs. 1 and 2 and Table I show for both polypeptide fragments a crowding of 1H chemical shift deviations from the random coil values [11,22] in the

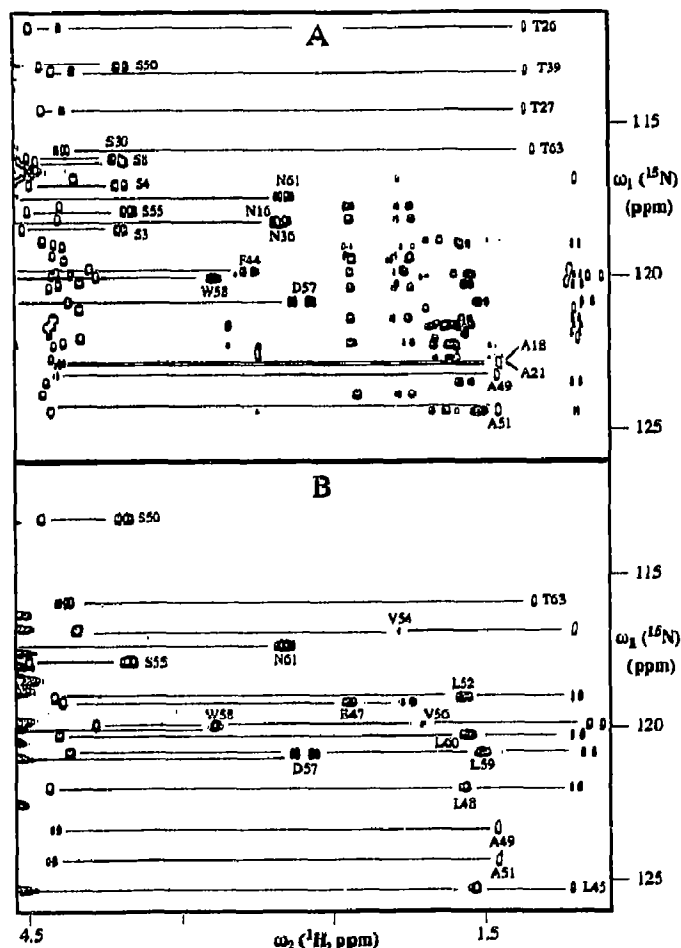


Fig. 2. (A) Spectral region showing the cross-peaks with the nonlabile protons in the same TOCSY-relayed $[^{15}N, ^1H]$ COSY spectrum as in Fig. 1A. The cross peaks of the spin systems of Ala, Thr, Asn, Asp, Phe, Trp and Ser have been connected with horizontal lines and identified with the one-letter amino acid code and the sequence position. (An expanded plot of the crowded region with the cross-peaks of the other side chains was previously presented for 434 repressor(1–69), with all assignments indicated [18]). (B) Same as (A) for 434 repressor(44–63). Assignments are given for all residues except Gly, Pro and the N-terminal Phe.

Table 1

 ^1H , ^{15}N and ^{13}C chemical shifts of 434 repressor(1-63) and 434 repressor(44-63) in the urea-unfolded form

Residue	Chemical shift (ppm)*				
	^{15}N	NH	αH	Other protons ^{a,h}	$^{13}\text{C}^{\text{b,c}}$
Ile-2	120.1	<u>8.71</u>	4.21	βH , 1.89; γH , 1.49, 1.19; γCH_3 , 0.93; δCH_3 , 0.89	
Ser-3	118.6	8.48	4.53	βH , 3.91, 3.86	
Ser-4	117.1	<u>8.44</u>	4.48	βH , 3.92, 3.86	
Arg-5	121.7	8.42	4.34	βH , <u>3.18</u> , <u>3.18</u> ; ϵNH , 7.18	
Val-6	119.9	8.14	4.09	βH , 2.08; γCH_3 , 0.94, 0.94	γCH_3 , 21.4, 20.9
Lys-7	124.6	8.49	4.34	βH , 1.84, 1.74; γH , 1.48; δH , 1.68; ϵH , 2.99, 2.99	
Ser-8	116.4	8.37	4.43	βH , 3.87	
Lys-9	122.4	8.50	4.32	βH , 1.82, 1.72; γH , 1.45; δH , 1.68; ϵH , 2.98, 2.98	
Arg-10	121.7	8.45	4.33	βH , 1.87, 1.75; γH , 1.63, <u>1.58</u> ; δH , <u>3.18</u> , <u>3.18</u> ; ϵNH , 7.18	
Ile-11	122.2	8.37	4.15	βH , 1.80; γH , 1.47, 1.17; γCH_3 , 0.89; δCH_3 , 0.86	
Gln-12	124.0	8.56	4.40	βH , 2.07, 1.97; γH , 2.33, 2.33	
Leu-13	123.6	8.45	4.38	βH and/or γH , 1.65, 1.57; δCH_3 , 0.92, 0.86	δCH_3 , 25.2, 23.4
Gly-14	108.1	8.49	4.00, 3.91		
Leu-15	120.1	8.16	4.31	βH and/or γH , 1.58; δCH_3 , 0.90, 0.86	δCH_3 , 25.1, 23.6
Asn-16	118.3	8.62	4.70	βH ; 2.87, 2.79	
Gln-17	119.6	8.42	4.25	βH , 2.12, 2.04; γH , 2.36, 2.36	
Ala-18	123.0	8.33	4.26	βCH_3 , 1.40	
Glu-19	118.3	8.29	4.29	βH , 2.07, 1.96; γH , 2.37, 2.37	
Leu-20	121.5	8.17	4.33	βH and/or γH , 1.65, 1.56; δCH_3 , 0.92, 0.87	δCH_3 , 25.1, 23.6
Ala-21	123.0	8.24	4.28	βCH_3 , 1.39	
Gln-22	117.9	8.27	4.27	βH , 2.07, 1.98; γH , 2.39, 2.34	
Lys-23	121.8	8.43	4.34	γH , 1.46, 1.39; ϵH , 2.98	
Val-24	120.3	8.29	4.15	βH , 2.07; γCH_3 , 0.96, 0.96	γCH_3 , 21.4, 20.7
Gly-25	111.2	8.52	4.06, 4.02		
Thr-26	111.9	8.16	<u>4.48</u>	βH , 4.27; γCH_3 , 1.22	
Thr-27	114.6	8.30	4.40	βH , 4.25; γCH_3 , 1.22	
Gln-28	121.5	8.51	4.33	βH , 2.10, 1.99; γH , 2.37, 2.37	
Gln-29	120.5	8.50	4.35	βH , 2.09, 1.98; γH , 2.37, 2.37	
Ser-30	116.3	8.47	4.49	βH , 3.94, 3.86	
Ile-31	121.2	8.29	4.16	βH , 1.87; γH , 1.47, 1.18; γCH_3 , 0.91; δCH_3 , 0.86	
Glu-32	122.4	8.43	4.26	βH , 2.02, 1.96; γH , 2.36, 2.36	
Gln-33	120.0	8.36	4.30	βH , 2.10, 2.01; γH , 2.37, 2.37	
Leu-34	121.6	8.26	4.34	βH and/or γH , 1.65, 1.56; δCH_3 , 0.91, 0.85	δCH_3 , 25.1, 23.5
Glu-35	119.4	8.46	4.32	βH , 2.08, 1.97; γH , 2.39, 2.34	
Asn-36	118.4	8.45	4.69	βH , 2.83	
Gly-37	107.5	8.41	3.97, 3.97		
Lys-38	119.1	8.14	4.39	βH , 1.87, 1.78; γH , 1.47, 1.39; δH , 1.68, 1.68; ϵH , 2.98, 2.98	
Thr-39	113.4	8.14	4.33	βH , 4.19; γCH_3 , 1.22	
Lys-40	122.9	8.44	4.34	βH , 1.80, 1.71; γH , 1.46; δH , 1.67; ϵH , 2.97, 2.97	
Arg-41	122.4	8.47	<u>4.58</u>	δH , <u>3.19</u> , <u>3.19</u> ; ϵNH , 7.20	
Pro-42					
Arg-43	120.1	8.43	4.33	βH , 1.85; δH , <u>3.13</u> , <u>3.13</u> ; ϵNH , 7.15	
Phe-44	120.0	8.30	4.67	βH , 3.29, <u>3.13</u> ; ring, 7.25, 7.29, 7.33	
Leu-45	124.6	8.31	(4.26) <u>4.59</u>	(βH , <u>3.16</u> ; ring, 7.24, 7.31, 7.35) βH , and/or γH , 1.55; δCH_3 , 0.91, 0.88	δCH_3 , 25.4, 23.75
	(125.4)	(<u>8.72</u>)	(<u>4.65</u>)	(βH , and/or γH , 1.55) (δCH_3 , 0.89, 0.89)	

Table 1 (continued)

Residue	Chemical shift (ppm)*				
	¹⁵ N	NH	αH	Other protons ^{a,b}	¹³ C ^{b,c}
Pro-46			(4.37)	γH, 1.99; δH, 3.67, 3.58 (βH, 2.33, 1.91; γH, 2.03) (δCH ₃ , 3.74, 3.62)	
Glu-47	119.1	8.56	4.25	βH, 2.04, 1.95; γH, 2.37, 2.37	
Leu-48	122.1	8.26	4.36	βH and/or γH: 1.60; δCH ₃ , 0.93, 0.87	δCH ₃ , 25.1, 23.6
Ala-49	123.4	8.35	4.31	βCH ₃ , 1.41	
Ser-50	113.2	8.22	4.40	βH, 3.91, 3.84	
Ala-51	124.5	8.33	4.35	βCH ₃ , 1.40	
Leu-52	119.2	8.15	4.32	βH and/or γH, 1.64, 1.59; δCH ₃ , 0.92, 0.87	δCH ₃ , 25.1, 23.5
Gly-53	107.8	8.31	3.95, 3.95		
Val-54	116.9	<u>7.91</u>	4.18	βH, 2.06; γCH ₃ , 0.89, 0.89	γCH ₃ , 21.4, 20.4
Ser-55	118.0	8.37	4.49	βH, 3.87, 3.81	
Val-56	120.2	8.20	<u>4.05</u>	βH, 1.91; γCH ₃ , 0.81, 0.72	γCH ₃ , 21.2, 20.4
Asp-57	121.0	8.32	<u>4.59</u>	βH, 2.74, <u>2.61</u>	
Trp-58	120.2	7.95	<u>4.59</u>	βH, 3.30, 3.24; 2H, 7.24; 4H, 7.59; 5H, 7.14; 6H, 7.23; 7H, 7.47; NH, 10.13	
	(120.1)	(7.95)	(4.59)	(βH, 3.28; 2H, 7.24) (4H, 7.59; 5H, 7.14; 6H, 7.23) (7H, 7.47; NH, 10.13)	
Leu-59	121.0	<u>7.99</u>	<u>4.22</u>	βH: <u>1.53</u> , 1.47, γH: <u>1.37</u> ; δCH ₃ , 0.85, 0.78	δCH ₃ , 25.1, 23.7
Leu-60	120.4	<u>8.04</u>	4.28	βH and/or γH, 1.62, 1.57; δCH ₃ , 0.92, 0.85	δCH ₃ , 25.1, 23.6
Asn-61	117.5 (117.5)	<u>8.36</u> (<u>8.36</u>)	4.73 (4.73)	βH, 2.84, 2.77 (βH, 2.84, 2.77) (δNH, 6.96, 7.92)	
Gly-62	107.9	8.33	3.97, 3.97		
Thr-63	116.0	7.77	4.29	βH, 4.23; γCH ₃ , 1.15	

*The values listed were measured for 434 repressor(1-63). A second entry, in parentheses, is given for those among the residues 44-63 for which any of the ¹H chemical shifts in 434 repressor(44-63) differs from those in 434 repressor(1-63) by more than 0.03 ppm, or for which additional resonances were assigned in 434 repressor(44-63). Chemical shifts relative to internal sodium 3-trimethylsilylpropionate, sodium salt (TSP), were measured in aqueous solution containing 7 M urea and 20 mM sodium perchlorate at pH 4.8 and T = 18°C. The proton chemical shifts are underlined if they deviate from the random coil values [22] by more than 0.1 ppm for non-labile protons or by more than 0.3 ppm for amide protons.

^aSide chain proton chemical shifts were measured in a TOCSY-relayed [¹⁵N, ¹H]COSY spectrum (Fig. 2) and individually assigned within their spin systems on the basis of the close coincidence with the random coil chemical shift values [22].

^bThe methyl groups of Val and Leu were stereospecifically assigned (see text) and are listed in the order γ¹, γ² and δ¹, δ², respectively.

^cThis column lists the ¹³C chemical shifts of the methyl groups of Val and Leu measured in the ([4-¹³C]Val,[5-¹³C]Leu)-labeled polypeptides.

region of residues 54-60. This coincides with previous observations in 434 repressor(1-69), where the presence of a residual folded structure in this region was further indicated by slowed amide proton exchange for several residues in the segment 51-60 [18]. (Protein dimerization in the urea-unfolded form was excluded as the cause of these effects by studies which showed that the ratio between native and unfolded form of the protein is independent of the total protein concentration [21].) The present observation that these 'non-random' chemical shifts are virtually identical in the three polypeptides 1-69, 1-63 and 44-63 now indicates that they reflect a localized residual structure that is preserved in the

shortest of the three fragments, and that this structure might be further characterized by a detailed investigation of 434 repressor(44-63).

A successful attempt to determine the residual 434 repressor structure in 7 M urea will depend on the ability to identify and assign a maximum possible number of conformational constraints. Since the aforementioned evidence indicates that the polypeptide segment of prime interest is Leu-48-Ala-Ser-Ala-Leu-Gly-Val-Ser-Val-Asp-Trp-Leu-Leu-60, NOEs with methyl groups of Val and Leu might represent an important part of the obtainable constraints. Because of the near-degeneracy of the ¹H chemical shifts of these methyls

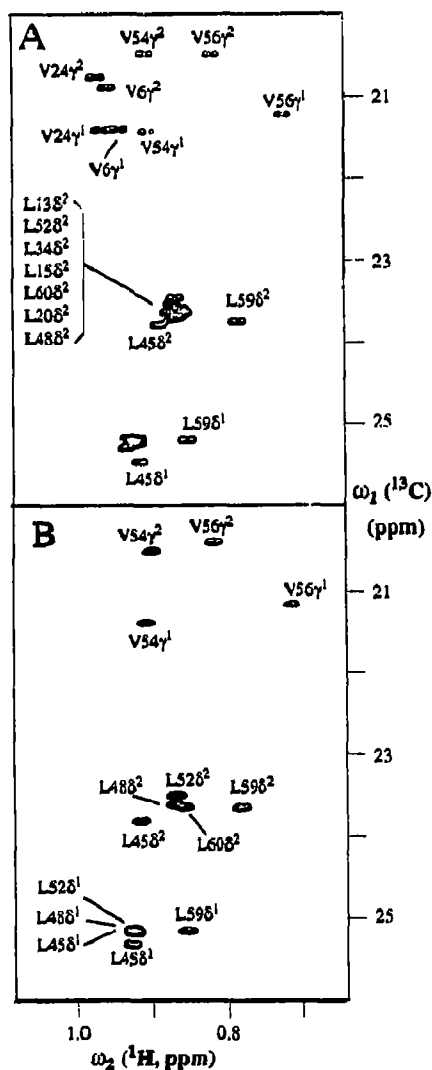


Fig. 3. (A) Spectral region containing the ^{13}C - ^1H cross-peaks of the Val and Leu methyl groups in a 2D [^{13}C , ^1H] COSY spectrum of urea-unfolded ([4- ^{13}C]Val,[5- ^{13}C]Leu)-434 repressor(1-63) (6 mM protein solution in D_2O containing 7 M urea and 20 mM sodium perchlorate, pD 4.8, $T = 18^\circ\text{C}$, proton frequency 600 MHz). Stereospecific assignments (see text) are given using the standard IUB-IUPAC nomenclature (γ^1 and δ^1 are *pro-R*, γ^2 and δ^2 are *pro-S*). (B) Same as (A), for ([4- ^{13}C]Val,[5- ^{13}C]Leu)-434 repressor(44-63) (3 mM solution in D_2O containing 7 M urea and 20 mM sodium perchlorate, pD 4.8, $T = 18^\circ\text{C}$, proton frequency 600 MHz).

(Table 1) we therefore decided to label the 434 repressor(1-63) with [4- ^{13}C]Val and [5- ^{13}C]Leu in order to spread the methyl resonances along the ^{13}C chemical shift axis. Furthermore, to obtain more meaningful NOE distance constraints [11] it was attractive to extend earlier stereospecific assignments of the isopropyl methyls in 434 repressor(1-69) [17] to the fragments 1-63 and 44-63.

The L-valine and L-leucine used for the biosynthetic labeling of 434 repressor(1-63) contain the ^{13}C label randomly distributed between the two diastereotopic

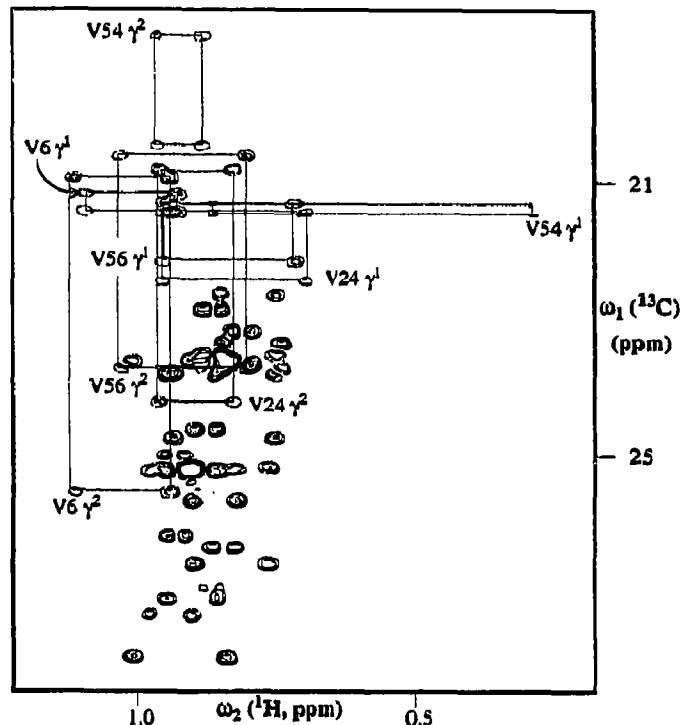


Fig. 4. Spectral region containing the cross-peaks of the Val and Leu methyl groups in a 2D exchange-relayed [^{13}C , ^1H] COSY spectrum of ([4- ^{13}C]Val,[5- ^{13}C]Leu)-434 repressor(1-63) (3 mM protein solution in D_2O containing 4.2 M urea and 20 mM sodium perchlorate, pD 4.8, $T = 18^\circ\text{C}$, proton frequency 500 MHz, mixing time = 100 ms). The rectangular patterns of cross peaks corresponding to the direct ^{13}C - ^1H correlation peaks of the native form and the urea-unfolded form of the protein and the two chemical-exchange peaks [20] are drawn for the eight methyl groups of Val. The stereospecific assignments (see text) are indicated next to the direct correlation peak corresponding to the native form of the protein, using standard IUB-IUPAC nomenclature (γ^1 and δ^1 are *pro-R*, γ^2 and δ^2 are *pro-S*). Positive and negative cross-peaks are both drawn with solid contour lines.

methyl positions, and in each Val and Leu molecule only one of the two methyl groups is labeled with ^{13}C . As a result all isopropyl methyls in the labeled protein give ^{13}C NMR signals of comparable intensity, for example in [^{13}C , ^1H] COSY. Fig. 3A shows that because the ^{13}C NMR lines are better dispersed than the corresponding ^1H resonances, the Val cross-peaks are well separated in the [^{13}C , ^1H] plane of 434 repressor(1-63). For example, the two cross peaks of the methyl groups of Val-54, for which the ^1H resonances are almost degenerate, are well separated in the [^{13}C , ^1H] plane. The Leu methyl cross-peaks are still only poorly resolved with the exceptions of Leu-59 and Leu-45. In 434 repressor(44-63), however, nearly complete separation of all methyl resonances of Val and Leu is achieved with this labeling strategy (Fig. 3B).

The methyl region of the [^{13}C , ^1H] COSY spectrum of native 434 repressor(1-63) is practically identical to that of native 434 repressor(1-69) [21]. The stereospecific assignments of the diastereotopic methyl groups of Val

and Leu obtained for 434 repressor(1-69) by biosynthetically directed fractional ^{13}C labeling [17] were therefore carried over to the native 434 repressor(1-63). An exchange-relayed [^{13}C , ^1H] COSY difference spectrum [20] was then recorded in a solution of 434 repressor(1-63) at 18°C and 4.2 M urea (Fig. 4). For each methyl group of Val or Leu this spectrum contains the direct correlation peaks for the folded and the urea-unfolded form of the protein, as well as two exchange cross-peaks, giving rise to typical rectangular patterns [20]. From this experiment the stereospecific assignments for urea-unfolded 434 repressor(1-63) given in Fig. 3A and Table I were established. Based on the near-identity of the methyl ^1H and ^{13}C shifts in the two fragments (Table I), the stereospecific assignments were also carried over to 434 repressor(44-63) (Fig. 3B), a process which was further based on a variety of data obtained from ^1H spin system identification and sequential ^1H NMR assignments in the fragment 44-63.

In conclusion, the experiments reported in this paper present evidence that in 7 M aqueous urea solution the region of residues 48-60 in all three polypeptide fragments, 1-69, 1-63 and 44-63, of 434 repressor retains an identical residual non-random structure. Sequence-specific ^1H NMR assignments and stereospecific assignments for the methyl groups of Val and Leu, now available for all three polypeptide fragments, provide a basis for further investigations of this residual structure in the urea-unfolded protein.

Acknowledgments: We thank Dr. K. Berndt for advise on the preparation of 434 repressor(44-63), and Mr. R. Marani for the careful processing of the manuscript. Financial support by the Schweizerischer Nationalfonds (project 31.25174.88) is gratefully acknowledged.

REFERENCES

- [1] Richards, F.M. (1991) *Sci. Am.* 264, 34-41.
- [2] Dobson, C.M. (1991) *Curr. Opin. Struct. Biol.* 1, 22-27.
- [3] Oas, T.G. and Kim, P.S. (1988) *Nature* 336, 42-48.
- [4] Roder, E., Elöve, G.A. and Englander, S.W. (1988) *Nature* 335, 700-704.
- [5] Matouschek, A., Kellis, J.T., Serrano, L. and Fersht, A. (1990) *Nature* 346, 440-445.
- [6] Miranker, A., Radford, S.E., Karplus, M. and Dobson, C.M. (1991) *Nature* 349, 633-636.
- [7] Shortle, D. and Meeker, A.K. (1989) *Biochemistry* 28, 936-944.
- [8] Hughson, F.M., Wright, P.E. and Baldwin, R.L. (1990) *Science* 243, 1544-1548.
- [9] Dyson, H.J., Rance, M., Houghton, R.A., Wright, P.E. and Lerner, R.A. (1988) *J. Mol. Biol.* 201, 201-217.
- [10] Mierlo, C.P.M., Darby, N.J., Neuhaus, D. and Creighton, T.E. (1991) *J. Mol. Biol.* 222, 353-371.
- [11] Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- [12] Wüthrich, K. (1989) *Science* 243, 45-50.
- [13] Mondragon, A., Subbiah, S., Almo, S.C., Drott, M. and Harrison, S.C. (1989) *J. Mol. Biol.* 205, 189-201.
- [14] Neri, D., Billeter, M. and Wüthrich, K. (1992) *J. Mol. Biol.* 223, 743-767.
- [15] Wharton, R.P. (1985) *The Binding Specificity Determinants of 434 Repressor*, Ph.D. Thesis, Harvard University, Cambridge, MA.
- [16] Anderson, J., Ptashne, M. and Harrison, S.C. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1307-1311.
- [17] Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry* 28, 7510-7516.
- [18] Neri, D., Wider, G. and Wüthrich, K. (1992) *Proc. Natl. Acad. Sci. USA* (in press).
- [19] Otting, G. and Wüthrich, K. (1988) *J. Magn. Reson.* 76, 569-574.
- [20] Wider, G., Neri, D. and Wüthrich, K. (1991) *J. Biomol. NMR* 1, 93-98.
- [21] Neri, D. (1992) *Use of NMR Spectroscopy for the Study of Small Proteins in the Native and the Urea-unfolded Form: Application to the N-terminal Domain of the Phage 434 Repressor*, Ph.D. Thesis No. 9687, ETH Zürich.
- [22] Bundi, A. and Wüthrich, K. (1979) *Biopolymers* 18, 285-298.