

Homology modelling and ^1H NMR studies of human leukaemia inhibitory factor

David K. Smith^{a,b,e}, Herbert R. Treutlein^{c,e}, Till Maurer^{a,e}, Catherine M. Owczarek^{d,e},
Meredith J. Layton^{d,e}, Nicos A. Nicola^{d,e}, Raymond S. Norton^{a,e,*}

^a*NMR Laboratory, Biomolecular Research Institute, 381 Royal Pde, Parkville 3052, Australia*

^b*School of Biochemistry, University of Melbourne, Parkville 3052, Australia*

^c*Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, Royal Melbourne Hospital, Parkville 3050, Australia*

^d*Walter and Eliza Hall Institute for Medical Research, Royal Melbourne Hospital, Parkville 3050, Australia*

^e*The Cooperative Research Centre for Cellular Growth Factors, Parkville 3050, Australia*

Received 1 June 1994; revised version received 7 July 1994

Abstract

Human leukaemia inhibitory factor (LIF) is a glycoprotein with a diverse range of activities on many cell types. A molecular model of LIF has been constructed based mainly on the structure of the related cytokine granulocyte colony-stimulating factor, and refined using simulated annealing and molecular dynamics in water. The model was stable during molecular dynamics refinement and is consistent with known stereochemical data on proteins. It has been assessed by comparison with ^1H NMR data on the ionization behaviour of the six histidine residues in LIF, the imidazolium pK_a values of which range from 3.6 to 7.4. These pK_a values were assigned to individual histidine residues from NMR studies on a series of His \rightarrow Ala mutants. The environments of the histidine residues in the model account very well for their observed ionization behaviour. Furthermore, the model is consistent with mutagenesis studies which have defined a group of amino acid residues involved in receptor binding.

Key words: Leukaemia inhibitory factor; Homology model; NMR; Histidine; pK_a ; Cytokine

1. Introduction

Leukaemia inhibitory factor (LIF) is one of a number of pleiotropic cytokines with overlapping functions [1,2]. It is active on a range of cell types including adipocytes, megakaryocytes, neuronal cells and osteoblasts [1]. On the basis of amino acid sequence, LIF has been shown to be most closely related to oncostatin M (OSM) and ciliary neurotrophic factor (CNTF) and is part of a larger group of cytokines that includes interleukin-6 (IL-6) and granulocyte colony-stimulating factor (G-CSF) [3]. The high affinity LIF receptor (LIFR) is composed of two subunits, an α -subunit that, by itself, binds LIF with low affinity and a β -subunit, gp130 [4]. The LIFR α -subunit is also a component of the OSM and CNTF receptor complexes, while gp130 is a component of the IL-6, IL-11, OSM and CNTF receptor complexes [5]. These shared receptor subunits may explain some of the overlapping functions of this group of cytokines.

The tertiary structures of several helical cytokines

have been determined by X-ray crystallography or NMR spectroscopy and have been reviewed by Sprang and Bazan [6]. They all show a similar up-up-down-down, left-handed, four helical bundle topology and have been classified into two structural groups [6]. The long-chain cytokines, which include G-CSF and growth hormone (GH), are larger and have α -helices that are longer on average and more tightly packed than in the short-chain group, which includes IL-4 and granulocyte-macrophage colony-stimulating factor (GM-CSF) [6]. These structures provide a base from which models of related cytokines can be constructed.

LIF consists of 180 amino acids and has a molecular mass of about 20 kDa. Its tertiary structure has not been determined, although crystallisation of a human-like LIF has been reported [7]. Determination of its structure in solution by NMR, which requires that the protein be labelled with ^{15}N and ^{13}C [8], has not yet been achieved. We have, therefore, examined the structure of human LIF (hLIF) by homology modelling and ^1H NMR. A model was constructed based on the structure of G-CSF (the most closely related long-chain cytokine for which a 3D structure was available) and refined by simulated annealing and molecular dynamics in water. The model was then assessed experimentally by ^1H NMR and site-directed mutagenesis studies, which allowed the pK_a values of the six histidine residues to be determined and assigned to specific residues in the protein. These data were then compared with the environments of the histidine residues in the model. The amino acid residues in

*Corresponding author. Fax: (61) (3) 903 9655.
E-mail: ray@mel.dbe.csiro.au.

Abbreviations: LIF, leukaemia inhibitory factor; LIFR, LIF receptor; CNTF, ciliary neurotrophic factor; OSM, oncostatin M; IL, interleukin; G-CSF, granulocyte-colony stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; GH, growth hormone; h, human; TOCSY, total correlation spectroscopy; RMSD, root mean square difference.

hLIF which have been shown to be involved in binding to the LIFR α -subunit [9] were also examined for consistency with the model.

2. Materials and methods

2.1. Model building

The cytokine four helical bundle structures are described by the helices of the bundle being designated A, B, C, D from the N- to the C-terminus, and the loops being designated by the helices they connect. The sequences of human, murine, rat, porcine and ovine LIF and murine and human G-CSF [10] were aligned using a multiple sequence alignment procedure [11,12] and the resulting alignment of hLIF and human G-CSF (hG-CSF), shown in Fig. 1, was examined for consistency with the heptad repeat motif observed in these cytokines [13]. In this alignment the AB loop in LIF has an insertion of three amino acids relative to G-CSF, and LIF has an elongated N-terminal tail and a truncated C-terminal region. This alignment is identical to that of Bazan [3] for the A and D helices, the start of the AB loop and the end of the CD loop, but is different for the B and C helices. The alignment used here shows a better correspondence with the heptad repeat pattern for the B and C helices and a better distribution of gaps. The model of LIF was then created using the Homology module of Insight (Biosym Inc., San Diego, CA). The coordinates of the four bundle helices and the BC and CD loops were taken from hG-CSF and the N-terminal tail was built using the Homology module. As both human GH (hGH) [14] and hG-CSF [15] have a small helix in the N-terminal region of the AB loop, and hGH and hLIF have a broadly analogous disulfide bridge in this region, the helix in the AB loop of hG-CSF was copied to the LIF model. Because this helix is not seen in porcine GH [16] it has been suggested that in hGH it may form as a consequence of receptor binding [14] and that the similarity with hG-CSF may not be real [6]. Recently, however, it has been observed in a crystal structure of an hGH mutant that is not receptor bound [17]. The remainder of the AB loop was taken from a search of structures in the Protein Data Bank [18]. The secondary structure prediction of Bazan [3], based on sequence alignments, defined the helices slightly differently from those in the structure of hG-CSF and in this work. Bazan excluded from the D helix the 'D1 motif' (D₁₅₄VFQKKKLG in hLIF), which is seen in LIF, OSM, CNTF and G-CSF. As this region is in the D helix in hG-CSF [15] it was included in the D helix of the LIF model. The region around Cys¹³¹ and Cys¹³⁴ was not included in the C helix by Bazan [3]; in our case it was modelled from the C helix of hG-CSF but not subjected to helical restraints in the refinement. The A helix was modelled to be the length observed in hG-CSF [15] rather than the shorter length predicted by Bazan [3]. Finally, the model was surrounded by four layers of water molecules [19]. In simulations carried out in the absence of water the helical bundle was not stable, as polar side chains interacted with the helical bundle.

2.2. Model refinement

Refinement of the model was performed in several stages by a simulated annealing and molecular dynamics protocol using the X-PLOR program [20]. Simulated annealing from 600K to 300K was performed with 0.5–1 ps of molecular dynamics at each interval of 50K, followed by 5 ps of dynamics at 300K, in each case with a time step of 0.5 or 1 fs. The three disulfide bridges in LIF [21] were defined and a CO₂ to HN₁₊₄ upper bound distance constraint of 2.2 Å was specified for residues of the helix bundle, except for residues 131–135, which were excluded from helix C in order not to constrain the disulfide bridges involving Cys¹³¹ and Cys¹³⁴. For the first stage of refinement the C α atoms of the bundle helices were held fixed and the loops and side chains were allowed to relax using the simulated annealing protocol. The next stage released the C α atoms but applied the upper bound constraint to the bundle helices and the simulated annealing protocol was repeated. This was followed by 100 ps of molecular dynamics with the helical constraints on and 64 ps of unrestrained dynamics at 300K. No restraints derived from NMR or biological data on LIF were used in the model building or the refinement. The stereochemical quality of the model was assessed and compared to a range of known protein crystal structures by using the PROCHECK suite of programs [22].

2.3. Protein preparation

Histidine-to-alanine point mutations for His¹⁶, His¹⁹, His¹¹², His¹³⁸ and His¹⁴¹ in the hLIF amino acid sequence were constructed using a PCR-based technique and verified as described previously [9]. Human LIF and the His \rightarrow Ala mutants were expressed in *Escherichia coli* NM522 as fusion proteins with glutathione S-transferase, and subsequently cleaved with thrombin. Recombinant proteins were purified, characterised and assayed as described previously [9].

2.4. NMR

Samples for NMR were prepared by dissolving lyophilized protein in either 100% ²H₂O or 90% H₂O/10% ²H₂O. The pH was adjusted with ²HCl or NaO²H. The pH titration was performed in 100% ²H₂O and pH values are reported as uncorrected meter readings. Measurements were made at 26°C, a protein concentration of about 1 mM, and pH intervals of approximately 0.5 units in the pH range 1.8–8.4; the protein precipitated at pH 9. ¹H NMR spectra were recorded on Bruker AMX-600 and AMX-500 spectrometers. Probe temperatures were maintained using a B-VT1000E control unit and a Haake cooling bath, and were calibrated using methanol and ethylene glycol [23]. Solvent suppression was performed by selective, low-power irradiation during the relaxation delay. Two-dimensional spectra were acquired using the time-proportional phase incrementation method [24] for quadrature detection in ω_1 . A total correlation spectrum (TOCSY) [25] was collected at each pH using the DIPSI-2 [26] pulse sequence with a spin-lock period of 30 ms. Spectra were processed using UXNMR and AURELIA software (Bruker AG, Karlsruhe).

2.5. pK_a values

The pK_a values of the six histidine residues were determined by fitting the chemical shifts of the C(2)H and C(4)H imidazole protons to the Henderson–Hasselbalch equation for a single titrating group or for a proton undergoing two independent titrations [27]. As the pK_a values were not corrected for isotope effects, values in H₂O may be slightly lower [28]. The histidine residues were sequentially assigned by identifying the missing imidazole proton resonances in 1D and 2D TOCSY spectra of each of the five His \rightarrow Ala mutants. Resonances from the sixth histidine residue were identified by inference.

3. Results

3.1. Model description

The refined model of human LIF has the characteristic four α -helix bundle topology of the known helical cytokine structures (Fig. 2). Two small α -helices are present in the AB loop. The N-terminal helix in this loop is consistent with helices in hG-CSF [15] and hGH [14] but is more regular than the overlapping 3₁₀/ α -helices in hG-CSF. This was modelled into the initial LIF structure as a helical region and, although changed, has remained a helix throughout the refinement process. The small helix at the C-terminal end of the AB loop formed during the refinement process and is analogous to the second helix in the AB loop of hGH. Cys¹³¹ has remained in the C-terminus of the C helix but the region around Cys¹³⁴ is not part of the helix. The two disulfide bridges these residues are involved in, Cys¹²–Cys¹³⁴ and Cys¹⁸–Cys¹³¹, restrict the conformational space available to the region between Cys¹² and Cys¹⁸, causing a loop to form in the N-terminal tail of the model between Cys¹² and Cys¹⁸. The beginning of the 'D1 motif' has lost its helical structure but the rest has remained in the D helix. The three consecutive lysine residues in this region (158–160) sit above the helix bundle and would not have a destabili-

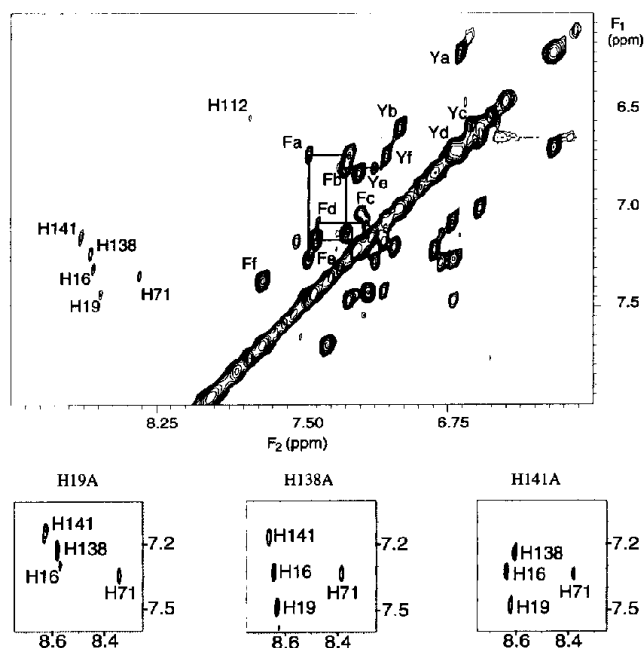


Fig. 3. Aromatic region of a TOCSY spectrum of hLIF in $^2\text{H}_2\text{O}$ at pH 4.5 and 26°C . The locations of all aromatic amino acid spin systems except that of one of the Tyr residues are shown. The lower panels show the region of His C(2)H to C(4)H cross-peaks in TOCSY spectra of the hLIF mutants H19A (pH 4.5), H138A (pH 3.4) and H141A (pH 3.4). Note that the different pH values for H138A and H141A give rise to slight differences in chemical shifts and cross-peak intensities. Substitution of either His16 or His19 by Ala caused a reduction in cross-peak intensity for the other, suggestive of an interaction between these residues. All spectra were recorded at 600 MHz.

demonstrated by the lack of change in the chemical shifts of the cross-peaks of other aromatic ring protons. Only one cross-peak (Tyr-c in Fig. 3) showed any significant movement with respect to pH. It appears to be monitoring the titration of a carboxyl group with a pK_a of 3.4.

4. Discussion

The model of hLIF based on hG-CSF proved to be stable during the simulated annealing and molecular dynamics refinement process and is consistent with known stereochemical data on proteins. It also accounts very well for the pK_a values of the six His residues, determined from the NMR data and identified by the His \rightarrow Ala mutants. The locations of the His residues in the model, and of nearby residues thought to influence their pK_a values, are shown in Fig. 5. His¹⁶ and His¹⁹ have slightly lowered pK_a values compared with those of 6.5–6.6 found in denatured proteins [30], indicating a slight destabilisation of their protonated forms. In the model the loop formed in the N-terminal tail brings His¹⁶ into line with the axis of the A helix and Arg¹⁵ close to His¹⁶ and His¹⁹ (which is at the N-terminus of helix A). The influence of the positive charges of the helix dipole and the

arginine side chain would be expected to lower the pK_a of these histidine residues [28,30], as was observed. His⁷¹ shows two independent titrations, the first reflecting a carboxyl titration (pK_a 2.6) and the second representing the imidazolium pK_a , which is slightly elevated at 7.4. This is also consistent with the model, where formation of the C-terminal helix in the AB loop brings His⁷¹ into close proximity to Glu⁷⁶. The negatively charged environment could stabilize the protonated form of His⁷¹, thus elevating its pK_a . His¹¹² has a significantly lowered pK_a value, suggesting that its imidazolium form is greatly destabilised. This is probably due to the cluster of positively charged lysine residues (102, 114, 153, 158, 159) around His¹¹² in the model, two of which come from the D1 motif [3]. His¹¹² has been shown to be involved in the binding of hLIF to its receptor α -subunit [9] and its low pK_a may be important for receptor binding. Finally, His¹³⁸ and His¹⁴¹ show essentially normal pK_a values, indicating that they are not affected by nearby charged groups and are probably exposed to solvent. In the model they are located in the CD loop but His¹³⁸ is close to Lys¹³⁶ and His¹⁴¹ to Asp¹⁴³. The fact that these nearby charged sidechains do not significantly affect the pK_a values of the histidines may reflect the possibility that this region of the structure is flexible in solution. In the X-ray structures of GH [14], IL-2 [31] and G-CSF [15,32] part of the CD loop cannot be seen in the electron density, suggesting that this is a mobile region in these molecules. Analysis of the last 22 ps of the dynamics trajectory of the hLIF model in water indicates that the N-terminus and the N-terminal region of the CD loop are amongst the most variable parts of the structure.

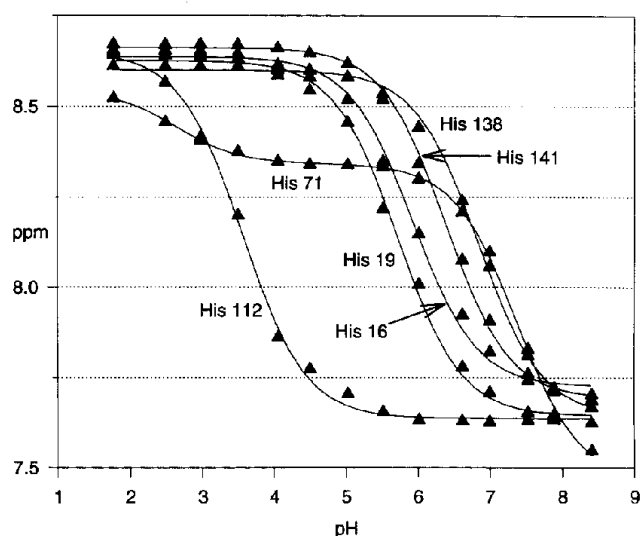


Fig. 4. pH titration curves for the six His residues of hLIF in $^2\text{H}_2\text{O}$ at 26°C . The pK_a values and, in parentheses, the C(2)H chemical shifts in ppm at the acidic and basic extremes, respectively, of the imidazolium titration were: His¹⁶, 5.9 (8.64, 7.72); His¹⁹, 5.7 (8.63, 7.64); His⁷¹, 7.4 (8.34, 7.46); His¹¹², 3.6 (8.65, 7.64); His¹³⁸, 6.8 (8.60, 7.64); His¹⁴¹, 6.4 (8.66, 7.69).

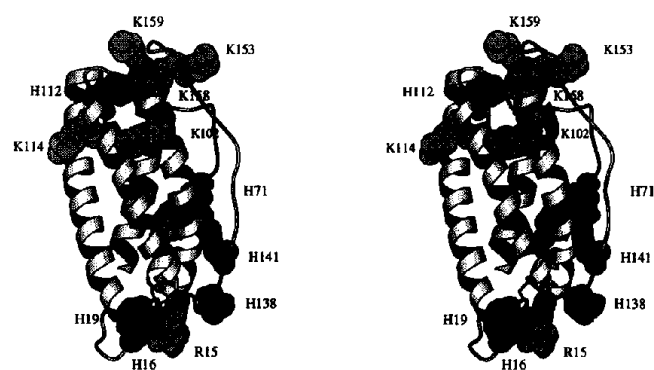


Fig. 5. Stereo views, generated using MOLSCRIPT [36], of the model of hLIF showing the His residues (dark shading) and other residues which may influence the His pK_a values (light shading). For clarity, the sidechain of Glu⁷⁶ is not labelled. This view is related to that in Fig. 2 by an approximately 180° rotation around the vertical axis.

Asparagines 9, 34, 63, 73, 96 and 116 have been identified as *N*-glycosylation sites in hLIF [33]. These residues are exposed in the model except for Asn⁹ and Asn⁹⁶. Asn⁹ is in the long N-terminal tail preceding the first disulfide bridge, a region which is expected to be mobile and thus unreliable in the model. Asn⁹⁶ is partially occluded by Tyr¹⁴⁶ in the CD loop, which, as noted above, is also likely to be mobile.

The two small helices in the AB loop of the hLIF model may be a conserved feature of the long-chain group of cytokines as they were either preserved or formed in the refinement process. Furthermore, formation of the more C-terminal of these helices allows the double titration of His⁷¹ to be explained. These helices are also present in the soluble form of an hGH mutant [17]. In hGH they are involved in receptor binding [14] and, if this is indeed a conserved feature, they may have similar functions in G-CSF and LIF. A study of mouse–human chimeras of LIF [9] identified His¹¹², Ser¹¹³, and some residues of the BC and CD loops as being involved in the binding of hLIF to its receptor. It is likely that residues in the D1 motif are involved in receptor binding as they form a contiguous surface on the model with His¹¹² and Ser¹¹³ and are involved in the low pK_a of His¹¹². If the D1 motif is involved in receptor binding, its conservation in G-CSF, CNTF and OSM would be consistent with a similar role in those molecules.

The NMR experiments described in this paper provide information on the pH stability and ionization behaviour of hLIF. Although His⁷¹ and His¹¹² participate in interactions which perturb their pK_a values significantly, in the case of His¹¹² by 3 pH units from the expected value, the overall structure of hLIF as monitored by 2D ¹H NMR is stable over the pH range 1.8–8.4. This implies that the electrostatic interactions involving these His residues, while important locally, do not play a major role in stabilizing the native structure.

In this study we have described a model of LIF which

is supported by the histidine pK_a values observed by ¹H NMR and is consistent with the known *N*-glycosylation sites and receptor binding site data. The atomic coordinates of the model are available on request.

Acknowledgements: We thank Andrew Wolfe, Greg Kemp, Chris Mardon and Jian-Guo Zhang for protein expression, Richard Mann for protein purification, Maria Harrison-Smith for technical assistance, Don Metcalf, Sandra Mifsud and Ladina Di Rago for biological assays, and Kerry Higgins for assistance with NMR. D.K.S. is supported by a Victorian Education Foundation scholarship. This is a contribution from the Cooperative Research Centre for Cellular Growth Factors, funded by the Australian Cooperative Research Centres Program. Additional support was from the National Institutes of Health, Bethesda, MD (Grant CA22556), and AMRAD Corporation.

References

- [1] Hilton, D.J. (1992) *Trends Biochem. Sci.* 17, 72–76.
- [2] Rose, T.M. and Bruce, A.G. (1991) *Proc. Natl. Acad. Sci. USA* 88, 8641–8645.
- [3] Bazan, J.F. (1991) *Neuron* 7, 197–208.
- [4] Gearing, D.P., Comeau, M.R., Friend, D.J., Gimpel, S.D., Thut, C.J., McGourty, J., Brasher, K.K., King, J.A., Gillis, S., Mosley, B., Ziegler, S.F. and Cosman, D. (1992) *Science* 255, 1434–1437.
- [5] Kishimoto, T., Taga, T. and Akira, S. (1994) *Cell* 76, 253–262.
- [6] Sprang, S.R. and Bazan, J.F. (1993) *Curr. Opin. Struct. Biol.* 3, 815–827.
- [7] Betzel, C., Visanji, M., Dauter, Z., Fourme, R., Weber, W., Marnitz, U., Boone, T., Pope, J., Miller, J., Hawkins, N. and Samal, B. (1993) *FEBS Lett.* 336, 236–238.
- [8] Clore, G.M. and Gronenborn, A.M. (1991) *Progr. NMR Spectrosc.* 23, 43–92.
- [9] Owczarek, C.M., Layton, M.J., Metcalf, D., Lock, P., Willson, T.A., Gough, N.M. and Nicola, N.A. (1993) *EMBO J.* 12, 3487–3495.
- [10] Minasian, E. and Nicola, N.A. (1992) *Protein Seq. Data Anal.* 5, 57–64.
- [11] Smith, D.K. (1986) MA Thesis, University of Canberra, ACT, Australia.
- [12] Hogeweg, P. and Hesper, B. (1984) *J. Mol. Evol.* 20, 175–186.
- [13] Parry, D.A.D., Minasian, E. and Leach, S.J. (1991) *J. Mol. Recog.* 4, 63–75.
- [14] de Vos, A.M., Ultsch, M. and Kossiakoff, A.A. (1992) *Science* 255, 306–312.
- [15] Hill, C.P., Osslund, T.D. and Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. USA* 90, 5167–5171.
- [16] Abdel-Meguid, S.S., Shieh, H.-S., Smith, W.W., Dayringer, H.E., Violand, B.N. and Bente, L.A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6434–6437.
- [17] Ultsch, M.H., Somers, W., Kossiakoff, A.A. and de Vos, A.M. (1994) *J. Mol. Biol.* 236, 286–299.
- [18] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* 112, 535–542.
- [19] Guenot, J. and Kollman, P.A. (1992) *Protein Sci.* 1, 1185–1205.
- [20] Brünger, A.T. (1992) X-PLOR, A System for Crystallography and NMR, Yale Univ., New Haven, CT.
- [21] Nicola, N.A., Cross, B. and Simpson, R.J. (1993) *Biochem. Biophys. Res. Commun.* 190, 20–26.
- [22] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Cryst.* 26, 283–291.
- [23] van Geet, A.L. (1970) *Anal. Chem.* 42, 679–680.
- [24] Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* 113, 967–974.

- [25] Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.* 53, 521–528.
- [26] Rucker, S.P. and Shaka, A.J. (1989) *Mol. Phys.* 68, 509–517.
- [27] Shrager, R.I., Cohen, J.S., Heller, S.R., Sachs, D.H. and Schechter, A.N. (1972) *Biochemistry* 11, 541–547.
- [28] Loewenthal, R., Sancho, J. and Fersht, A.R. (1991) *Biochemistry* 30, 6775–6779.
- [29] Harris, N.L., Presnell, S.R. and Cohen, F.E. (1994) *J. Mol. Biol.* 236, 1356–1368.
- [30] Loewenthal, R., Sancho, J. and Fersht, A.R. (1992) *J. Mol. Biol.* 224, 759–770.
- [31] McKay, D.B. (1992) *Science* 257, 412–413.
- [32] Lovejoy, B., Cascio, D. and Eisenberg, D. (1993) *J. Mol. Biol.* 234, 640–653.
- [33] Schmelzer, C.H., Harris, R.J., Butler, D., Yedinak, C.M., Wagner, K.L. and Burton, L.E. (1993) *Arch. Biochem. Biophys.* 302, 484–489.
- [34] Bacon, D.J. and Anderson, W.F. (1986) *J. Mol. Biol.* 191, 153–161.
- [35] Kabsch, W. and Sander, C. (1983) *Biopolymers* 22, 2577–2637.
- [36] Kraulis, P.J. (1991) *J. Appl. Cryst.* 24, 946–950.