Complete assignment of the ^{1}H NMR spectrum and secondary structure of the DNA binding domain of GAL4

Paresh L. Gadhavi, Andrew R.C. Raine, Peter R. Alefounder and Ernest D. Laue

Department of Biochemistry, Tennis Court Road, Cambridge, CB2 1QW, UK

Received 3 October 1990

Complete ^{1}H NMR resonance assignments are presented for the cysteine rich region of the DNA binding domain of the yeast transcriptional activator GAL4. The protein contains short helical regions between Asp-12 and Leu-19 and between Lys-30 and Trp-36. It is clearly distinct from the C_2H_2 class of zinc finger protein typified by the *Xenopus laevis* transcription factor (TF)IIIA. We also find that the first SP(X)(X) sequence, a recently proposed DNA binding motif (residues 41 to 44), appears to be tightly packed against the metal binding domain.

GAL4; 2D ^{1}H NMR; Sequence-specific resonance assignment; Secondary structure; DNA binding; Zinc finger

1. INTRODUCTION

During the last few years it has emerged that eukaryotic transcription factors recognise and bind to DNA via a number of different structural motifs. One such motif, the metal binding or zinc finger, consists of a small protein domain stabilised by the interaction with zinc ions [1,2]. The yeast gene regulatory protein GAL4 activates transcription of the GAL1 and GAL10 genes by binding to 4 related 17 bp DNA sequences within the upstream activating sequence for galactose (UASg) [3]. It has been proposed [1,4] that 4 of the 6 cysteines within the DNA binding domain of the protein (GAL4, 1–147) form a structure analogous to the zinc fingers of TFIIIA [5] where each zinc ion has been shown by EXAFS to be co-ordinated by two cysteine and two histidine (C_2H_2) residues [6]. Following EXAFS studies we have recently suggested that the GAL4 protein contains a two metal ion cluster where each metal ion is coordinated by 4 cysteine residues [7]. Similar conclusions have resulted from ^{113}Cd NMR studies by Pan and Coleman [8].

In order to further study the structure of the zinc containing DNA binding domain by ^{1}H NMR spectroscopy we required the simplest possible protein domain containing the cysteine-rich region. A 43 amino acid fragment (GAL4 (7–49)) was purified from *E. coli* and found to contain two zinc ions per protein monomer by both spectrophotometric titration with cadmium and electrospray ionisation mass spectrometry; it also bound to a concensus 17 bp DNA binding site, albeit much more weakly than the intact DNA binding domain GAL4 (1–147) [9]. In this paper we give a preliminary account of our NMR studies and report the sequence specific ^{1}H NMR resonance assignments and identification of secondary structure. Even at this stage, prior to the calculation of the 3-D structure, it is clear that the fragment has a different structure to that determined for the C_2H_2 glass of zinc finger protein typified by TFIIIA [10–12].

Correspondence address: E.D. Lane, Department of Biochemistry, Tennis Court Road, Cambridge, CB2 1QW, UK

2. MATERIALS AND METHODS

2.1. *Protein purification*

A GAL4 DNA binding domain, GAL4 (7–49), was expressed in *E. coli* under the control of the *tac* promoter and purified by chromatography on Fast-S and FPLC Mono-S columns; as judged by reversed phase HPLC and gel electrophoresis the protein was >95% pure. The amino acid composition, the N- and C-terminal amino acid sequences and the molecular mass as determined by electrospray ionisation mass spectrometry, corresponded to that predicted [9]. Samples for NMR spectroscopy were concentrated to 2–3 mM and adjusted to pH 5.4 in the presence of 10% ^2H_2O and 5 mM zinc chloride. Finally they were sealed under nitrogen.

2.2. *NMR Spectroscopy*

2D NMR spectra were recorded at 283, 288, or 293 K on a Bruker AMX 600 spectrometer. Data processing was carried out using either the manufacturer's software or the FTNMR package kindly provided by Dr Dennis Hare.

All spectra were acquired in the phase sensitive mode with quadrature detection in the t_1 dimension using time-proportional phase incrementation (TPPI) [13,14]. In all cases the carrier was placed on the ^1H_2O resonance which was suppressed by presaturation during the relaxation delay and additionally in the mixing time in NOESY experiments. The receiver reference phase [15] and the delay between the opening of the receiver gate and acquisition of the first data point [16] was optimised to obtain a flat baseline. During data processing, baseline correction in ω_2 was carried out using a polynomial of order 3.

Double quantum filtered COSY (2QF-COSY) spectra were acquired using the standard phase cycle [17]; spectra were recorded using either 16 or 32 scans for each t_1 point and the data size was 4 K (t_2) × 1 K (t_1) points. A double quantum (2QS) spectrum was recorded with a mixing time of 32 ms [18,19] and NOESY spectra were

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

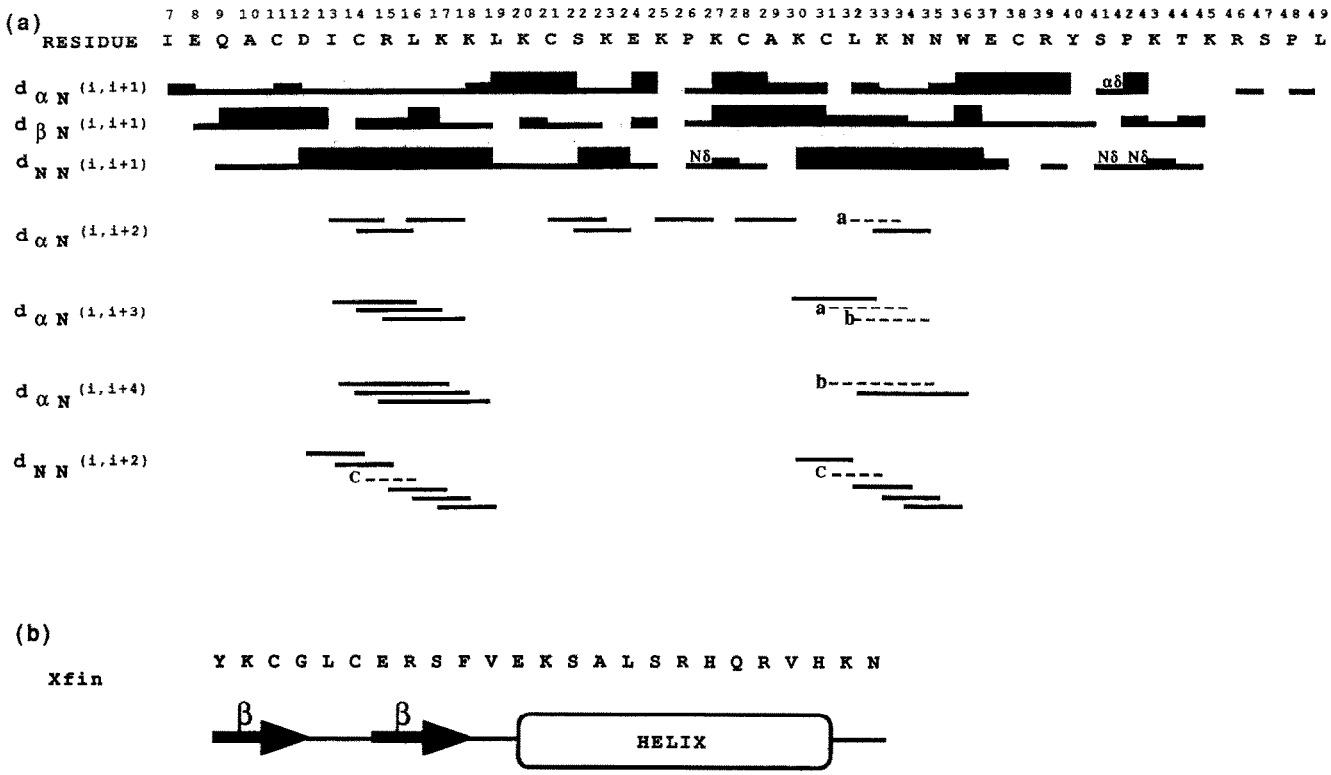

Fig. 1. (a) Summary of the sequential and medium-range NOE connectivities observed for the GAL4 (7–49) protein. Sequential NOE's are represented by boxes where the height of the box is a qualitative measure (strong, medium or weak) of NOE intensity in a 200 ms NOESY spectrum recorded at 283 K. Medium range NOE's are presented by bars connecting the appropriate residues. The letters **a**, **b** and **c** indicate ambiguities; for example, either one or both of the two medium range NOE's labeled **a** is present. (b) Diagrammatic illustration of the secondary structure, in relation to the residues co-ordinating the zinc, determined for the Xfin-31 zinc finger by Lee et al. [11,12].

recorded with mixing times of 30, 60, 90, 120, 200 and 300 ms. TOCSY spectra were recorded with mixing times of 30 and 100 ms using a spin locking field of 5 kHz. The TOCSY experiments employed the DIPSI-2 mixing sequence [20] and were acquired using the method of Rance [21]. The NOESY, TOCSY and 2Q spectra were acquired with a data size of 4 K (t_2) × 512 (t_1) points and 32 scans for each t_1 point. Finally a relayed COSY (RELAY) spectrum was recorded with a mixing time of 36 ms, a data size of 4 K (t_2) × 1 K (t_1) points and 64 scans for each t_1 point.

3. RESULTS

A complete assignment of the ^{1}H NMR spectrum of GAL4 (7–49) at pH 5.4 and 283 K was obtained using standard procedures. Assignments within a given spin system were made using 2QF COSY, RELAY, TOCSY and 2Q spectra. Sequential assignment was achieved in the normal way by analysis of the NOESY spectrum. Although the protein is small, the high lysine/arginine content complicated the analysis. The excellent chemical shift dispersion of the NH signals (10.10–7.02 ppm) enabled assignment of most of these long side chain spin systems in TOCSY spectra recorded with a mixing time of 100 ms. Some spectra were also recorded at 288 K and 293 K in an attempt to resolve the spin systems of lysines 20, 25 and 45 but assignments of the side chains of these 3 residues remain tentative. These spectra did, however, confirm that the structure of the protein is stable over this range. Finally, the side chain aromatic and NH_2 groups were assigned using the NOESY spectrum.

The sequential NOE connectivities observed are summarised in Fig. 1a and the assignments are given in Table I. A sequence of NOESY spectra with mixing times of between 30 and 120 ms were also recorded to help quantitate the distance constraints for calculation of the three dimensional structure.

All the samples prepared for NMR spectroscopy underwent spontaneous degradation over a period of days resulting in the appearance of additional cross-peaks in the spectra. It seems likely that these are due to forms of the protein that have either lost one of the two metal ions or perhaps some of the cysteines may no longer be correctly co-ordinated. Because of this, the lack of sequential NOE's between the C-terminal residues of the protein prevents a completely unambiguous identification.

4. DISCUSSION

The chemical shift dispersion of signals within the ^{1}H NMR spectrum clearly demonstrates the presence of a well defined folded structure; the upfield shifts of the

Table I

^{1}H NMR Chemical Shifts of GAL4 (7–49) at pH 5.4 and 283 K (relative to internal TSP)

	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	Others
7 Ile	–	3.87	1.96	1.22,1.51,0.98	0.91	
8 Glu	8.91	4.45	1.91,2.10	2.26,2.33		
9 Gln	8.69	4.51	2.10,2.20	2.56		7.02,7.77 ($N^{\epsilon}H$)
10 Ala	8.44	4.65	1.56			
11 Cys	9.42	4.17	2.21,2.93			
12 Asp	9.20	4.14	2.63			
13 Ile	8.10	3.73	1.78	0.91,1.55,0.68	0.68	
14 Cys	7.86	3.81	2.82,3.27			
15 Arg	8.57	4.12	1.84,1.96	1.58	3.21,3.36	7.33 ($N^{\epsilon}H$)
16 Leu	8.01	4.08	1.77	1.62	0.86	
17 Lys	8.27	4.14	1.52,1.62	1.29,1.35	1.97	2.37,2.49 ($C^{\epsilon}H$)
18 Lys	7.73	3.93	1.88,2.16	1.37	1.69,1.76	3.04 ($C^{\epsilon}H$) 7.59 ($N^{\zeta}H$)
19 Leu	8.72	4.67	1.52,1.86	1.60	0.90,1.10	
20 Lys	8.51	4.37	1.77	1.86	1.65	3.21 ($C^{\epsilon}H$) 7.25 ($N^{\zeta}H$)
21 Cys	8.95	4.72	2.92			
22 Ser	9.26	4.52	4.03,4.33			
23 Lys	9.57	4.58	1.97,2.05	1.43,1.47	1.58	2.92,3.00 ($C^{\epsilon}H$)
24 Glu	7.25	4.11	1.91,2.05	2.26,2.33		
25 Lys	8.52	4.54	1.58,1.75	1.24,1.33	2.03	2.39 ($C^{\epsilon}H$)
26 Pro	–	4.49	2.23,2.52	1.89,2.08	3.46,3.80	
27 Lys	7.92	5.49	1.66	(1.53 –	1.60)	3.05,3.11 ($C^{\epsilon}H$)
28 Cys	10.10	4.67	3.58,3.70			
29 Ala	8.37	4.03	1.52			
30 Lys	8.34	3.99	2.03	1.47,1.55	1.70	2.97,3.07 ($C^{\epsilon}H$)
31 Cys	8.01	4.02	2.80,3.32			
32 Leu	8.71	4.03	1.78,1.52	1.52	0.91	
33 Lys	7.87	3.93	1.69,1.84	1.33	1.58	2.92 ($C^{\epsilon}H$)
34 Asn	7.02	4.19	0.57,1.18		6.48,7.09 ($N^{\delta}H$)	
35 Asn	7.51	4.43	2.68,3.18		6.85,7.61 ($N^{\delta}H$)	
36 Trp	8.54	5.23	3.20,3.40		7.00	8.05 ($C^{\epsilon}H$) 10.15 ($N^{\epsilon}H$) 6.86,7.40 ($C^{\zeta}H$) 7.07 ($C^{\eta}H$)
37 Glu	8.83	4.30	1.94,1.98	2.19,2.30		
38 Cys	8.25	4.41	2.58,3.18			
39 Arg	8.15	4.74	1.69,1.75	1.53,1.60	3.17,3.25	7.31 ($N^{\epsilon}H$)
40 Tyr	9.04	4.76	2.78,2.81		6.94	6.90 ($C^{\epsilon}H$)
41 Ser	9.13	4.74	3.87,3.94			
42 Pro	–	4.55	1.97,2.34	2.07	3.78,3.93	
43 Lys	8.65	4.31	1.78,1.85	1.47	1.72	3.02 ($C^{\epsilon}H$)
44 Thr	8.26	4.33	4.21	1.22		
45 Lys	8.51	4.33	1.75,1.84	1.46	1.70	3.01 ($C^{\epsilon}H$)
46 Arg	8.42	4.33	1.73,1.83	1.63	3.21	7.29 ($N^{\epsilon}H$)
47 Ser	8.15	4.56	3.75			
48 Pro	–	4.47	2.01,2.29	2.04	3.77,3.84	
49 Leu	7.93	4.16	(1.58)		0.89,0.93	

$C^{\beta}H$'s of Asn-34 and the downfield shifts of the NH and $C^{\beta}H$'s of Cys-28 are particularly notable. One dimensional spectra of the apoprotein recorded after dialysis at low pH in the absence of zinc show substantial changes (data not shown) supporting the evidence [9] from mass spectrometry and other methods that the protein contains two zinc ions per monomer.

The observed patterns of sequential and medium range NOE connectivities (see Fig. 1a) provide information about the secondary structure of the GAL4 (7–49) fragment. Most striking are the two continuous stretches of strong d_{NN} connectivities running from Asp-12 to Leu-19 and from Lys-30 to Tyr-36, together with a number of $d_{\alpha N}$ (i,i+2), d_{NN} (i,i+2), $d_{\alpha N}$ (i,i+3) and $d_{\alpha N}$ (i,i+4) connectivities. The appropriate region of the NOESY spectrum illustrating the d_{NN} NOE connectivities is shown in Fig. 2. These patterns of NOE connectivities are indicative of helix although the overlap of the $C^{\alpha}H$ resonances precludes a distinction from a series of turns at this stage [22,23].

Strong sequential $d_{\alpha N}$ and weak d_{NN} NOE connectivities are observed from Leu-19 to Ser-22 and from Lys-27 to Ala-29 which possibly suggests the presence of extended chain conformation at these residues. The observation of several NOE's between Ser-22, Lys-23, Glu-24 and Cys-28 suggests that these regions pack against each other. Consistent with this the strong d_{NN} NOE's between Ser-22 and Glu-24 and the $d_{\alpha N}$ (i,i+2) NOE's in this region are also indicative of turns [22,23].

The residues SP(X)(X) have recently been proposed

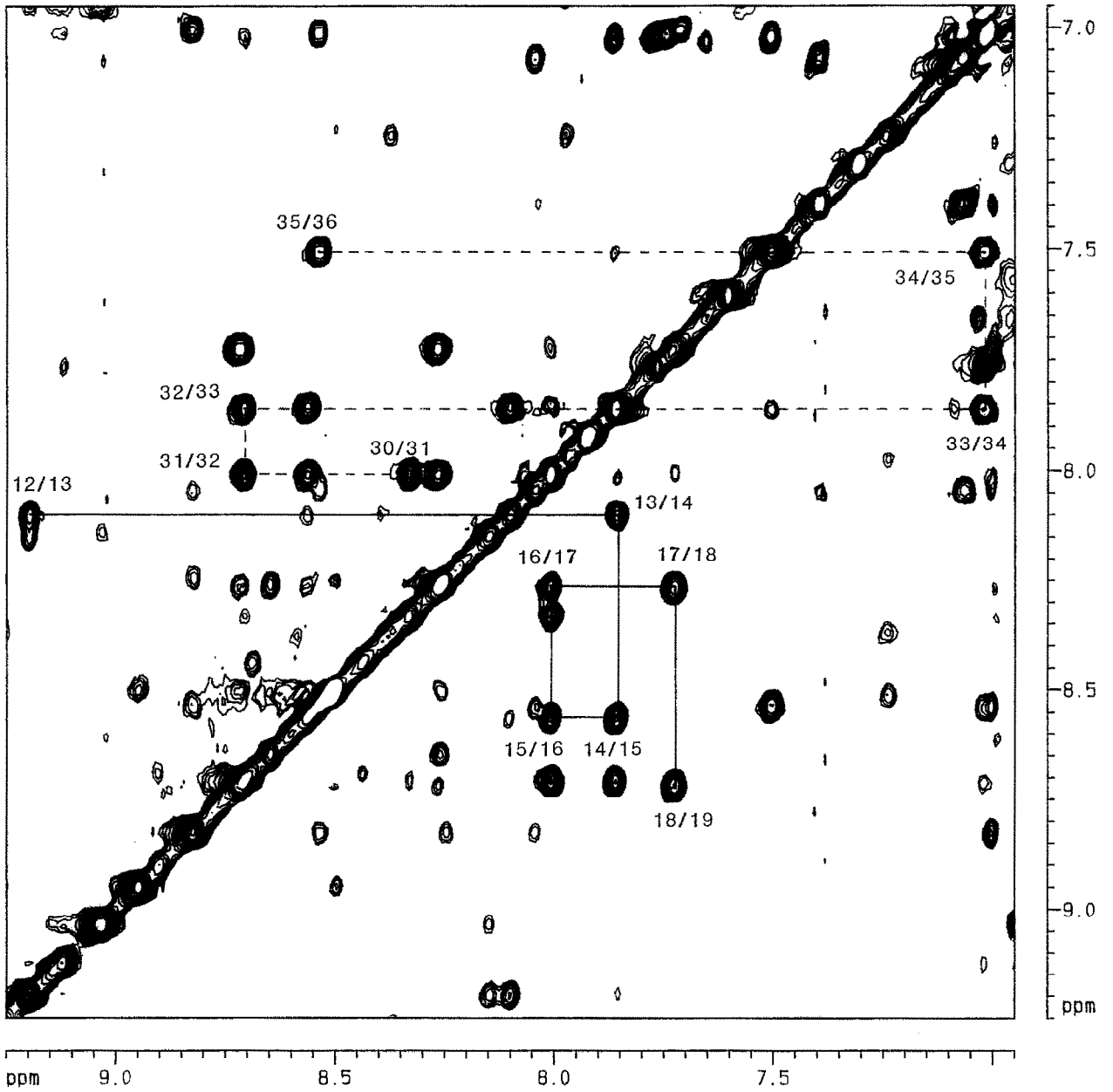

Fig. 2. One region of the NOESY spectrum (T_m = 200 ms, 283 K) of GAL4 (7–49) containing the two continuous stretches of strong sequential d_{NN} connectivities running from Asp-12 to Leu-19 (—) and Lys-30 to Trp-36 (---).

to be a new DNA binding motif [24,25]. Analysis of a number of protein sequences suggested that such motifs would be found on either side of specific DNA recognising structures but not within them [25]. Interestingly we find a large number of NOEs between the Tyr-40 adjacent to the first such motif (residues 41 to 44) and the cysteine rich region of the DNA binding domain indicating that it is buried inthe interior and the SP(X)(X) motif is tightly packed against the domain.

If the protein existed in a structure similar to that found in the C_2H_2 type zinc finger then one would have expected to find a helix from approximately Lys-21 to Lys-34, but extended chain in the amino terminal region of the protein (see Fig. 1b). Comparison of this secondary structure with the results obtained, suggests that the structure of the cysteine rich region of the DNA binding domain is different from that found in the C_2H_2 zinc finger and that GAL4 may represent a novel DNA binding motif. This conclusion is supported by the observation of several NOE's between residues near the N and C termini of this fragment, for example between Gln-9 and Tyr-40. This suggests that, in the cysteine rich region of the DNA binding domain of GAL4, the N and C termini of the fragment are close to each other; in the C_2H_2 zinc finger, they are at opposite ends of the domain. The calculation of the three dimensional struc-

ture of GAL4 (7-49) using distance geometry and restrained molecular dynamics is presently in progress in this laboratory and will be described in detail elsewhere. This structure should provide some insight into the mode of interaction of this class of protein with DNA.

Acknowledgements: We thank Jennifer Barna, Adrian Davis and Jane Povey for help with computing NMR and biochemical experiments, Caroline Hill for useful discussions, Brian Green for mass spectrometry, Juliette Jacoby for amino acid analysis, Len Packman for protein sequence analysis and Hitoshi Kakidani and Mark Ptashne for expression plasmids and *E. coli* strains. This work was supported by SERC and its Cambridge Centre for Molecular Recognition.

REFERENCES

[1] Klug, A. and Rhodes, D. (1987) Trends Biochem. Sci. 12, 464-469.
[2] Evans, R.M. and Hollenberg, S.M. (1988) Cell 52, 1-3.
[3] Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Cell 40, 767-774.
[4] Johnston, M. (1987) Nature 328, 353-355.
[5] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609-1614.
[6] Diakun, G.P., Fairall, L. and Klug, A. (1986) Nature 324, 698-699.
[7] Povey, J.F., Diakun, G.P., Garner, C.D., Wilson, S.P. and Laue, E.D. (1990) FEBS Lett. 266, 142-146.
[8] Pan, T. and Coleman, J.E. (1990) Proc. Natl. Acad. Sci. USA 87, 2077-2081.
[9] Gadhavi, P.L., Alefounder, P.R., Green, B. and Laue, E.D. (in preparation).
[10] Parraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. (1988) Science 241, 1489-1492.
[11] Lee, M.S., Cavanagh, J. and Wright, P.E. (1989) FEBS Lett. 254, 159-164.
[12] Lee, M.S., Gippert, G.P., Kizhake, V.S., Case, D.A. and Wright, P.E. (1989) Science 245, 635-637.
[13] Bodenhausen, G., Vold, R.I. and Vold, R.R. (1980) J. Magn. Reson. 37, 93-106.
[14] Marion, D. and Wüthrich, K. (1983) Biochem. Biophys. Res. Comm. 113, 967-974.
[15] Marion, D. and Bax, A. (1988) J. Magn. Reson. 79, 352.
[16] Hoult, D.I., Chen, C.N., Eden, H. and Eden, M. (1983) J. Magn. Reson. 51, 110.
[17] Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) J. Am. Chem. Soc. 104, 6800-6801.
[18] Rance, M. and Wright, P.E. (1986) J. Magn. Reson. 66, 372-378.
[19] Dalvit, C., Rance, M. and Wright, P.E. (1986) J. Magn. Reson. 69, 356-361.
[20] Shaka, A.J., Lee, C.J. and Pines, A. (1988) J. Magn. Reson. 77, 274-293.
[21] Rance, M. (1987) J. Magn. Reson. 74, 557-564.
[22] Wüthrich, K., Billeter, M. and Braun, W. (1984) J. Mol. Biol. 180, 715-740.
[23] Wagner, G., Neuhaus, D., Wörgötter, E., Vašák, M., Kägi, J.H.R. and Wüthrich, K. (1986) J. Mol. Biol. 187, 131-135.
[24] Suzuki, M. (1989) EMBO J. 8, 797-804.
[25] Suzuki, M. (1989) J. Mol. Biol. 207, 61-84.