Journal of Biomolecular NMR, **12:** 89–107, 1998.
KLUWER/ESCOM

Complete assignment of ^{1}H, ^{13}C and ^{15}N chemical shifts for bovine β-lactoglobulin: Secondary structure and topology of the native state is retained in a partially unfolded form

Stanislava Uhrínová[a], Dušan Uhrín[a], Helen Denton[b,*], Mark Smith[c], Lindsay Sawyer[b] and Paul N. Barlow[a,**]

[a]*The Edinburgh Centre for Protein Technology, Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, U.K.;* [b]*Structural Biochemistry Group, Swann Building, Mayfield Road, Edinburgh EH9 3JR, U.K.;* [c]*New Zealand Dairy Research Institute, Private Bag 11 029 Pamerston North, New Zealand*

Received 14 November 1997; Accepted 17 February 1998

Key words: β-lactoglobulin; methyl-containing residues; modified HCCH-TOCSY; partially folded form; side-chain assignment

Abstract

Although β-lactoglobulin (β-LG) has been studied extensively for more than 50 years, its physical properties in solution are not yet understood fully in terms of its three-dimensional (3D) structure. For example, despite a recent high-resolution crystal structure, it is still not clear why the two common variants of bovine β-LG which differ by just two residues have different aggregation properties during milk processing. We have conducted solution-state NMR studies on a recombinant form of the A variant of β-LG at low pH conditions where the protein is partially unfolded and exists as a monomer rather than a dimer. Using a ^{13}C,^{15}N-labelled sample, expressed in *Pichia pastoris*, we have employed the standard combination of 3D heteronuclear NMR techniques to obtain near complete assignments of proton, carbon and nitrogen resonances. Using a novel pulse sequence we were able to obtain additional assignments, in particular those of methyl groups in residues preceding proline within the sequence. From chemical shifts and on the basis of inter-residue NOEs, we have inferred the secondary structure and topology of monomeric β-LG A. It includes eight antiparallel β-strands arranged in a barrel, flanked by an α-helix, which is typical of a member of the lipocalin family. A detailed comparison with the crystal structure of the dimeric form (for a mixture of A and B variants) at pH 6.5 reveals a close resemblance in both secondary structure and overall topology. Both forms have a ninth β-strand which, at the higher pH, forms part of the dimer interface. These studies represent the first full NMR assignment of β-LG and will form the basis for a complete characterisation of the solution structure and dynamics of this protein and its variants.

Abbreviations: β-LG, β-lactoglobulin; CSI, chemical shift index; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy

Introduction

β-Lactoglobulin (β-LG) is a small (18350 Da) soluble, acid-stable protein found in the whey of milk of ruminants and other species (Hambling et al., 1992). Despite its abundance, its function remains something of a mystery. It is a member of the lipocalin family (Flower et al., 1993) whose structurally characterised members include odorant binding protein (Tegoni et al., 1996), epididymal retinoic acid binding protein (Newcomer et al., 1984), retinol binding

*Present address: Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, U.K.
**To whom correspondence should be addressed.

protein (Cowan et al., 1990), bilin-binding protein (Huber et al., 1987), major urinary protein from mouse (Bocskei et al., 1992) and β-LG itself (Brownlow et al., 1997). Like other members of the family, β-LG binds tightly small hydrophobic molecules such as fatty acids for which it might act as a transporter (Perez and Calvo, 1995).

Bovine β-LG is dimeric above pH 3.5, but exists as a monomer at low pH and low salt concentration. It has been reported that under these conditions the protein is locally unfolded and may represent a kinetic intermediate in the folding pathway (Molinari et al., 1996). There are two major variants (A and B) in bovine milk differing by two amino acids (Asp^{64} in variant A is Gly in B; Val^{118} in A is Ala in B) (Hambling et al., 1992). A recent, high-resolution crystal structure (of an A/B mixture), at pH 6.5, confirmed that the protein is a typical lipocalin, consisting of a β-barrel of eight antiparallel β-strands with $(+1)_8$ topology together with a flanking α-helix (Brownlow et al., 1997). The barrel has the shape of a flattened cone. One curved face is formed from a twisted sheet containing the N-terminal half of strand A, together with B, C and D. The other curved face is a second twisted sheet containing strands E, F, G, H and the C-terminal half of strand A. Strand A has a 90° bend at its midpoint, Leu^{22}. This structure closely resembles that of RBP but a ninth strand (I) is present in β-LG which extends the FGHA sheet. A short 3_{10}-helix precedes strand A and a second 3_{10}-helical turn lies in the long AB loop and forms part of the dimer interface along with strand I. β-turns link B,C, D, E and F. Strands F and G are connected via a γ-turn. Short 3_{10}-helical turns are found between strands G and H, and within the C-terminal region. A three-turn α-helix lies in the sequence between strands H and I. There are two disulphides (Cys^{66}–Cys^{160}, and Cys^{106}–Cys^{119}) and a free thiol (Cys^{121}).

As a consequence both of its ready availability and its commercial importance in milk processing, β-LG has been the subject of extensive physicochemical studies over the last 60 years (e.g. Tanford et al. (1959), reviewed in Hambling et al. (1992)). These have revealed significant differences between the A and B variants in terms of their solubility, calorimetric parameters, and gelation and aggregation properties (Jakob and Puhan, 1992). The two variants also differ in their propensity for disulphide interchange (involving Cys^{106}–Cys^{121}) which is an important factor in the thiol-mediated, heat-induced denaturation of milk. However, it is difficult to interpret these differences in terms of the high-resolution protein crystal structure. For example, the Asp^{64}→Gly substitution is in an ill-defined loop between the third and fourth β-strands. The substitution of Val^{118} of A by Ala in B would appear to involve the loss of two methyl groups from a hydrophobic environment but according to differential scanning calorimetry and proteolysis measurements, the B variant is more stable than the A variant; however, studies of thermal aggregation and urea denaturation indicate the opposite to be true (see the review by Fox (1995)). A better understanding of the relationship between sequence, conformation and solution properties would almost certainly follow from detailed NMR studies. Previous NMR work has been hampered by the lack of a suitable isotopically labelled sample and only partial resonance assignments based on homonuclear studies have been possible (Molinari et al., 1996). This study was carried out at pH 2. The protein is monomeric under these conditions and locally unfolded (Molinari et al., 1996). The homonuclear studies revealed the presence of a structured 'β-sheet core' together with a more disordered region giving rise to broad overlapping peaks centred at the random coil position. A more recent study from the same laboratory (Ragona et al., 1997), also at pH 2, reported the presence of a buried cluster of 11 hydrophobic residues which was stable even at 323 K.

We now report the first complete resonance assignment for β-LG (variant A). This was achieved at pH 2.65, using a double (^{13}C,^{15}N-labelled) sample and a range of established and novel NMR experiments. Despite the fact that under these conditions the protein is known to be partially folded, we show that its secondary structure and topology are highly similar to that of the high-pH native dimeric form.

Methods

Sample preparation and characterisation

Bovine β-LG (variant A) was expressed as a recombinant protein from a genetically modified strain of *Pichia pastoris* (kindly made available by Professor Carl Batt) as described elsewhere (Kim et al., 1997). Purification of β-LG and preparation of ^{15}N- and ^{15}N,^{13}C-labelled samples of the protein will be described in full elsewhere. In summary, cells were grown on minimal media supplemented with 1% ^{15}N-ammonium sulphate and 1% ^{13}C-glucose and expression was induced with 1% ^{13}C-methanol;

subsequently, protein was purified from the culture supernatant using ion exchange chromatography followed by gel exclusion chromatography. As a consequence of vector construction, the recombinant material was expected to include three extra residues (Glu-Ala-Glu) at its N-terminus. In addition, the first two residues of wild-type β-LG (Leu-Ile) are replaced with Ala-Tyr. The N-terminal sequence of the recombinant protein used in this study should thus be Glu-Ala-Glu-Ala-Tyr-Val-Thr-Gln-Thr-Met-Lys. N-terminal analysis and mass spectrometry revealed that the N-terminus is ragged, a common problem with this expression vector; there are several N-terminal sequences present including Glu-Ala-Tyr-Val-Thr-Gln, Ala-Tyr-Val-Thr-Gln and Tyr-Val-Thr-Gln and Val-Thr-Gln. There was some batch-to-batch variation in the relative quantities of these species. However, for the double-labelled sample, N-terminal analysis and assignment of the NMR spectra confirmed that the full-length, uncleaved sequence was the major (>80%) component.

The NMR sample contained 1 mM protein in 50 mM potassium phosphate adjusted to pH 2.65, in 550 μ of H_2O/D_2O (90/10 v/v). Under these conditions, the sample is monomeric and was judged to be stable for months at 4 °C, and for many days at 37 °C (the temperature at which NMR experiments were run) on the basis of there being little change in appearance of the (^1H,^{15}N) heteronuclear single quantum coherence (HSQC) spectrum. However, the intensity of some unidentified signals gradually increased throughout the first few triple-resonance experiments before stabilising.

NMR data collection and processing

All NMR spectra were recorded at 37 °C on a Varian-INOVA 600 spectrometer equipped with a 5 mm z-gradient triple-resonance probe. A set of triple-resonance experiments was acquired, amongst which the most useful were CBCA(CO)NH, HNCACB (Muhandiram and Kay, 1994), HBHA(CBCACO)NH (Grzesiek and Bax, 1993), HBHA(CBCA)NH (Wang et al., 1994), HNCO (Kay et al., 1994), HN(CA)CO (Engelke and Rüterjans, 1995). Three-dimensional (3D) CBCA(CO)NH and HNCACB spectra were recorded with spectral widths of 8000 Hz in F3 (N^1H), 2015 Hz in F2 (^{15}N) and 10 560 Hz in F1 (^{13}C). The size of the acquired matrix was (t_3,t_2,t_1) 512 × 48 × 64 complex data points. Three-dimensional HBHA(CBCACO)NH and HBHA(CBCA)NH spectra had the same spectral widths in F3 (N^1H) and F2 (^{15}N) dimensions, while the spectral width in F1 (^1H) was 2940 Hz using the folding of CH_3 protons. The number of t_1 and t_2 increments was 32 and 48, respectively, in both experiments. The carbonyl spectral widths (F1) in HNCO and HN(CA)CO experiments were set to 2225 Hz, with ^1H and ^{15}N spectral widths set as above. The size of the acquired matrices was 512 × 32 × 32 complex data points. The aliphatic side-chain chemical shifts were obtained from H(C)(CO)NH total correlation spectroscopy (TOCSY) and (H)C(CO)NH-TOCSY (Grzesiek and Bax, 1992; Grzesiek et al., 1993; Logan et al., 1993) experiments, respectively. In these experiments parameters were used as above, with the exception of the F1 (^1H) spectral width in the H(C)(CO)NH-TOCSY where 7200 Hz was used and the size of the acquired matrix was 512 × 48 × 80 complex data points. All these experiments involved detection of NH protons and the sensitivity-enhanced protocol was used (Kay et al., 1992). Several modifications were made to the original pulse sequences, including combined use of nonselective and shaped ^{13}C pulses, return of the magnetisation of the water protons to the z-axis prior to acquisition without the aid of selective pulses, and elimination of Bloch-Siegert shifts during the t_1 period. These modifications will be discussed elsewhere. A ^{15}N-edited 3D NOESY-HSQC (Sklenář et al., 1993; Mori et al., 1995) was recorded using a 70 ms mixing time. The spectral widths were 5000 Hz in F3 (N^1H), 2015 Hz in F2 (^{15}N) and 7200 Hz in F1 (^1H). The size of the acquired matrix was 512 × 48 × 96. In most of the experiments, the ^1H carrier was placed on the H_2O resonance at 4.644 ppm, the ^{15}N carrier was placed at 117 ppm, and the ^{13}C carrier was set depending on the experiment, at 43 ppm for C^α and C^β carbons, 58 ppm for C^α carbon, 176 ppm for carbonyls and 20 ppm for methyls. The indirectly detected dimensions were referenced as described in Live et al. (1984), and Bax and Subramanian (1986). Data sets were processed using home-written macros within the package FELIX95 (Biosym Technologies/Molecular Simulations, San Diego, CA, U.S.A.). Post-acquisition water signal removal was effected using a sine-bell convolution function with a 10-point window size and linear prediction of tails (according to the FELIX95 manual, Biosym Technologies/Molecular Simulations). In general, a Lorentzian-to-Gaussian window function was used for resolution enhancement in the acquired dimension, followed by zero filling. Phase-shifted-sine bell window functions were used in the indirectly

detected dimensions. In constant-time experiments, mirror image linear prediction (Zhu and Bax, 1990) was used to extrapolate the FID by up to 50%. The FIDs were then zero-filled prior to Fourier transform. Final 3D matrices typically consisted of 1024 × 128 × 128 real data points. Transformed data sets were converted into strip plots within FELIX95, then displayed and analysed using the program XEASY (Bartels et al., 1995). The chemical shift index (CSI) was calculated using the CSI program (Wishart and Sykes, 1994).

Results and Discussion

Backbone assignments

Figure 1 shows the $^{1}H,^{15}N$ HSQC spectrum of β-LG. Although dispersion is, in general, as expected for a protein containing a high proportion of β-sheet, a number of low-intensity peaks appear in an overlapped region in the centre of the spectrum. These peaks are not a consequence of the recombinant nature of the protein since they were also detected in a natural abundance $^{1}H,^{15}N$ spectrum of wild-type bovine β-LG obtained commercially (not shown). The intensities of these minor signals is about one-quarter of the major ones. It was possible to assign about 20 out of the 40 minor peaks to segments within a minor form (or forms) of the protein, including Thr^{4}–Leu^{10}, Ala^{16}–Trp^{19}, Ala^{26}–Asp^{29}, and Lys^{138}–Leu^{143}. It was impossible to extend these assignments due to low signal-to-noise and extensive overlap. Over time, some peaks, including the broad ones, became more intense but the overall appearance of the $^{1}H,^{15}N$ HSQC did not alter.

Signals due to the NH_2 groups of asparagines and glutamines were unequivocally identified in a two-dimensional (2D) heteronuclear triple quantum edited experiment (Schmidt and Rüterjans, 1990) and are shown (indicated by a horizontal dotted line and/or a label) in Figure 1. Sequential assignments were achieved on a uniformly $^{15}N,^{13}C$-labelled sample using three independent connectivity pathways derived from the six 3D triple-resonance experiments: CBCA(CO)NH, HNCACB, HBHA(CBCACO)NH, HBHA(CBCA)NH, HNCO and HN(CA)CO. The CBCA(CO)NH and HBHA(CBCACO)NH spectra correlate the amide proton and nitrogen chemical shifts of one residue with those of α and β carbons or protons of the preceding residue, while the HNCACB and HBHA(CBCA)NH spectra additionally provide corresponding intra-residual correlations. Sequential assignments obtained from these experiments were verified using the HNCO and HN(CA)CO experiments. The carbonyl chemical shifts of the $i-1$ residues were matched with the intra-residue correlations obtained in the HN(CA)CO experiment. The sequential assignment of the first 14 residues obtained in this manner is shown in Figure 2. The strip number corresponds to the position of the residue in the amino acid sequence and the sequential connectivities are depicted with horizontal lines.

Despite the relatively good dispersion of H^{α} and C^{α} signals, which is characteristic of predominantly β-structures, several difficulties were encountered during the backbone assignment. The backbone amide of Leu^{22} was not identified, however proton and side chain resonances of Leu^{22} were assigned based on correlations with the NH-group of Ala^{23}. No assignments at all could be made for Glu^{157}, Glu^{158} or Ile^{162} at the C-terminus of the molecule. Cross peaks for Trp^{61} and Ala^{67} were weak. In total, complete assignment of C^{α}, C^{β}, H^{α} and CO was achieved for 161 out of 165 residues, and all but four of the expected amide nitrogens (and all but four of their attached protons) were assigned. Based on these six experiments we assigned 97% of C^{α}, C^{β} and H^{α} resonances. Due to overlap we assigned only 90% of H^{β} resonances.

Sidechain assignments

Experiments aimed at backbone assignments also furnished assignments for all but three of the expected C^{β}resonances. They also provided all but 16 H^{β} resonance assignments for the first 154 residues beyond which overlap is severe and several backbone assignments are missing. Additional side-chain assignments were obtained from (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY techniques and two novel CC-TOCSY-type experiments. The first two experiments correlate NH protons and nitrogens with the carbons or protons of the preceding residue (see Figure 3).

The new experiments employed in this study were designed to facilitate the assignment of side-chain resonances of residues containing methyl groups. These experiments take advantage of a good separation of methyl cross peaks in proton/carbon correlated spectra of proteins, both from one another and from the signals of CH and CH_2 groups (Figure 4). The initial parts of the proposed pulse sequences (see Figures 5a and b) are similar to those of the (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY experiments (Logan et al.,

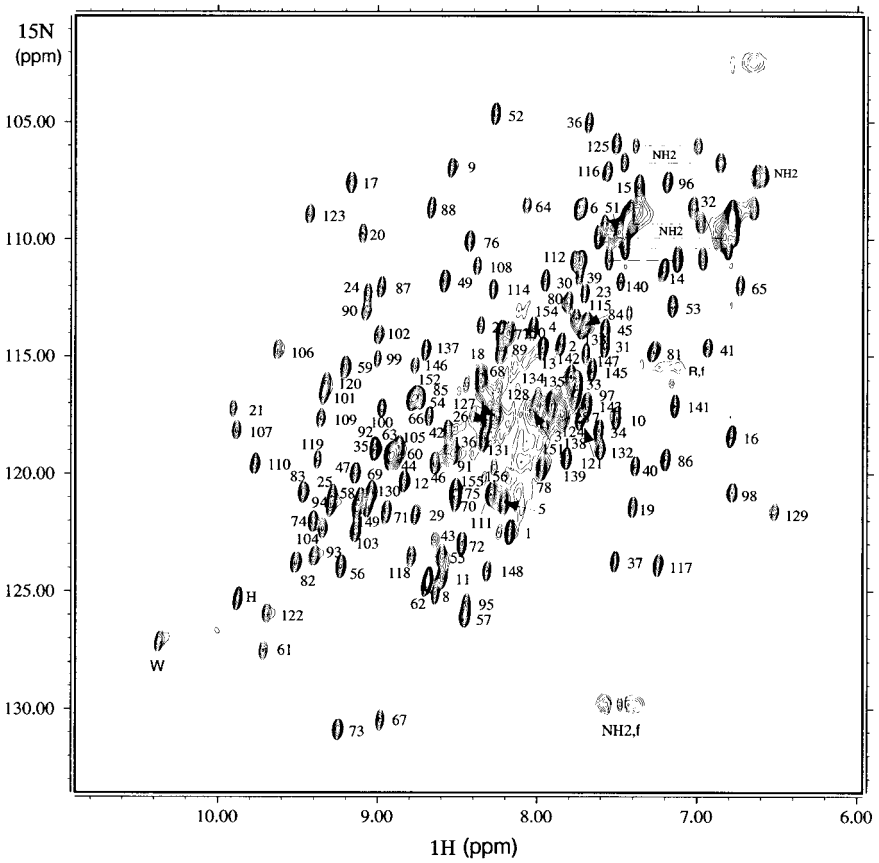

Figure 1. 2D ^{1}H-^{15}N HSQC spectrum of β-LG. The numbers correspond to the position of the residue in the amino acid sequence; 'f' indicates folded signals.

1993; Gardner et al., 1996) and will not be discussed here. The new experiments differ only after the carbon-carbon spin-lock period. Here, the carbon magnetisation is not directed towards NH protons but, instead, the carbon chemical shift is labelled during the constant-time interval and the magnetisation is immediately transferred to protons for detection. The soft-pulse WATERGATE (Piotto et al., 1992) was used during the refocusing interval for water suppression. We refer to these two experiments as CT-(H)CCH-TOCSY and CT-H(C)CH-TOCSY, the first letter not in parentheses indicating the chemical shift of the nucleus which is sampled during the first indirectly detected dimension. The transfer function of the constant-time interval T of the proposed experiments has the form

$$I = n \sin(\pi {}^{1}J_{CH}T) \times \cos^{n-1}(\pi {}^{1}J_{CH}T) \times II \cos(\pi {}^{1}J_{CC}T)$$

where n is equal to 1, 2 and 3 for CH, CH_2 and CH_3 groups, respectively, and the product runs through all the carbon-carbon couplings of the appropriate carbons. Assuming $^{1}J_{CH} = 123$ Hz and $^{1}J_{CC} = 35$ Hz for CH_3 groups, a local maximum of -0.57 is located at $T = 9.6$ ms. At this time the transfer function for CH and CH_2 groups, calculated assuming three and two carbon-carbon interactions respectively, is -0.11 for CH and $+0.20$ for CH_2 groups. As already indicated, only CH_3 groups are of interest in these experiments. Given the high value of the transfer function for these groups together with favourable relaxation properties of methyl carbons, this is a very sensitive experiment. The proposed use of a 9.6 ms constant-time chemical shift labelling period yields better digital resolution than variable-time experiments and, after a mirror image linear prediction, provides very good digital resolution. Nonetheless, the unambiguous interpretation of the spectra relies on sufficient separation of CH_3 cross peaks. Conveniently, CH_3 strips

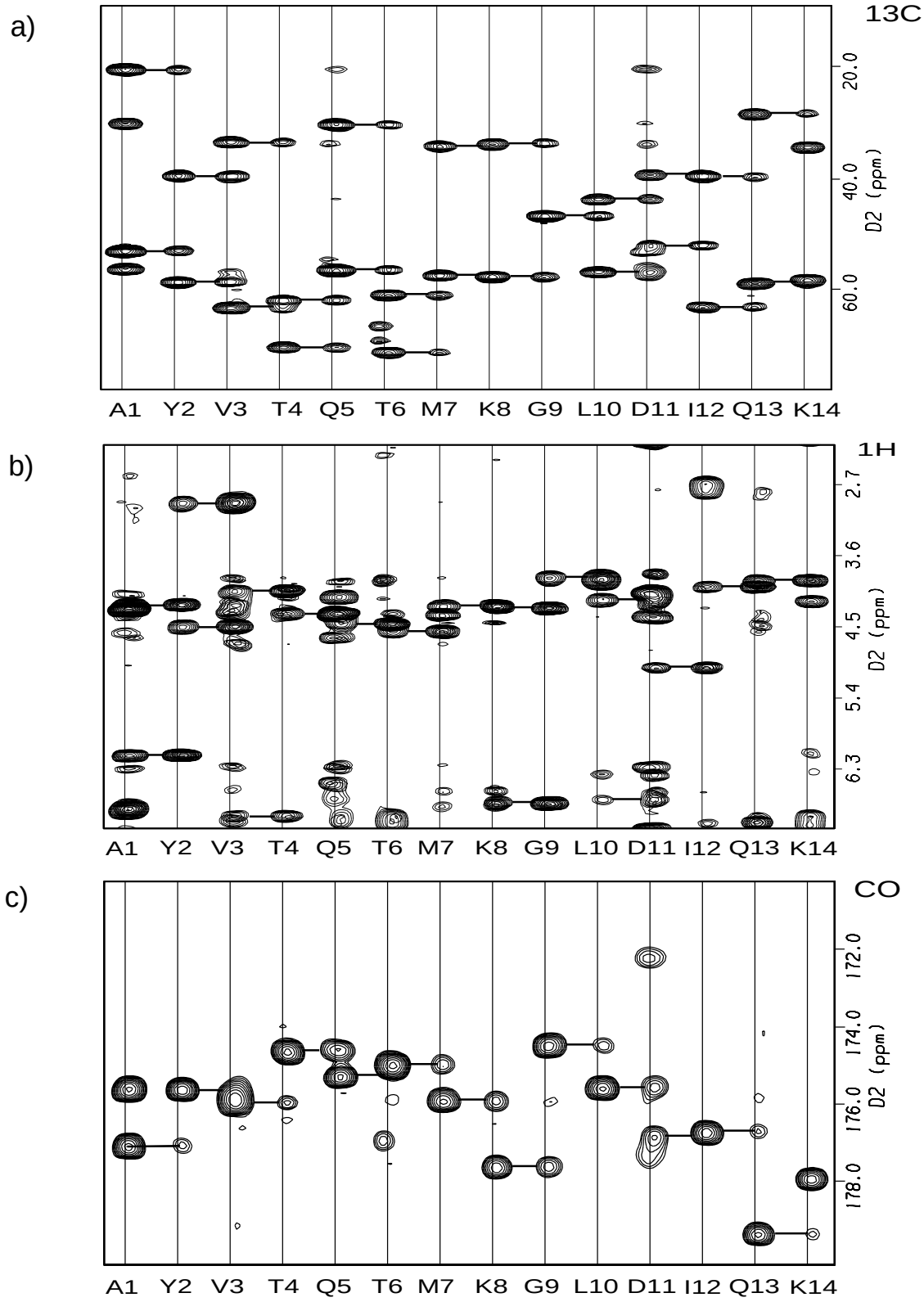

Figure 2. Sequential assignment of the first 14 residues, Ala^{1}-Lys^{14}, of β-LG. The strip number corresponds to the position of the residue in the amino acid sequence. (a), (b) and (c) show $C^{\alpha,\beta}$, $H^{\alpha,\beta}$ and CO cross peaks extracted from HNCACB, HAHB(CACB)NH and HN(CA)CO spectra, respectively. The sequential connectivities in each spectrum are depicted with horizontal lines.

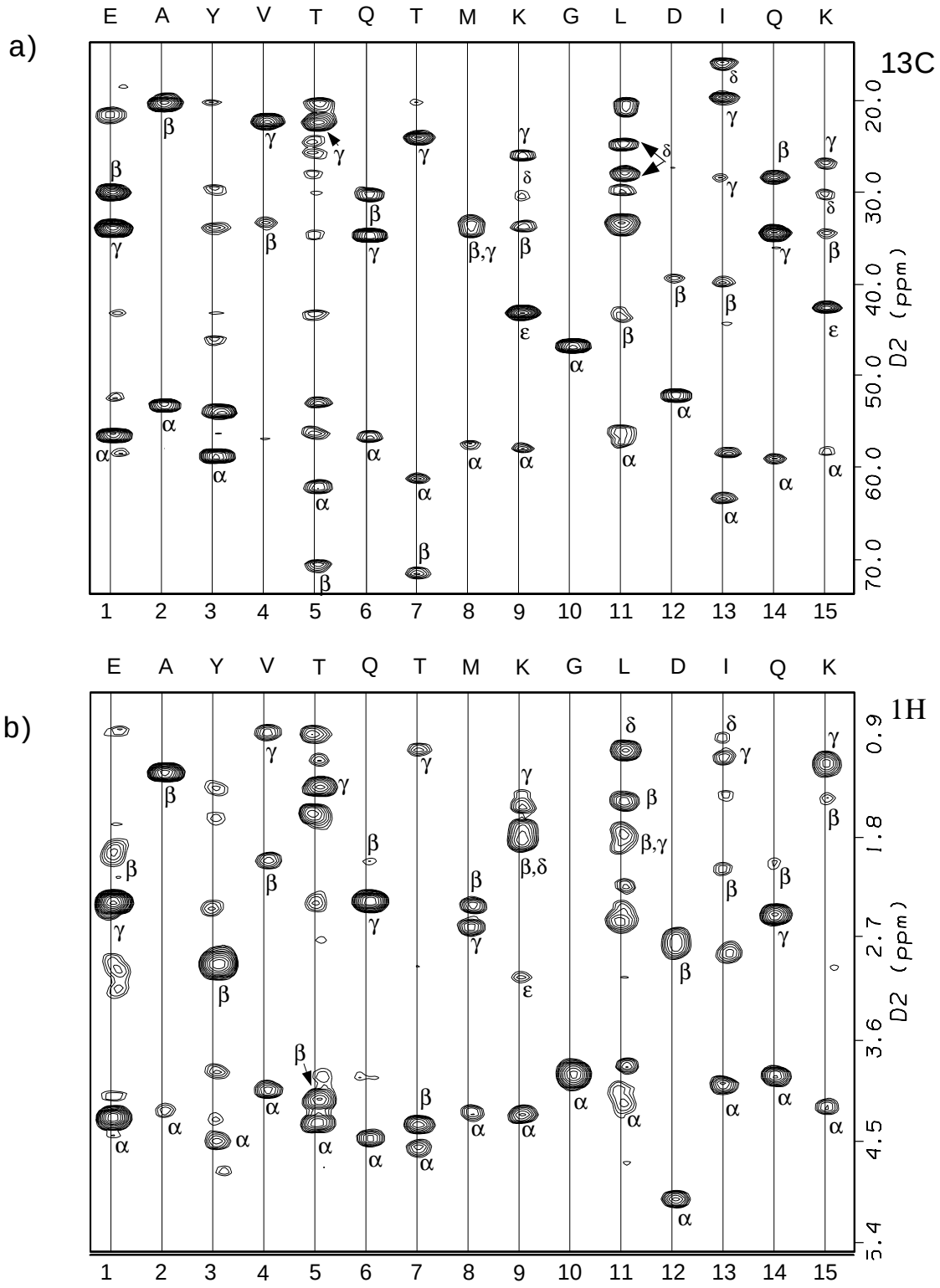

Figure 3. Selected strips from the (a) (H)C(CO)NH-TOCSY and (b) H(C)(CO)NH-TOCSY spectra of β-LG. The strip number corresponds to the position of an amino acid in the sequence; the signals in the strips belong to the n−1 residue indicated above.

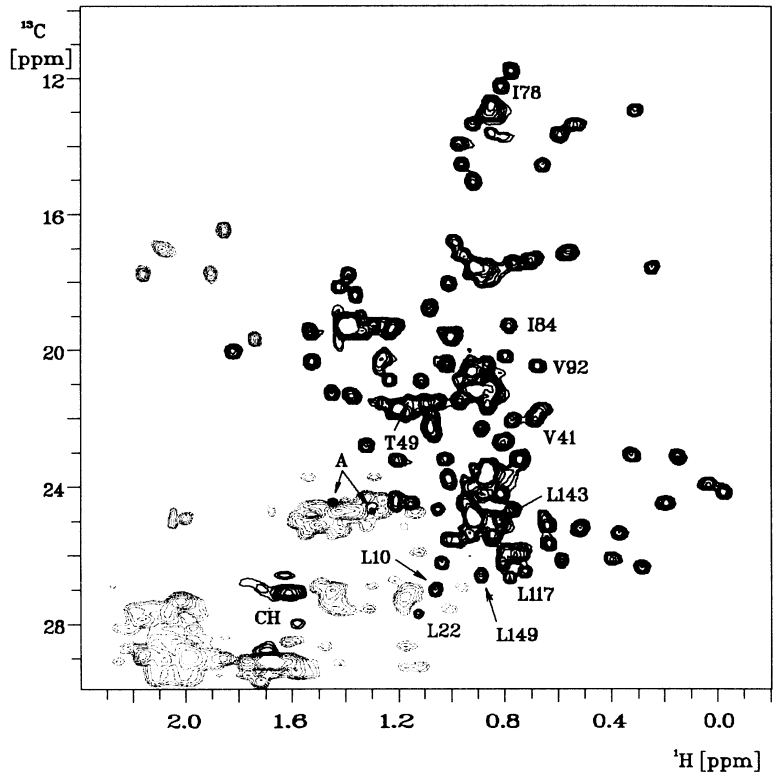

Figure 4. Partial 2D CT-HSQC (Vuister and Bax, 1992) spectrum of β-LG acquired using the constant-time interval of 29.4 ms for labelling of the ^{13}C chemical shifts. Negative cross peaks are plotted in grey. Signals of Ala^{25} and Ala^{26} overlapping with CH_2 signals are indicated by a double arrow. Positive cross peaks of γ-CH groups of leucine are visible at approximately 1.6 and 27 ppm (^{1}H/^{13}C). The strips shown in Figure 6 were taken at positions of the labelled CH_3 cross peaks.

can be extracted from CH_3 carbon planes of the 3D spectra as shown in Figure 6. In the rare event of overlap, CH_3 cross peaks should be easily distinguishable from CH_2 ones by their opposite sign. Out of 165 residues in this β-LG sample, 68 contain a total of 108 methyl groups. Of these, only Ala^{25} and Ala^{26} CH_3 cross peaks overlapped with CH_2 resonances (Figure 4). No attempt was therefore made to select only CH_3 protons during the acquisition. It is conceivable that a multiple-quantum editing (Domke and Leibfritz, 1990) or filtration (Shaw et al., 1997) could be employed to eliminate all but methyl resonances. However, such pulse sequences are likely to have more pulses and a shorter available sampling of carbon chemical shifts.

The new pulse sequences proved useful in three ways. Firstly, there were three instances in β-LG where methyl-containing residues preceded a proline. Because the magnetisation from side chains in (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY experiments is detected on NH protons of the subsequent residue (see Figure 3), side-chain resonances of residues preceding prolines are not accessible in these experiments. However, the side-chain resonances of these residues, namely Thr^{49}, Ile^{78} and Leu^{143}, were identified in CT-H(C)CH-TOCSY and CT-(H)CCH-TOCSY spectra (see Figure 6) based on a comparison of their α and β proton and carbon chemical shifts obtained from HNCACB and HBHA(CBCA)NH experiments used for the backbone assignment. A second circumstance where the new experiments were helpful was the overlap of NH cross peaks in (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY spectra such as illustrated for overlapping NH cross peaks of Leu^{10} and Ala^{34} (see Figure 1). The (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY NH strips contain reso-

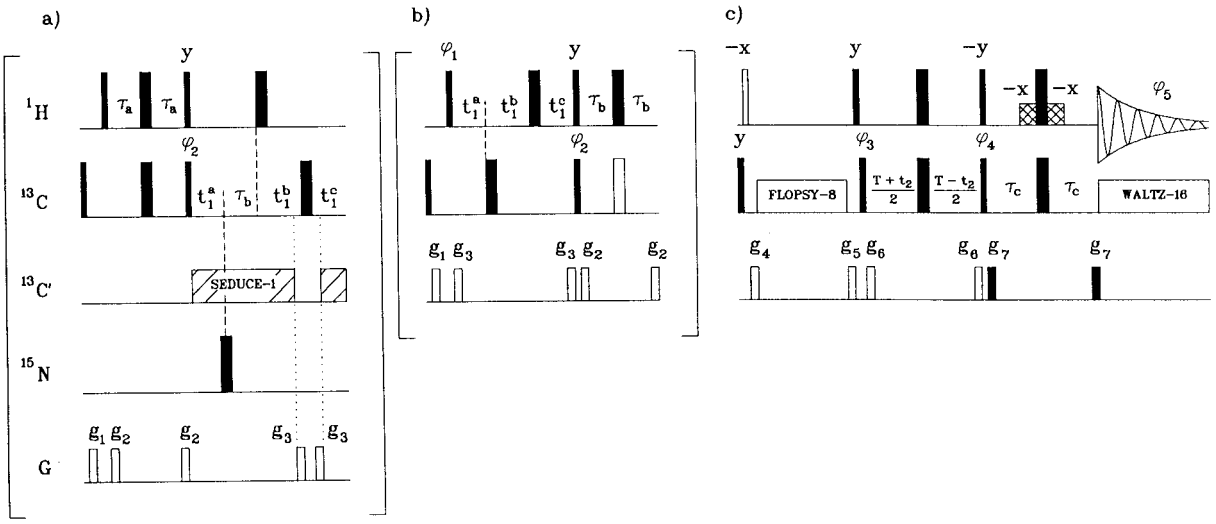

Figure 5. Pulse sequences of CT-(H)CCH-TOCSY (a+c) and CT-H(C)CH-TOCSY experiments (b+c). Narrow and wide bars represent 90° and 180° x-axis pulses unless indicated otherwise. The open rectangle corresponds to a 180° ^{13}C pulse in (b) which is optimised so that its application minimises excitation of carbonyl spins (Kay et al., 1990). Seven cycles of the FLOPSY-8 sequence (Mohebbi and Shaka, 1991) are employed using an 8.45 kHz field. All the other carbon pulses are applied using a field strength of 18.1 kHz. The proton pulses immediately before and after the FLOPSY-8 mixing are inserted solely for the purpose of returning the water magnetisation to the z-axis before acquisition. The first of these two pulses (represented by an open rectangle) was not applied during the acquisition of the imaginary t_1 point of the CT-H(C)CH-TOCSY (b+c) experiment. Rectangular water-selective proton pulses (cross-hatched in the figure) were 1.4 ms long. The SEDUCE-1 (McCoy and Mueller, 1992) was 130 ppm cosine-modulated using 305 μs 90° pulses. The ^{1}H, ^{15}N, ^{13}C and ^{13}C′ (carbonyl) carriers are centred at HOD frequency, 120, 43 and 179 ppm, respectively. The ^{13}C frequency is changed to 20 ppm after the FLOPSY-8 sequence. All gradient pulses are applied along the z-direction with the exception of WATERGATE gradient pulses which are applied along the z and y-axes. The delays employed are as follows: $\tau_a = 1.8$ ms, $\tau_b = 0.9$ ms, $\tau_c = 2.0$ ms. In experiment (a+c) the additional delays are $t_1^a = t_1/2$, $t_1^b = t_1/2 + \delta$ and $t_1^c = \tau_b + \delta$; $\delta = gt3 + 2pw - n\tau_b/(N-1)$, where gt3 is the duration of the gradient g3, pw is the ^{1}H 90° pulse, N is the number of complex points in the ^{13}C dimension, n=0,1,2, …, (N−1) (Grzesiek and Bax, 1993; Logan et al., 1993). In experiment (b+c) the additional delays are $t_1^a = t_1/2 + \tau_a$, $t_1^b = t_1/2 + \delta$ and $t_1^c = \tau_a + \delta$; $\delta = 2pwc - n(\tau_a\text{-}gt3)/(N-1)$, where pwc is the ^{13}C 90° pulse. Quadrature detection in F1 is achieved via States-TPPI (Marion et al., 1989) of φ_2 in (a) and States procedures (States et al., 1982) in (b). The axial peak displacement is not applied in (b) in order to avoid the saturation of the water. In F2 the States-TPPI procedure is applied to the phase φ_4. The phase cycling is $\varphi_1 = x$; $\varphi_2 = x,-x$, $\varphi_3 = 2x,2(-x)$, $\varphi_4 = 4y,4(-y)$ and $\varphi_5 = x,2(-x),x,-x,2x,-x$. The duration and the strengths of gradients are g1 = 0.5 ms, 7 G/cm, g2 = 0.5 ms, 5 G/cm, g3 = 0.2 ms, 15 G/cm, g4 = 0.294 ms, 10 G/cm, g5 = 0.5 ms, 10 G/cm, g6 = 0.1 ms, 16 G/cm, g7 = 0.3 ms , 30 G/cm (z-axis) and 20 G/cm (y-axis).

nances of both residues (see Figure 3) while the CT-H(C)CH-TOCSY and CT-(H)CCH-TOCSY CH_3 strips of Leu[10] show only signals due to this residue. A third advantage accrues from the fact that there is, inherently, proton-carbon correlation information for CH_3 groups in the carbon detected experiments which is not present in the NH detected spectra. This is important for assigning proton and carbon methyl resonances for residues with two CH_3 groups. We note that this can also be obtained from 2D CT-HSQC spectra (Vuister and Bax, 1992) as shown in Figure 4. In the new experiments, resonances of Val, Leu and Ile (which each contain two methyl groups) are present in two different strips, except where both methyls have the same carbon and proton shifts. This allows verification of the assignments and removal of possible ambiguities arising from overlap.

A list of assigned nuclei is presented in Table 1. Overall, using the set of experiments described above, approximately 98% of the carbons and 95% of the protons in aliphatic side chains were assigned. The ϵ-CH_3 groups of the four methionines were not assigned. Partial assignment of aromatic protons came from 2D experiments correlating the C^{β} atom of aromatic amino acids with their δ and ϵ protons (Yamazaki et al., 1993) and from the ^{13}C-edited NOESY experiments (data not shown). This analysis led to δ proton assignments for three out of four Phe residues and four out of five tyrosines. Wherever possible, we compared proton shifts with the very limited assignments available (Ragona et al., 1997) – the agreement was ±0.02 ppm.

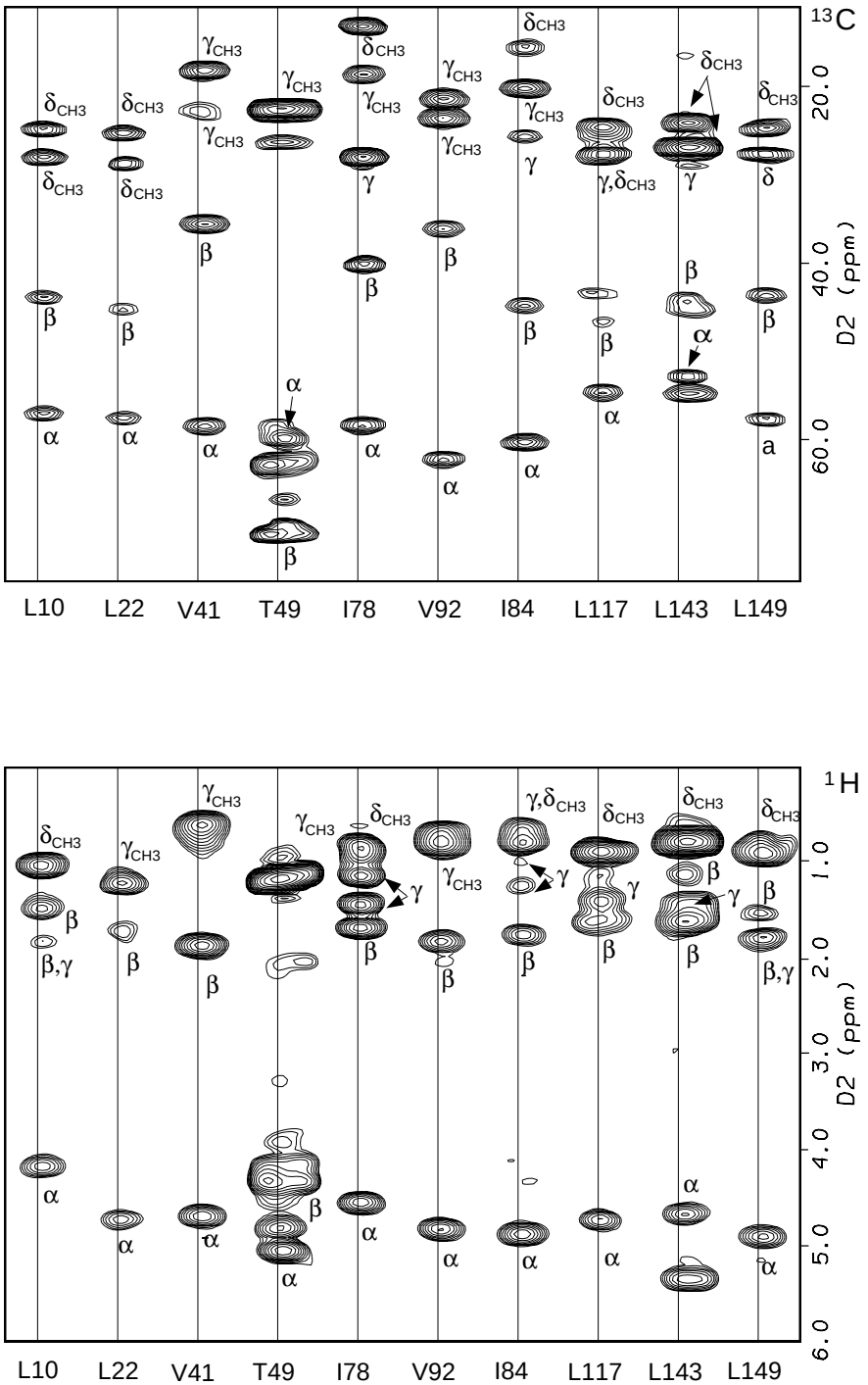

Figure 6. Selected (see Figure 4) strips from (a) CT-(H)CCH-TOCSY and (b) CT-H(C)CH-TOCSY spectra of β-LG acquired using the pulse sequences of Figure 5. The number of t_1 and t_2 increments were 64 and 56 in both experiments. The respective carbon and proton spectral widths in F1 were 10 560 and 6200 Hz. The F2 carbon spectral width was 6200 Hz. The number of scans per FID was 8 and the total acquisition time was 34 h per spectrum.

Table 1. Chemical shift assignments[a] for β-LG at pH 2.4 and 37 °C.

Residue	HN	N	CO	C^{α}	C^{β}	H^{α}	H^{β}	C-other	H-other
E-3			172.06	55.59	29.25	4.09	2.20	32.79	2.55γ
A-2	8.66	124.55	177.21	53.03	19.74	4.39	1.40		
E-1	8.29		175.50	55.96	29.50	4.35	2.41,1.97	33.46	
A1	8.15	122.48	176.97	52.73	19.86	4.26	1.26		
Y2	7.84	114.52	175.51	58.16	38.95	4.51	2.96		7.03δ
V3	7.93	117.85	175.76	63.16	32.91	4.08	2.03	21.94	0.88γ
T4	8.01	113.50	174.52	61.69	70.27	4.34	4.14	22.20	1.09γ
Q5	8.19	121.17	175.18	56.12	29.79	4.48	2.04	34.25	2.39γ
T6	7.70	108.71	174.88	60.75	70.96	4.54	4.36	23.61	1.04γ
M7	7.72	117.79	175.81	57.23	33.73	4.26	1.69,1.90	32.87	2.42γ,2.63γ
K8	8.61	125.00	177.54	57.31	33.12	4.27	1.85	25.49,29.69,42.7	1.51γ,1.77δ,3.05ϵ
G9	8.50	106.92	174.38	46.30		3.90			
L10	7.49	117.64	175.47	56.55	43.37	4.17	1.49,1.81	24.28,27.40	1.83γ,1.06δ,1.05δ
D11	8.58	124.47	176.76	51.59	38.98	5.04	2.76		
I12	8.81	120.37	176.64	62.89	39.39	4.02	2.10	27.90,19.24,15.46	1.47γ,1.09δ,0.91δ
Q13	7.97	114.73	179.30	58.58	27.75	3.93	2.02,2.14	33.97	2.51
K14	7.21	111.37	177.86	57.76	33.84	4.19	1.22,1.48	26.39,29.74,42.18	1.16
V15	7.36	107.83	174.14	61.52	31.79	4.81	2.57	20.03,22.90	1.06γ,0.99γ
A16	6.78	118.51	178.70	53.18	20.71	4.25	1.54		
G17	9.16	107.61	172.97	44.84		4.61,3.94			
T18	8.33	115.80	173.80	64.6	70.27	4.40	3.91	22.50	1.10γ
W19	7.40	121.47	172.39	55.95	37.51	4.77	3.24,2.22		
Y20	9.01	109.82	177.80	57.23	42.48	5.16	2.81	133.6δ	7.09δ,6.79ϵ
S21	9.87	117.32	173.03	60.64	63.54	4.84			
L22			174.59	56.98	44.80	4.72	1.71,1.91	28.24,24.61	1.17δ
A23	7.70	112.39	175.92	51.14	23.70	5.21	1.21		
M24	9.05	112.39	173.13	54.92	37.76	5.73	1.84,2.37	32.36	2.53γ
A25	9.28	120.97	174.58	52.00	25.16	5.28	1.31		
A26	8.29	117.33	177.64	51.74	24.98	5.42	1.44		
S27	8.35	113.75	174.18	60.75	64.87	4.82	3.98,4.16		
D28	7.83	117.34	176.32	52.00	40.25	5.08	2.93,2.57		
I29	8.75	121.80	177.70	66.24	39.05	3.67	1.87	29.02,18.55,14.35	1.15γ,0.99δ
S30	7.94	111.84	175.17	60.75	63.24	4.23	3.95		
L31	7.57	114.68	177.23	57.15	41.88	4.18	1.41,1.76	25.93,22.15	1.76γ,0.95δ
L32	7.03	108.72	177.17	54.80	48.75	4.65	1.31,1.64	26.30,24.95	1.64γ,0.77δ
D33	7.77	115.99	173.78	56.54	38.07	4.27	3.12,2.83		
A34	7.61	118.22	177.92	52.03	21.67	4.57	1.46		
Q35	9.01	119.06	176.69	59.43	28.81	4.12	1.89	33.19	2.11γ
S36	7.67	105.09	174.43	57.50	64.06	4.85	4.26,3.95		
A37	7.50	123.77	177.77	52.15	18.68	4.48	1.44		
P38			176.69	66.19	32.74	4.27	2.42,1.94	28.43,50.94	
L39	7.73	111.62	175.84	52.55	41.48	4.67	1.54,2.08	28.43,25.79,24.05	1.88γ,1.05δ
R40	7.38	119.70	172.54	55.38	29.45	3.97	2.24,1.43	28.36,43.93	1.73γ,1.25γ,3.29δ
V41	6.94	114.67	173.07	57.81	35.01	4.69	1.86	22.51,17.56	0.70γ
Y42	8.56	118.23	176.30	55.57	39.04	4.85	3.14,2.75		
V43	8.63	122.87	174.51	64.62	32.06	3.79		22.36	
E44	8.91	119.49	176.41	55.49	32.43	5.11	1.92,2.04	33.57	2.38γ
E45	7.57	113.89	173.88	55.17	33.69	5.24	2.27	33.82	2.99γ
L46	8.63	119.65	176.37	53.91	44.40	5.16	2.11,1.06	27.96,25.32	1.41γ,0.23δ,0.55δ
K47	9.13	120.07	173.36	53.68	35.64	5.13	1.63,1.87	29.45,40.61	1.35γ,1.69δ,2.92ϵ

Table 1 continued.

Residue	HN	N	CO	C^{α}	C^{β}	H^{α}	H^{β}	C-other	H-other
P48			176.65	62.48	32.81	5.21	2.05	28.21,51.14	2.21γ,3.30δ,3.99δ
T49	8.57	111.82	175.84	59.33	69.02	5.07	4.82	21.89	1.17γ
P50			177.67	65.88	32.31	4.38	2.47,1.91	28.47,51.39	2.09γ,3.91δ
E51	7.49	109.77	176.13	56.62	28.47	4.41	2.27,1.98	34.58	
G52	8.25	104.70	174.51	46.04		4.31,3.62			
D53	7.15	112.90	173.38	53.35	40.06	5.14	3.11,2.73		
L54	8.77	116.99	174.26	54.42	46.54	4.66	1.13,0.65	26.45,23.93	0.96γ,0.06δ,0.34δ
E55	8.62	123.56	174.90	55.24	29.54	4.89	1.98,2.11	33.82	2.23γ
I56	9.22	124.02	174.25	61.26	40.34	4.36	1.64	26.96,17.52,13.49	0.56γ,0.35δ
L57	8.45	126.17	175.84	55.02	42.91	5.16	1.83,1.61	29.11,24.99	1.61γ,0.84δ
L58	9.04	120.89	175.10	53.55	46.68	5.42	1.38,1.52	26.18	1.40γ,0.63δ
Q59	9.20	115.55	175.38	53.89	32.45	5.44	2.17	33.65	2.26γ,2.40γ
K60	8.91	119.18	173.98	55.95	37.85	4.69	1.60	24.90,42.91	1.42γ,1.60δ,3.02ϵ
W61	9.69	127.51	175.91	58.01	30.47	5.19	3.43,3.06		
E62	8.67	125.06	175.03	55.17	31.42	4.47	1.76,1.96	33.48	
N63	8.86	119.16	174.32	55.43	37.42	4.13	2.92,2.73		
D64	8.04	108.70	172.83	54.40	37.94	4.09	3.14,2.86		
E65	6.74	112.08	172.30	54.32	31.76	4.13	1.80,1.94	32.02	2.36γ,2.20γ
C66	8.66	117.63	173.42	55.17	40.08	5.01			
A67	8.98	130.43	176.39	51.83	21.72	4.58	1.41		
Q68	8.34	116.32	176.44	55.17	30.99	5.15	1.97,2.13	34.68	2.31γ,2.47γ
K69	9.10	121.15	174.12	54.66	36.22	4.72	1.52,1.69	24.90,34.18,42.31	1.28γ,2.54ϵ,2.77ϵ
K70	8.49	121.15	175.67	55.43	34.93	5.21	1.67	25.16,30.04,42.22	1.17γ,1.29γ,1.65δ,2.89
I71	8.93	121.67	174.31	59.89	41.97	4.45	1.68	27.30,17.69,14.01	1.33γ,0.98γ,0.69δ
I72	8.45	123.00	173.91	60.32	38.02	4.46	1.91	28.07,17.87,12.12	0.72γ,0.60δ
A73	9.25	130.94	176.21	50.03	20.01	5.07	1.02		
E74	9.39	122.09	176.44	55.43	30.56	4.62	2.03,2.23	33.82	2.54
K75	8.50	120.74	175.66	57.66	32.96	4.12	1.99	24.21,30.13,42.8	1.31γ,1.56γ,1.75δ,3.03
T76	8.41	110.11	174.54	60.49	72.59	4.84	4.49	21.38	1.25
K77	8.21	113.99	176.34	57.83	32.96	4.18	1.84,1.97	25.58,42.57	1.52γ,1.74δ,2.45ϵ,3.02ϵ
I78	7.97	119.84	175.06	57.89	39.59	4.55	1.69	27.36,18.06,12.1	1.41γ,1.13γ,0.93γ,0.82δ
P79			175.95	64.87	32.87	4.27	2.48,2.01	27.99,51.92	
A80	7.80	112.75	174.52	52.86	20.52	4.12	1.83		
V81	7.27	114.93	175.20	61.95	33.99	4.99	1.75	20.02,22.67	1.07γ,0.90γ
F82	9.51	123.83	175.08	56.12	44.28	5.24	2.79,2.51	131.45δ6.87δ,7.16ϵ	
K83	9.46	120.87	176.60	55.69	35.10	5.01	1.81	25.33,29.96,42.4	1.42γ,1.59γ,1.79δ,3.04ϵ
I84	7.72	113.94	174.80	59.72	44.20	4.89		25.16,19.75,14.78	0.79γ,0.66δ
D85	8.74	116.74	173.54	52.86	38.79	5.05	2.75,2.94		
A86	7.19	119.44	175.92	52.09	23.18	4.65	1.33		
L87	8.96	112.11	175.95	56.12	38.71	3.86	1.68,1.91	26.96,23.01	1.41γ,0.81δ
N88	8.65	108.73	174.66	55.35	38.36	4.24	3.19,2.88		
E89	8.22	114.91	176.09	54.23	32.45	5.07	2.02	30.13	2.35γ,2.75γ
N90	9.06	113.15	176.82	53.29	40.42	5.77	2.84,3.44	107.1(N)	7.45,6.85(NH_2)
K91	8.55	119.17	175.41	55.95	38.19	5.57	1.46,1.71	25.41,30.13,42.74	1.48γ,2.90ϵ
V92	9.00	118.98	173.98	61.61	35.45	4.83	1.83	23.01,20.95	0.72γ,0.81γ
L93	9.39	123.54	176.37	54.12	44.97	5.36	1.42	28.33,26.53,8.78	1.58γ,0.81δ
V94	9.27	121.44	175.13	63.58	31.85	4.36	2.30	21.04	0.85γ
L95	8.43	125.57	176.82	58.09	43.00	4.31	1.49,1.82	29.19,25.76,23.70	1.73γ,0.84δ,0.81δ
D96	7.18	107.66	174.19	54.83	44.02	4.99	3.21,2.34		
T97	7.75	116.50	169.58	60.75	69.67	4.16		18.04	0.29γ

Table 1 continued.

Residue	HN	N	CO	C^{α}	C^{β}	H^{α}	H^{β}	C-other	H-other
D98	6.78	120.88	179.23	53.03	42.22	4.64	3.22,2.49		
Y99	8.99	115.10	176.81	65.38	36.99	3.81	3.65,3.11		6.89δ
K100	8.97	117.29	176.63	57.32	35.02	4.32	1.85,1.96	25.58,29.53,42.57	1.58γ,3.10ϵ
K101	9.34	116.74	176.51	59.46	36.05	4.81	2.81,2.13	25.93,29.62,42.74	1.55γ,1.11γ,1.87δ,3.12ϵ
Y102	8.98	114.14	174.35	57.06	44.37	6.57	3.12,2.82		
L103	9.13	122.54	173.67	57.15	44.45	4.39	0.38,1.02	27.99,26.70	
L104	9.33	122.47	176.42	53.29	44.45	5.78	1.51,2.26	27.81,26.61,24.90	2.09γ,1.09δ,1.05δ
F105	8.86	118.90	172.41	56.38	42.82	5.80	3.23,2.70		
C106	9.62	114.82	172.93	56.38	49.86	5.72	3.12,2.70		
M107	9.87	118.21	173.51	54.83	37.85	5.71	2.06	33.48	2.67γ
E108	8.37	111.22	174.12	55.26	32.79	4.79	2.16	32.94	2.05γ
N109	9.34	117.74	177.58	52.09	39.48	5.17	3.24,2.97		
S110	9.76	119.62	174.73	61.61	63.58	3.96	4.11,3.77		
A111	8.28	120.91	178.57	53.97	19.92	4.37	1.43		
E112	7.72	110.94	174.94	53.68	28.55	4.86	1.97,2.22	31.84	2.38γ
P113			177.94	66.67	32.51	4.18	2.16,1.87	28.07,50.37	1.47γ
E114	8.27	112.19	178.08	59.29	28.33	4.13	2.16	34.42	2.53γ
Q115	7.75	113.42	176.21	57.58	29.44	4.39	2.02,2.26	34.76	2.47γ
S116	7.56	107.19	173.98	58.95	65.64	4.57	4.72,3.87		
L117	7.24	123.99	176.37	57.23	43.17	4.93	1.79	28.07,27.21,24.21	1.79γ,0.82δ,0.91δ
V118	8.78	123.52	174.94	60.66	35.96	5.20	2.02	22.41	1.11γ,0.92γ
C119	9.37	119.46	173.33	55.86	50.37	5.67	3.03,2.66		
Q120	9.32	116.13	173.37	55.95	33.22	5.07	4.71,4.05	38.34	
C121	7.84	117.76	173.00	57.24	27.82	4.63	1.24,2.23		
L122	9.68	126.00	177.44	52.78	45.14	5.95	1.70,1.92	27.04,25.42	1.71γ,1.08δ,0.38δ
V123	9.42	109.02	177.99	59.20	37.41	6.16	2.87	24.98,19.67	1.21γ,1.30γ
R124	7.90	117.24	176.18	57.23	33.73	3.93	1.74	26.36,43.17	1.67γ
T125	7.50	105.98	172.81	57.59	70.42	4.87	4.26	22.07	1.19γ
P126			177.23	63.16	32.34				
E127	8.25	117.54	175.50	55.17	31.25	4.37	1.39	33.13	
V128	8.30	116.64	175.13	63.67	31.50	3.92	2.23	21.64,21.46	1.12γ,0.98γ
D129	6.52	121.70	175.53	53.29	40.68	4.84	2.89		
D130	9.04	121.41	177.56	57.15	38.54	4.39	2.85		
E131	8.32	118.56	178.15	59.38	27.99	4.07	2.11,2.22	33.82	2.48γ
A132	7.61	119.07	178.97	55.69	19.67	3.29	1.23		
L133	7.72	113.94	179.18	58.18	41.97	4.04	1.64	27.30,25.67,23.87	0.65δ
E134	8.01	117.31	179.63	59.38	27.81	4.35		33.73	
K135	7.91	117.30	179.74	59.38	32.53	3.92	1.55,1.71	25.67,29.10,42.82	
F136	8.51	119.13	175.86	60.49	39.65	4.36	3.38,3.21	131.5δ	7.13δ,6.90ϵ
D137	8.68	114.77	178.85	56.89	38.71	4.03	2.92,2.75		
K138	7.90	117.05	179.08	59.81	32.87	3.94	1.88,1.67	25.67,29.70,42.65	1.45γ,2.97δ
A139	7.81	119.39	179.37	54.92	18.04	4.12	1.39		
L140	7.47	111.85	178.76	55.78	41.88	3.90	1.29,1.40	27.04,24.64,23.78	1.35γ,0.19δ,0.05δ
K141	7.14	117.17	177.07	59.63	32.96	3.93	1.89	25.16,29.53,42.5	1.34γ,1.56δ,1.69δ,2.98ϵ
A142	7.77	115.99	176.86	52.09	19.75	4.46	1.41		
L143	7.69	117.15	176.23	52.24	43.53	4.67	1.13,1.64	26.26,25.93,23.54	1.57γ,0.76δ,0.77δ
P144		174.96	62.47	28.24	4.62		49.77		
M145	7.65	115.68	177.09	54.57	33.31	4.16	2.27		
H146	8.75	115.49	172.46	56.29	31.59	4.95	3.61,2.85		
I147	7.69	114.92	171.97	60.58	40.68	4.63	29.61,17.78,14.95		1.49γ,0.79γ,0.97δ
R148	8.30	124.23	174.22	55.43	33.39	4.97	1.65,1.78	34.50,43.85	1.47,3.05δ[a]

Table 1 continued.

Residue	HN	N	CO	C^{α}	C^{β}	H^{α}	H^{β}	C-other	H-other
L149	9.12	121.67	175.72	53.97	46.25	4.73	1.21,1.42	27.21,24.04,21.38	0.91
S150	8.16	113.99	172.88	57.83	65.90	4.88	3.79		
F151	8.05	117.49	174.08	57.44	42.40	5.06	3.18,2.75		7.44δ
N152	8.72	116.86	174.31	51.43	38.96	5.01	2.84,2.72		
P153			176.13			4.38			
T154	8.11	113.25	174.68	62.47	70.11	4.35		22.01	1.21γ
Q155	8.31	120.12	176.32	56.49	33.48	4.37			
L156	8.17	120.15	176.30	57.55	39.23	4.30			
E157				–					
E158				–					
Q159	8.23	122.51	175.51	56.27	29.39	4.28			
C160	8.22	118.99	177.28	53.24	38.78	4.73	2.88		
H161	8.13	117.58	175.97	56.37	29.39	4.37			
I162									

The water signal was set to 4.644 at 37 °C, and the indirectly detected dimensions were referenced as described in Live et al. (1984), and Bax and Subramanian (1986).

[a] Amide proton chemical shifts are ±0.01 ppm, nitrogen chemical shifts are ±0.03 ppm, carbon chemical shifts are ±(0.30–0.55 ppm), and indirectly detected protons are ±(0.02–0.08 ppm) depending on the experiment.

Secondary structure determination and topology

Figure 7 illustrates the outcome of chemical shift analysis according to Wishart and Sykes (1994). Eight β-strands (strands A–H) followed by an α-helix and then a ninth β-strand (strand I) are predicted. The postulated strands are between 4 and 10 residues in length, and the helix appears to encompass 12 residues. These results are largely consistent with the analysis of sequential and short-range NOEs (see Figure 8) for strands C, D, E, G, H and I which shows blocks of strong, sequential αN NOEs, coincident with an absence of strong sequential NN NOEs, for residues 54–63 (strand C), 65–77 (D), 81–86 (E), 102–109 (G), 117–123 (H) and 147–149 (I). Other strands are less strongly indicated by the sequential NOE data: strand A is less well defined by sequential NOEs due to the lack of an assignment for the NH of Leu^{22}; within the CSI-predicted strand B, no αN(i,i+1) is observed for Glu^{44}; and within the predicted strand F, strong NN(i,i+1) NOEs are observed for residues 90 and 95. The CSI prediction of helix for residues 128–141 is also largely consistent with the run of medium or strong sequential NN NOEs involving residues 129–141, and seven αN(i,i+3)/NN(i,i+2) NOEs in this region. The sequential NOE diagram also indicates the presence of several αN(i,i+2) NOEs consistent with a region of 3_{10}-helix encompassing residues 11–15. The presence of NN(i,i+4) connectivities from residues 49 and 98 is consistent with tight turns between β-strands.

Figure 9 illustrates the NOE connectivities observed between β-strands. These are consistent with a β-barrel composed of eight antiparallel β-strands with $(+1)_8$ topology. The pattern of inter-strand NOEs is atypical in places: for example, within strand F and between strands F and G. This indicates that strand F is twisted. Strand I appears to lie adjacent to the C-terminal portion of strand A and to run antiparallel to it.

Comparison with X-ray structure

The recently determined high-resolution X-ray structure of dimeric β-LG at pH 6 identified nine β-strands (Brownlow et al., 1997) as shown in Figure 8 (grey symbols). The closed nature of the β-sheet formed in β-LG arises from kinks in two of the β-strands. Strand A has a marked bend at residue 22, a feature common to the lipocalin family, which is reflected in a significant deviation of the phi/psi angles from those of a β-strand (observed, −81/−41°; ideal, −140/+135°. DSSP still reports this region as an extended strand and does not report a break. In the case of strand F, residue 90 has a psi torsion angle of 5° while residue 94 has phi and psi angles of −84° and −42°. Once again, DSSP reported strand F as continuous.

These observations largely coincide with the CSI-predicted strands for monomeric β-LG in solution at

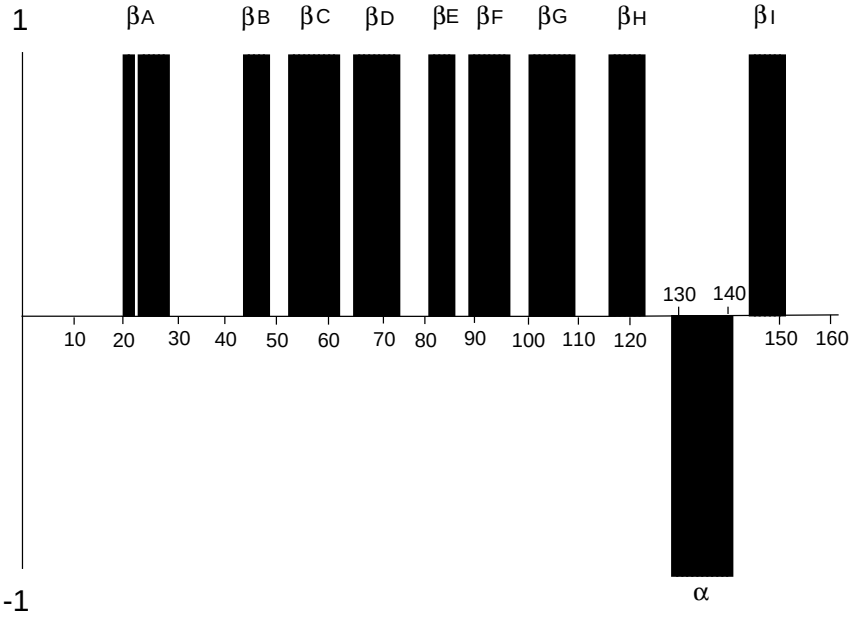

Figure 7. Consensus CSI for β-LG derived from deviations of H^{α} , C^{α} and C^{β} and carbonyl carbon chemical shifts from random coil values (Wishart and Sykes, 1994) indicating secondary structure elements. In the figure, a stretch of four or more sequential residues with an index of +1 indicates a β-strand region and, similarly, a stretch of residues with a −1 index indicates a region of α-helicity. Residues were assigned as a helix or a β-strand only if three out of the four H^{α}, C^{α}, C^{β}, CO CSIs indicated the same structural element.

pH 2, although strand lengths differ by one or two residues in most cases and strand A (from CSI) appears to be two residues shorter at its N-terminal end and two residues longer at its C-terminal end. Differences of one residue in strand length can be explained both by the limited reliability of the CSI program in predicting the precise boundaries of secondary structure, and the criteria adopted within the program (DSSP (Kabsch and Sander, 1983) used to ascribe secondary structure within the crystal structure of β-LG. However, the differences (±2 residues) between crystal and NMR data for strand A are difficult to explain in this way and might reflect a genuine small difference in secondary structure. The kink in strand A occurs at Leu^{22} which was one of a very few residues for which no amide proton or ^{15}N was assigned. Further, it is interesting to note that it was the length of strand A and the subsequent loop which led to the misthreading of the initial β-LG X-ray structure (Papiz et al., 1986; Brownlow et al., 1997).

The sequential NOE data also agree with the X-ray-derived secondary structure. Comparison with the appropriate distances measured in the X-ray-derived structure indicated that the ^{15}N-edited NOESY experiment was not very sensitive in this instance, since NOEs were observed only for protons less than about 3.7 Å apart. This was probably a consequence of using a doubly labelled sample of β-LG for this experiment. However, given this limitation, most expected NOEs were observed and none of the observed NOEs clashed with what would be expected from the X-ray structure. For example the presence (already alluded to) in strand F (residues 89–96 according to CSI or 89–97 according to DSSP) of strong NN(i,i+1) NOEs for residues 90 and 95 is consistent with the NN distances (1.9 and 2.5 Å) in the crystal structure. The presence of a tight γ-turn between strands F and G noted in the crystal structure is consistent with the observation of a medium-strength NN(i,i+4) from residue 98. One of the four segments of the 3_{10}-helix described for the X-ray structure is clearly indicated in the NMR data by the presence of three αN(i,i+2) connectivities in the region 11–15; the other 3_{10}-helical turns are not as clearly characterised by the NMR data.

Inter-strand NOEs are sparse due to the aforementioned poor sensitivity of the ^{15}N-edited NOESY as collected on a double-labelled sample and some loss under residual water of α protons in the ^{13}C-edited NOESY. However all those NOEs identified are consistent with distances measured in the X-ray structure. These include the atypical ones mentioned earlier. Some NOEs that might be expected on the basis of

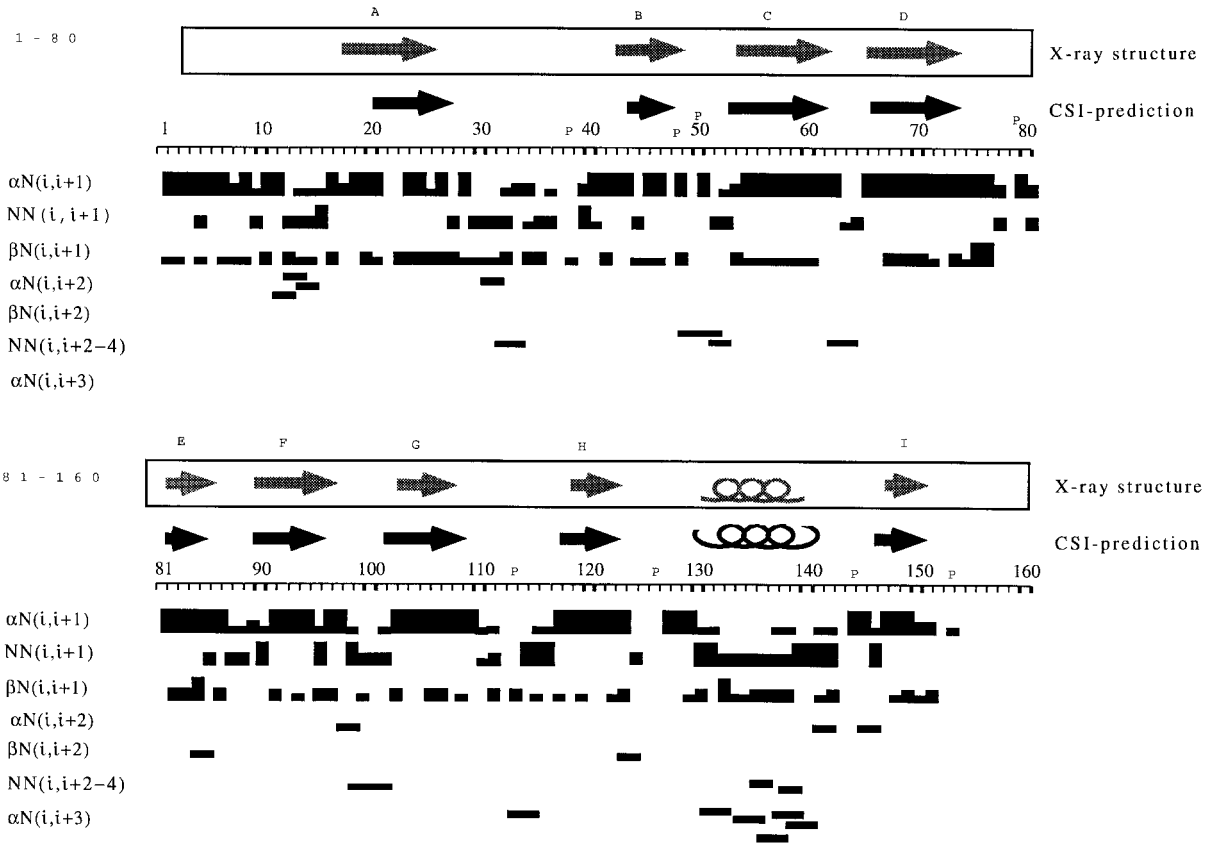

Figure 8. Schematic representation of (i,i+1), (i,i+2), (i,i+3), and (i,i+4) NOEs for β-LG. The observation of an NOE of the appropriate category is indicated by a horizontal line connecting the two residue positions indicated by the ruler. Line thickness is approximately proportional to NOE strength (classified as weak, medium or strong). The occurrence of a proline (no amide proton) in the sequence is indicated by the letter 'P'. A schematic representation of the secondary structure as indicated by the CSI is shown above the ruler as black symbols (arrow = β-strand, loops = α-helix). For comparison, secondary structural elements, as derived from the X-ray structure and classified using DSSP, are shown as stippled symbols.

the X-ray structure are absent (not as a consequence of overlap). These include a connectivity in the CD loop (62 NH to 67 NH); two inter-strand connectivities between strands A and I (24 NH to 149 NH; 26 NH to 147 NH); and a connectivity between strands A and H (25 NH and 119 NH). It is possible that the absence of these NOEs reflects small local changes of structure in this region of the monomer which is close to the dimer interface.

Significance for understanding the conformation of β-LG at low pH

Urea denaturation studies of β-LG at pH 2 revealed this protein to undergo no unfolding transition below 4 M (Fink et al., 1994) and it was the most acid-resistant of 20 proteins used in the study. The studies of Molinari et al. (1996) and Ragona et al. (1997) employed ^{1}H NMR, near UV-CD and measurements of ANS fluorescence to demonstrate that β-LG displays some characteristics of a partially unfolded protein at pH 2, exhibiting a β-sheet core, a stable hydrophobic cluster of 11 residues, but a sufficiently expanded hydrophobic core to allow access to ANS. The results of our experiments indicate that both the secondary structure and the topology of β-LG are largely the same when it is a monomer at low pH as when it is a dimer at neutral pH. We did make some observations that might be explained by disorder or conformational heterogeneity, and we were unable to detect NOEs in a few cases where they might have been expected on the basis of the neutral pH structure. However, with the exception of some local differences close to the potential dimer interface, and the apparent differences in the extent of strand A, we have been unable so far

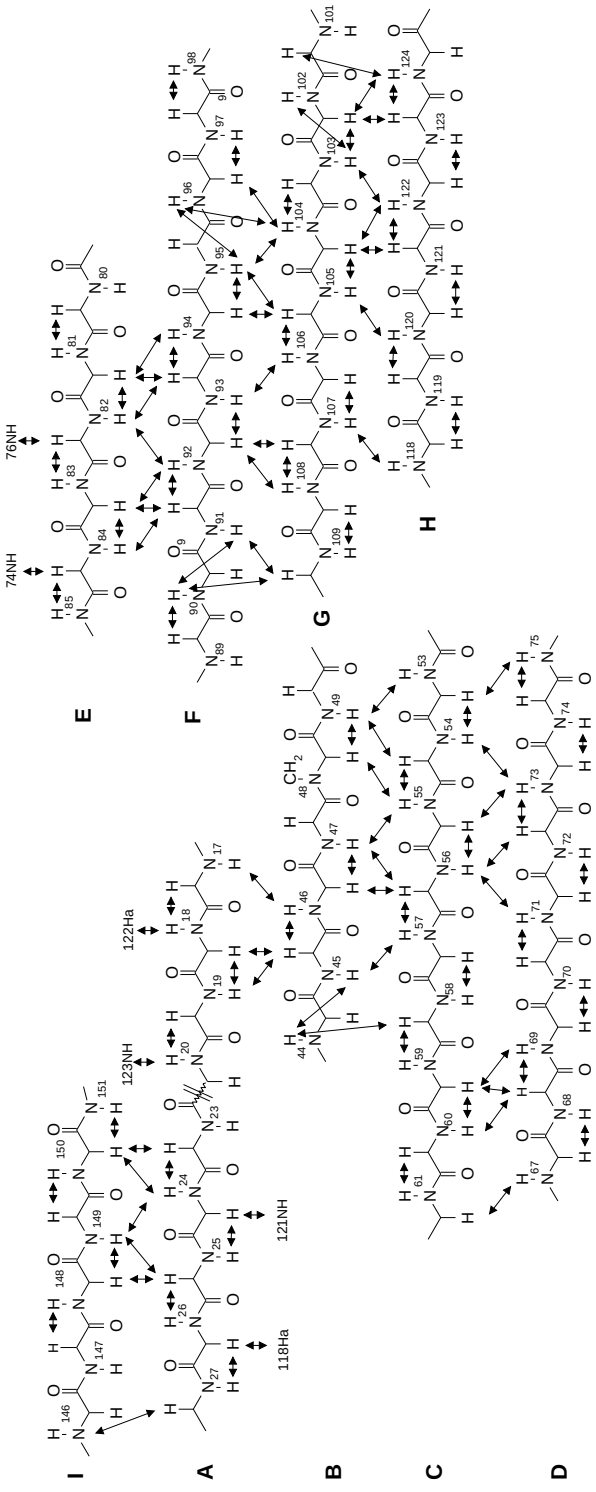

Figure 9. Summary of the intra- and inter-strand NOEs observed for β-LG. NOE connectivities are indicated by the double-headed arrow; strands are labelled A–I as in Figure 8

to identify any particular region of the structure which is obviously different from the native form at pH 6.5. Taking together the current results with those of Molinari et al., this suggests that the structure of the β-LG monomer at low pH, while not fully folded, is very close to its structure in the dimer.

Conclusions

We have used heteronuclear NMR to probe the secondary structure and topology of β-LG under extreme acid conditions. Despite the fact that the molecule is not fully folded, nearly all of the native secondary structure and topology is retained. Further studies will be aimed at solving the solution structure of β-LG at pH 2.6, and at a comprehensive dynamic analysis using relaxation-based techniques. These studies will be of particular value if this form of the protein is indeed a kinetic intermediate of the folding pathway, as has been suggested (Molinari et al., 1996). The quality of the spectra is such that it should also be possible to study the monomer–dimer transition and the formation of the native structure. The assignments presented here will also enable an exploration of the hydrophobic binding site on the basis of ligand-induced chemical shift perturbations in an isotopically labelled sample.

Acknowledgements

We are most grateful to Carl Batt (Cornell University) for providing us with the overexpressing strain used in this work. This work was supported by the Wellcome Trust (S.U., D.U. and P.N.B.), the Biomolecular and Biotechnological Sciences Research Council (through the support of the Edinburgh Centre for Protein Technology and to L.S.) and a research grant from the Foundation for Research, Science and Technology, New Zealand, Contract Number DRI 403 (M.S.).

References

Bartels, C.H., Xia, T.-H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **5**, 1–10.

Bax, A. and Subramanian, S. (1986) *J. Magn. Reson.*, **67**, 565–569.

Bocskei, Z., Groom, C.P., Flower, D.R., Wright, C.E., Cavaggioni, A., Findlay, J.B. and North, A.C.T. (1992) *Nature*, **360**, 186–188.

Brownlow, S., Cabral, J.H.M., Cooper, R., Flower, D.R., Yewdall, S.J., Polikarpov, I., North, A.C.T. and Sawyer, L. (1997) *Structure*, **5**, 481–495.

Cowan, S.W., Newcomer, M.E. and Jones, T.A. (1990) *Proteins*, **8**, 44–61.

Domke, T. and Leibfritz, D. (1990) *J. Magn. Reson.*, **88**, 401–405.

Engelke, J. and Rüterjans, H. (1995) *J. Magn. Reson.*, **B109**, 318–322.

Fink, A.L., Calciano, L.J., Goto, Y., Kurotsu, T. and Palleros, D. (1994) *Biochemistry*, **33**, 12504–12511.

Flower, D.R., North, A.C.T., and Attwood, T.K. (1993) *Protein Sci.*, **2**, 753–761.

Fox, P.F. (1995) *Heat Induced Changes in Milk*, 2nd ed., International Dairy Federation, Brussels.

Gardner, K.H., Konrat, R., Rosen, M.K. and Kay, L.E. (1996) *J. Biomol. NMR*, **8**, 351–356.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440

Grzesiek, S., Anglister, J. and Bax A. (1993) *J. Magn. Reson.*, **B101**, 114–119.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Hambling, S.G., McAlpine, A.S. and Sawyer, L. (1992) In *Advanced Dairy Chemistry I* (Ed., Fox, P.F.), Elsevier, Amsterdam, pp. 141–190.

Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H. (1987) *J. Mol. Biol.*, **198**, 499–513.

Jakob, E. and Puhan, Z. (1992) *Int. Dairy J.*, **2**, 157–178.

Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A (1990) *J. Magn. Reson.*, **89**, 496–514.

Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.

Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) *J. Magn. Reson.*, **A109**, 129–133.

Kim, T.R., Goto, Y., Hirota, N., Kumata, K., Denton, H., Wu, S.-Y., Sawyer, L. and Batt, C.A. (1997) *Protein Eng.*, **10**, 1339–1345.

Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. (1984) *J. Am. Chem. Soc.*, **106**, 1939–1941.

Logan, T.M., Olejniczak, E.T., Xu, R.X. and Fesik, S.W. (1993) *J. Biomol. NMR*, **3**, 225–231.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

McCoy, M.A. and Mueller, L. (1992) *J. Am. Chem. Soc.*, **114**, 2108–2112.

Mohebbi, A. and Shaka, A.J. (1991) *J. Chem. Phys.*, **178**, 374–377.

Molinari, H., Ragona, L., Varani, L, Musco, G., Consonni, R., Zetta, L. and Monaco, H. (1996) *FEBS Lett.*, **381**, 237–243.

Mori, S., Abeygunawardana, C., Johnson, M.O. and van Zijl, P.C.M. (1995) *J. Magn. Reson.*, **B108**, 94–98.

Muhandiram, D.R. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 203–216.

Newcomer, M., Jones, T.A., Agvist, J., Sundelin, J., Eriksson V., Rask, L. and Peterson, P.A. (1984) *EMBO J.*, **3**, 1451–1454.

Papiz, M.Z., Sawyer, L., Eliopoulus, E.E., North, A.C.T., Findlay, J.B.C., Sivaprasadarao, R., Jones, T.A., Newcomer, M.E. and Kraulis, P.J. (1986) *Nature*, **324**, 383–385.

Perez, M.D. and Calvo, M. (1995) *J. Dairy Sci.*, **78**, 978–988.

Piotto, M., Saudek, V. and Sklenář, V., (1992) *J. Biomol. NMR*, **2**, 661–665.

Ragona, L., Pustera, F., Zetta, L., Monaco, H.L., and Molinari, H. (1997) *Folding Design*, **2**, 281–290.

Schmidt, J.M. and Rüterjans, H. (1990) *J. Am. Chem. Soc.*, **112**, 1279–1280.

Shaw, G.L., Müller, T., Mott, H.R., Oschkinat, H., Campbell, I.D. and Mitschang, L. (1997), *J. Magn. Reson.*, **124**, 479–483.

Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.*, **A102**, 241–245.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.
Tanford, C., Bunville, L.G. and Nozaki, Y. (1959) *J. Am. Chem. Soc.*, **81**, 4032–4036.
Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. and Cambillau, C. (1996) *Nat. Struct. Biol.*, **3**, 863–867.
Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.
Wang, A.C., Lodi, P.J., Qin, J., Vuister, G.W., Gronenborg, A.M. and Clore, M. (1994) *J. Magn. Reson.*, **B105**, 196–198.
Wishart, D.S. and Sykes, B.D., (1994) *J. Biomol. NMR*, **4**, 171–180.
Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 11054–11055.
Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.