

Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy

James G. Omichinski¹, G. Marius Clore¹, Kazuyasu Sakaguchi², Ettore Appella² and Angela M. Gronenborn¹

¹Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA and ²Laboratory of Cell Biology, Building 37, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received 16 August 1991

A 39-residue peptide (p7-DF) containing the two zinc binding domains of the p7 nucleocapsid protein was prepared by solid-phase peptide synthesis. The solution structure of the peptide was characterized using circular dichroic and nuclear magnetic resonance spectroscopy in both the presence and absence of zinc ions. Circular dichroic spectroscopy indicates that the peptide exhibits a random coil conformation in the absence of zinc but appears to form an ordered structure in the presence of zinc. Two-dimensional nuclear magnetic resonance spectroscopy indicates that the two zinc binding domains within the peptide form stable, but independent, units upon the addition of 2 equivalents of ZnCl₂ per equivalent of peptide. Structure calculations on the basis of nuclear Overhauser (NOE) data indicate that the two zinc binding domains have the same polypeptide fold within the errors of the coordinates (~0.5 Å for the backbone atoms, the zinc atoms and the coordinating cysteine and histidine ligands). The linker region (Arg¹⁷–Gly²³) is characterized by a very limited number of sequential NOEs and the absence of any non-sequential NOEs suggest that this region of the polypeptide chain is highly flexible. The latter coupled with the occurrence of a large number of basic residues (four out of seven) in the linker region suggests that it may serve to allow adaptable positioning of the nucleic acid recognition sequences within the protein.

Nucleocapsid protein p7: HIV-1: NMR: Solution structure

1. INTRODUCTION

Infectious human immunodeficiency virus type 1 (HIV-1) particles contain almost exclusively unspliced, viral genomic RNA [1]. Mutational analysis of viral sequences has demonstrated that major *cis*-acting packaging signals are located near the 5' end of the RNA, 3' to the splice donor site [2–5] with additional packaging signals present in other regions [4,6]. The sole *trans*-acting factor required for packaging is the *gag* polyprotein, since mutations in *gag* frequently prevent assembly [7]. The *gag* precursor polyprotein is cleaved into mature proteins at the time of virus budding,

releasing amongst others a 7 kDa nucleocapsid protein (p7) which contains a tandem repeat of the Cys-X₂-Cys-X₄-His-X₄-Cys sequence motif that is conserved in all retroviruses [8,9]. Deletions or changes of these conserved cysteine or histidine residues impair or alter packaging of RNA into virions [10], supporting the hypothesis that the retroviral Cys₂HisCys motif may be important for nucleic acid binding, analogous to the classical Cys₂His₂ zinc finger motif found in transcription factors. Recently, preferential binding of recombinant HIV-1 *gag* polyprotein to two 200-nucleotide-long regions of RNA derived from the 5' end of the *gag* coding region was shown to involve the N-terminal Cys₂HisCys motif [11]. Although there is still some controversy about the involvement of zinc in retroviral nucleocapsid proteins [12,13], it has been shown unambiguously for the case of HIV-1 p7 that protein isolated from virions, as well as a cloned nucleocapsid protein of 140 residues comprising the p7 sequence, binds two equivalents of zinc tightly and stoichiometrically [14,15]. In addition, small retroviral zinc finger-like synthetic peptides have been shown to form folded domains upon binding of zinc [16–18]. In this report we present a structural investigation by CD and NMR spectroscopy of a synthetic 39 residue peptide comprising residues 13 to 51 of the p7 nucleocapsid protein of

Abbreviations: HIV-1, human immunodeficiency virus type 1; p7-DF, 39-residue peptide comprising residues 13–51 of the p7 nucleocapsid protein and containing two zinc binding domains; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; P.COSY, two-dimensional primitive correlated spectroscopy; PE.COSY, two-dimensional primitive exclusive correlated spectroscopy; HOHAHA, two-dimensional homonuclear Hartmann-Hahn spectroscopy; HPLC, high-performance liquid chromatography.

Correspondence address: G.M. Clore or A.M. Gronenborn, Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

HIV-1 and encompassing the two zinc binding domains (Fig. 1A) with the aim of assessing the structural similarity between both domains and of determining whether the two domains have a specific spatial orientation relative to each other.

2. MATERIALS AND METHODS

The 39-residue peptide comprising the two zinc binding domains of HIV-1 p7 was prepared by solid-phase peptide synthesis [19] on an Applied Biosystems (Foster City, CA) 430A peptide synthesizer. Protected *t*-butoxycarbonyl-L-amino acids were purchased from Peptides International (Lexington, KE) and protected phenylacetamidomethyl resin from Applied Biosystems (Foster City, CA). The final protected resin was cleaved using a modified low/high HF cleavage method [20]. The extracted peptide (5% aqueous acetic acid/1 mM dithiothreitol) was purified by gel filtration chromatography on a G-50 (Pharmacia Fine Chemicals, Piscataway, NJ) column and by HPLC on a Vydac (Hesperia, CA) C₄-semipreparative (10 mm × 25 cm) reverse-phase column. The peptide was eluted from the HPLC column with an acetonitrile-water solvent gradient system containing 0.1% trifluoroacetic acid, and was >95% pure. Identity of the synthetic peptide was confirmed by amino acid analysis, as well as by analysis of the NMR data.

CD spectra were recorded from 190 to 260 nm on a Jasco J-500 spectropolarimeter at 25°C. The peptide (7.5 μM) was dissolved in aqueous ZnCl₂ (15 μM) at pH 5.8. The spectra are presented as plots of mean residue ellipticity and have been corrected for background interference.

Samples for NMR contained 4.3 mM peptide and 9 mM ZnCl₂ in either 99.996% D₂O or 90% H₂O/10% D₂O at pH 5.8. The following spectra were recorded in both H₂O and D₂O at 600 MHz on a Bruker AM 600 spectrometer at 15°C and 25°C: NOESY [21] (with mixing times of 50 and 150 ms), HOHAHA [22] (with mixing times ranging from 32 to 55 ms), P.E.COSY [23] and P.COSY [24].

3. RESULTS AND DISCUSSION

The CD spectrum of p7-DF in the absence of zinc is characteristic of a random coil conformation (Fig. 1B). The addition of 2.1 mol of zinc per mol of peptide induces a major change in the CD spectra of p7-DF at pH 5.8 (Fig. 1B). In the presence of zinc a positive maximum is observed at ~214 nm, and there is a significant decrease in the negative ellipticity as well as a small red shift of the minimum. This result indicates that upon complexation with zinc, the p7-DF peptide adopts a different fold from that observed for the free peptide. In addition, the NMR spectrum of p7-DF in the absence of zinc is characterized by a very narrow chemical shift dispersion, confirming that the free peptide exists in a random conformation (data not shown). Subsequent NMR studies were therefore conducted with peptide (4.3 mM) in the presence of ZnCl₂ (9 mM) at pH 5.8.

The complete sequential assignment of the p7-DF peptide-Zn complex was accomplished using conventional 2D NMR methodology [25,26]. To resolve ambiguities arising from chemical shift degeneracy, all spectra were recorded at two temperatures (15 and 25°C). Examples of selected regions of the NOESY spectrum illustrating NOE connectivities involving the

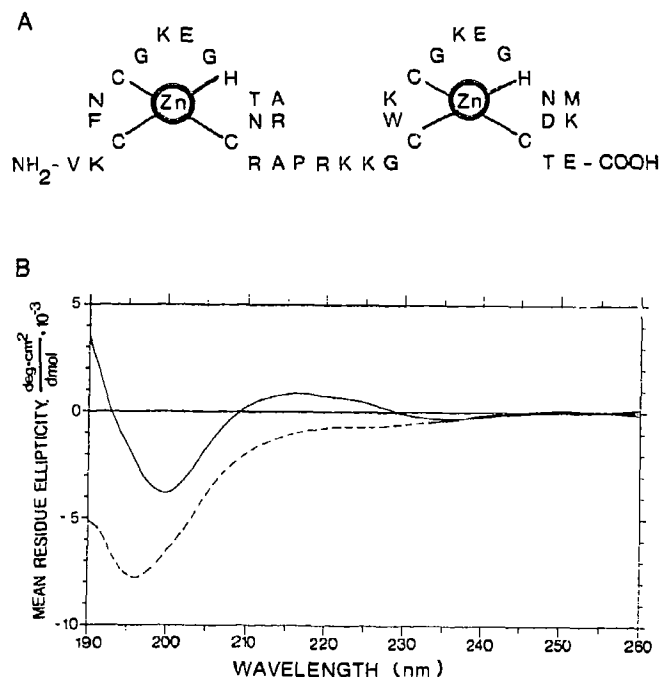


Fig. 1. (A) Amino acid sequence of the p7-DF peptide corresponding to positions 13–51 of the p7 nucleocapsid protein. (B) Circular dichroism (CD) spectra of p7-DF in the presence (—) and absence (-----) of zinc ions.

backbone NH protons are shown in Fig. 2, a summary of the short range NOE data involving the NH, C^αH and C^βH protons is given in Fig. 3, and the complete list of assignments is presented in Table I.

Qualitative interpretation of the sequential and short-range NOE data presented in Fig. 3 reveals striking similarities between the two zinc binding domains (Cys³–Cys¹⁶ and Cys²⁴–Cys³⁶). There are NH(i)–NH(i+1) cross peaks observed from Cys³ to Lys⁸ and from Cys²⁴ to Lys²⁹ with the strongest cross peaks occurring in the regions comprising Asn⁵ to Gly⁷ and from Lys²⁶ to Gly²⁸. Likewise, NH(i)–NH(i+1) connectivities from Ala¹³ to Cys¹⁶ in the N-terminal domain are matched by the corresponding NOEs in the C-terminal one from Met³⁴ to Cys³⁷. No NH(i)–NH(i+1) cross peaks are observed for the linker region between the two domains (Arg¹⁷ to Gly²³). Medium and long range NOE cross peaks also support the view that both domains are structurally very similar. Both N-terminal cysteines, Cys³ from the N-terminal domain and Cys²⁴ from the C-terminal one, exhibit the same pattern of C^βH(i)–NH(i+3, i+4 and i+5) connectivities. The two conserved zinc binding histidines, His¹¹ and His³², exhibit numerous connectivities from the imidazole ring protons (C^{δ2}H and C^{ε1}H) to residues i-3 (Lys⁸, Lys²⁹), i+4 (Asn¹⁵, Asp³⁶) and i+5 (Cys¹⁶, Cys³⁷), the latter also being involved in zinc chelation.

Interestingly, we did not observe any NOE cross-peaks between the two zinc binding domains for any of

the conditions used in our NMR experiments. This is consistent with the notion of a flexible linker region characterized by a paucity of sequential NOEs and no non-sequential NOEs. The presence of a mobile joint between the two zinc binding domains of the p7-DF peptide is in contrast to solution NMR data obtained for a double zinc finger peptide of the classical Cys₂His₂ type (J.G.O., G.M.C. and A.M.G., unpublished data) derived from a human enhancer binding protein [27,28] and for the two Cys₄ zinc binding domains of the glucocorticoid receptor [29]. In both these protein fragments the linker region is much more rigid and the two domains display a specific orientation with respect to each other. In addition, results from X-ray diffraction studies on a Cys₂His₂ type zinc finger/DNA complex demon-

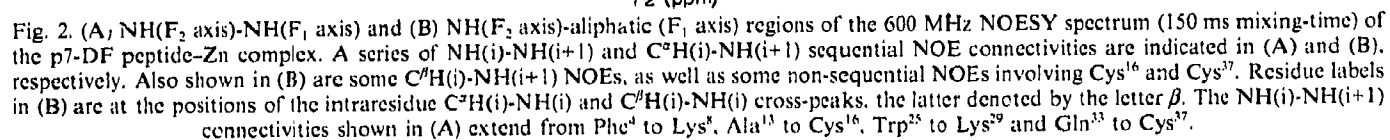
strate hydrogen bonding between adjacent fingers [30], a feature clearly absent in the p7 peptide structure.

To obtain a better idea of the polypeptide fold of the two zinc binding domains, we proceeded to carry out a series of structure calculations on the two separate domains (residues 1–17 and 22–39) using the hybrid distance geometry-simulated annealing method [31]. The calculations were based on 147 and 148 approximate NOE derived interproton distance restraints for the N- and C-terminal domains, respectively, which were classified into three ranges, 1.8–2.7 Å, 1.8–3.3 Å and 1.8–5.0 Å, corresponding to strong, medium and weak NOEs [26]. In addition, ϕ backbone torsion angle restraints (–10° to –175°) were employed to ensure that the ϕ angles for all residues, except glycine, were located in

Table I
Proton resonance assignments of the p7-DF peptide-zinc complex at 15°C and pH 5.8

Residue	Chemical shift (ppm) ^a			
	NH	C ^α H	C ^β H	Others
Val ¹		3.77	2.01	C ^γ H ₃ 0.83, 0.83
Lys ²	8.52	4.41	1.50, 1.44	C ^γ H 1.01, 1.01; C ^δ H 1.07, 1.07; C ^ε H 2.86, 2.86
Cys ³	8.31	4.06	2.78, 1.79	
Phe ⁴	8.84	4.47	3.22, 3.11	C ^δ H 7.29; C ^ε H 7.34; C ^ζ H 7.34
Asn ⁵	9.56	4.71	3.10, 2.75	NH ₂ 8.08; 6.83
Cys ⁶	8.83	4.91	3.23, 2.48	
Gly ⁷	7.94	4.09, 3.77		
Lys ⁸	8.14	4.41	1.87, 1.81	C ^γ H 1.72, 1.72; C ^δ H 1.50, 1.50; C ^ε H 3.02, 3.02
Glu ⁹	8.44	4.52	2.11, 1.86	C ^γ H 2.37, 2.21
Gly ¹⁰	8.56	4.47, 3.74		
His ¹¹	7.22	4.80	3.21, 3.21	C ^{δ2} 7.06; C ^{ε1} 7.41
Thr ¹²	8.27	4.68	4.48	CH ₃ 1.15
Ala ¹³	8.93	4.63	1.43	
Arg ¹⁴	8.11	4.10	1.82, 1.69	C ^γ H 1.61, 1.61; C ^δ H 3.16, 3.16; NH 7.21
Asn ¹⁵	7.93	4.98	3.00, 2.58	NH ₂ 7.80, 6.93
Cys ¹⁶	7.52	3.86	3.34, 2.95	
Arg ¹⁷	8.36	4.23	1.85, 1.74	C ^γ H 1.64, 1.64; C ^δ H 3.12, 3.12; NH 7.20
Ala ¹⁸	8.26	4.54	1.30	
Pro ¹⁹		4.24	2.18, 1.78	C ^γ H 1.93, 1.93; C ^δ H 3.74, 3.54
Arg ²⁰	8.42	4.25	1.78, 1.70	C ^γ H 1.58, 1.58; C ^δ 3.08, 3.08; NH 7.12
Lys ²¹	8.30	4.24	1.78, 1.69	C ^γ H 1.33, 1.33; C ^δ H 1.55, 1.55; C ^ε H 2.86, 2.86
Lys ²²	8.39	4.33	1.77, 1.70	C ^γ H 1.41, 1.37; C ^δ H 1.64, 1.64; C ^ε H 2.93, 2.93
Gly ²³	8.24	3.75, 3.56		
Cys ²⁴	8.17	4.00	2.86, 2.01	
Trp ²⁵	8.60	4.48	3.43, 3.40	C ^{γ1} H 7.34; C ^{ε3} H 7.60; C ^{ε1} H 10.13; C ^{ε2} H 7.50; C ^{ε3} H 7.14; C ^γ H 7.23
Lys ²⁶	9.44	4.25	2.24, 1.66	C ^γ H 1.31, 1.31; C ^δ H 1.82, 1.82; C ^ε H 2.94, 2.94
Cys ²⁷	8.69	4.94	3.22, 2.53	
Gly ²⁸	8.19	4.07, 3.85		
Lys ²⁹	8.47	4.37	1.88, 1.82	C ^γ H 1.77, 1.60; C ^δ H 1.41, 1.41; C ^ε H 3.02, 3.02
Glu ³⁰	8.55	4.22	1.99, 1.86	C ^γ H 2.22, 2.11
Gly ³¹	8.67	4.34, 3.69		
His ³²	7.14	4.78	3.19, 3.19	C ^{δ2} H 6.84; C ^{ε1} H 7.48
Gln ³³	8.95	4.60	1.92, 1.92	C ^γ H 2.44, 2.37; NH ₂ 7.66, 6.88
Met ³⁴	8.94	4.83	2.15, 2.15	C ^γ H 2.51, 2.51; C ^ε H ₃ 2.09
Lys ³⁵	8.63	4.24	1.88, 1.73	C ^γ H 1.22, 1.22; C ^δ H 1.33, 1.33; C ^ε H 3.00, 3.00
Asp ³⁶	7.94	4.89	2.96, 2.54	
Cys ³⁷	7.69	3.68	3.29, 2.87	
Thr ³⁸	8.32	4.45	4.42	C ^γ H 1.08
Glu ³⁹	8.62	4.06	2.13, 1.86	C ^γ H 2.37, 2.21

^aChemical shifts are reported relative to 4,4-dimethyl-4-silapentane-1-sulfonate.



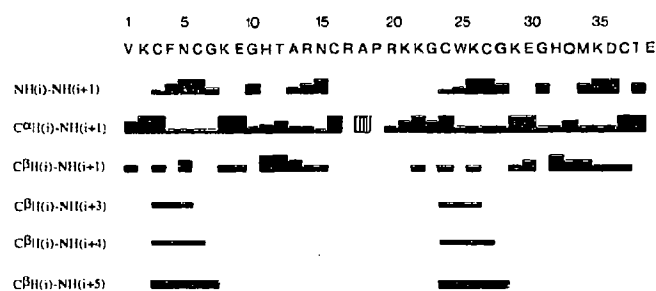


Fig. 3. Summary of short-range NOEs involving the NH, C^αH and C^βH protons, as well as the C^αH protons of proline, observed for the p7-DF peptide-Zn complex. The NOEs are classified into strong, medium and weak according to the thickness of the lines. The C^αH (Ala¹⁸)-C^βH (Pro¹⁹) NOE is indicated by the hatched box along the same line as the C^αH(i)-NH(i+1) NOEs.

the allowed region of the Ramachandran plot [32], and the coordination geometry of the zinc was restrained to be tetrahedral [27]. Initial structure calculations excluding the zinc atom indicated unambiguously that the zinc had to be coordinated to the N^{ε2} atom of the histidine ligand. The final set of simulated annealing structures, 15 for each domain, are not high resolution structures, since neither stereospecific assignments nor extensive torsion angle restraints were derived from the data [33], but are intended to supply information about the polypeptide fold. All the structures satisfy the experimental restraints within the errors of the data (no interproton distance violations greater than 0.3 Å), display very small deviations from idealized covalent geometry and have good non-bonded contacts. The atomic rms distribution of the individual structures about the mean coordinate positions is ~0.5 Å for the backbone atoms of residues 3/24 to 16/37, the zinc atoms and the cysteine and histidine side chains. The atomic rms difference for this same set of atoms between the two regularized mean structures is 0.47 Å. Thus, the structures of the two fingers can be considered to be identical within the errors of the present coordinates. A smooth backbone atom representation of the two fingers is shown in Fig. 4A and the dotted line indicates the linker region whose conformation is not determined by the data. A superposition of the backbone, the zinc atoms and the histidine and cysteine side chains for residues 3/24 to 16/37 of the two regularized mean structures is shown in Fig. 4B. The structure of the two domains basically consists of a β-hairpin-like structure at the N-terminus containing the first two metal bridging cysteines, followed by a loop which is tethered to the zinc atom by the histidine and last cysteine side chain.

In summary, the p7-DF peptide, comprising residues 13–51 of the p7 nucleocapsid protein of HIV-1, is characterized by a random coil conformation in the absence of zinc but forms a highly ordered structure upon addition of equimolar amounts of zinc. The solution structure of the zinc complex consists of two nearly identical zinc binding domains (Cys³-Cys¹⁶ and Cys²⁴-Cys³⁷)

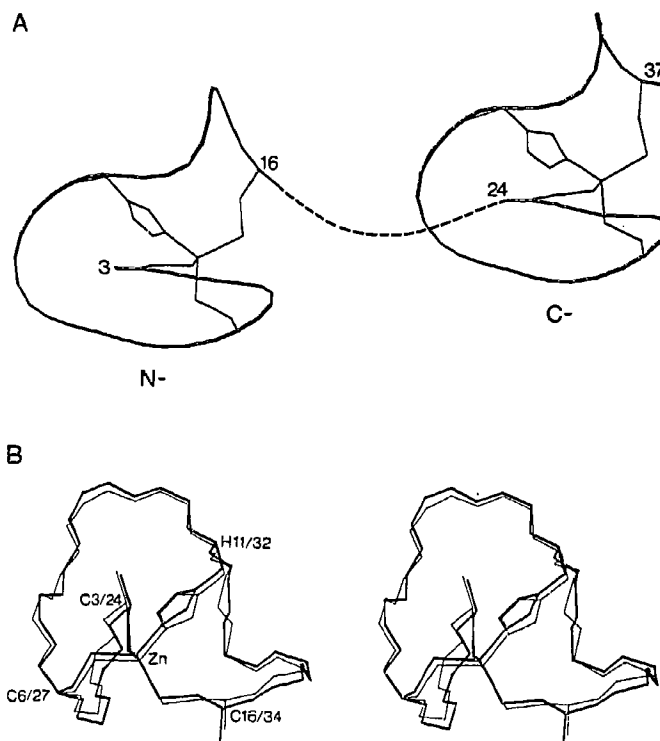


Fig. 4. (A) Smooth backbone atom representation of the N- and C-terminal zinc binding domains of the p7-DF peptide-Zn complex. The linker region from residues 17 to 23, whose conformation could not be determined, is shown as a dotted line. (B) Best-fit superposition of the regularized mean structures of the N- (thick line) and C-terminal (thin line) zinc binding domains.

which are structurally very similar to the single domain peptides previously determined by NMR [16,18]. They are separated by a very flexible linker segment (Arg¹⁷-Gly²³) which contains four basic residues (2 Arg and 2 Lys) and a conserved proline. Thus, it may be possible that this basic region is involved in contact with the nucleic acid. Further experiments are being carried out to examine in more detail how the complete p7 recognizes nucleic acids and exerts its biological effect.

Acknowledgements: This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C., A.M.G. and E.A.)

REFERENCES

- [1] Aronoff, R. and Linial, M. (1991) *J. Virol.* 65, 71–80.
- [2] Aldovini, A. and Young, R. (1990) *J. Virol.* 64, 1920–1926.
- [3] Clavel, F. and Orenstein, J.M. (1990) *J. Virol.* 64, 5230–5234.
- [4] Lever, A., Gottlinger, H., Haseltine, W. and Sodroski, J. (1989) *J. Virol.* 63, 4085–4087.
- [5] Mann, R., Mulligan, R.C. and Baltimore, D. (1983) *Cell* 33, 153–159.
- [6] Mann, R. and Baltimore, D. (1985) *J. Virology* 54, 401–407.
- [7] Gottlinger, H.G., Sodroski, J.C. and Haseltine, W.A. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5781–5785.
- [8] Covey, S.N. (1986) *Nucleic Acids Res.* 14, 623–633.
- [9] Copeland, T.D., Morgan, M.A. and Oroszlan, S. (1984) *Virology* 133, 137–145.

- [10] Gorelick, R.J., Nigida, S.M., Bess, J.R., Arthur, L.O., Henderson, L.E. and Rein, A. (1990) *J. Virol.* 64, 3207-3211.
- [11] Luban, J. and Goff, S. (1991) *J. Virol.* 65, 3203-3212.
- [12] Jentoft, J.E., Smith, L.M., Fu, X., Johnson, M. and Leis, J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7094-7098.
- [13] Gorelick, R.J., Henderson, L.E., Hanser, J.P. and Rein, A. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8420-8424.
- [14] South, T.L., Blake, P.R., Sowder, R.C., Arthur, L.O., Henderson, L.E. and Summers, M.F. (1990) *Biochemistry* 29, 7786-7789.
- [15] Fitzgerald, D.W. and Coleman, J.E. (1991) *Biochemistry* 30, 5195-5201.
- [16] Summers, M.F., South, T.L., Kim, B. and Hare, D.R. (1990) *Biochemistry* 29, 329-340.
- [17] Green, L.M. and Berg, J. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6403-6408.
- [18] South, T.L., Blake, P.R., Hare, D.R. and Summers, M.F. (1991) *Biochemistry* 30, 6342-6349.
- [19] Merrifield, R.B., Vizili, L.D. and Boman, H.G. (1982) *Biochemistry* 21, 5020-5031.
- [20] Tam, J.P., Heath, W.F. and Merrifield, R.B. (1983) *J. Am. Chem. Soc.* 105, 6442-6455.
- [21] Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.* 71, 4546-4553.
- [22] Bax, A. (1989) *Methods Enzymol.* 176, 151-168.
- [23] Mueller, L. (1987) *J. Magn. Reson.* 72, 191-196.
- [24] Marion, D. and Bax, A. (1988) *J. Magn. Reson.* 80, 528-533.
- [25] Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley, New York.
- [26] Clore, G.M. and Gronenborn, A.M. (1987) *Protein Eng.* 1, 275-288.
- [27] Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. and Gronenborn, A.M. (1990) *Biochemistry* 29, 9324-9334.
- [28] Sakaguchi, K., Appella, E., Omichinski, J.G., Clore, G.M. and Gronenborn, A.M. (1991) *J. Biol. Chem.* 266, 7306-7311.
- [29] Hard, T., Kellenbach, E., Boelens, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K.R., Gustafsson, J.-A. and Kaptein, R. (1990) *Science* 249, 157-160.
- [30] Pavletich, N.P. and Pabo, C.O. (1991) *Science* 252, 809-817.
- [31] Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.* 229, 317-324.
- [32] Richardson, J. (1981) *Adv. Prot. Chem.* 34, 167-339.
- [33] Clore, G.M. and Gronenborn, A.M. (1991) *Science* 252, 1390-1398.