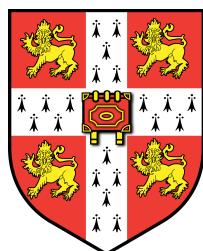


Developing a Solid-State NMR Toolkit for Structural Analysis of Collagen

Part III Project

Alex Massey
Robinson College



Lent Term 2011

Declaration

This dissertation is submitted in partial fulfilment of the requirements for Part III Chemistry. It describes work carried out in the Department of Chemistry in the Michaelmas Term 2010 and the Lent Term 2011. Unless otherwise indicated, the research described is my own and not the product of collaboration.

Alex Massey

Acknowledgements

Acknowledgement text goes here

Contents

1	Introduction	5
2	Recoupling of Anisotropic Interactions	7
2.1	Chemical Shift and Shielding Tensors	7
2.2	Recoupling Experiments	9
2.2.1	Types of Recoupling Experiment	10
3	The 2DCSA Experiment	12
3.1	Theory	12
3.1.1	Pulse sequence for idealised pulses	15
3.2	Phase Cycling	16
3.2.1	Nested Phase Cycling	16
3.2.2	Cogwheel Phase Cycling	17
3.3	Optimised and Constant-Time 2DCSA Experiments	17
3.3.1	Constant-Time Experiment	17
3.3.2	Optimised Experiment	18
4	Experimental Details	19
4.1	Recording of Spectra	19
4.2	Samples	20
4.3	Fitting of 2DCSA Lineshapes	20
4.3.1	Fitting of Lineshapes with Multiple Components	21
4.4	Error Analysis	22
5	NMR Studies of Poly(L-Proline)	25
5.1	^{13}C Cross-Polarisation (CP) Spectrum of Polyproline	25
5.2	γ -Carbon Splitting	25
5.2.1	Implications for Collagen	28
5.3	^{13}C 2DCSA of Polyproline	29

6 NMR Studies of Poly(Pro-Gly-Pro)	31
6.1 ^{13}C CP Spectrum of Poly(Pro-Gly-Pro)	31
6.2 Solution-State Studies of Poly(Pro-Gly-Pro)	31
6.3 ^{13}C 2DCSA of Poly(Pro-Gly-Pro)	33
6.3.1 Fitting of 2DCSA Spectrum with Multiple Components .	34
7 ^{13}C 2DCSA Studies of Triple-Helical Peptides	40
7.1 Collagen Model Peptides	40
7.2 Other Triple-Helical Peptides	43
7.2.1 Sheep Osteoblast Collagen	43
7.2.2 Poly(Pro-Gly-Pro)	44
7.2.3 Aluminium Tanned Leather	45
7.3 Ala-Pro-Gly	45
7.4 Error Analysis	47
7.5 Chemial Shift Prediction Programs	48
7.6 Discussion	48
7.6.1 Glycine Carbonyl Chemical Shifts	48
7.6.2 Proline Carbonyl Chemical Shifts	49
7.6.3 Aliphatic Tensors	50
7.7 Hypothesis Testing	50
7.8 ^{13}C 2DCSA Measurements of Gelatine	54
8 Conclusions and Further Work	58
8.1 Other Model Compounds	58

1 Introduction

Collagen is an extremely important and abundant structural protein in the human body, accounting for 25% of the protein mass [1]. Indeed, it is the most abundant protein in the world [2]. Collagen is secreted by cells in connective tissue, forming the environment in which those cells live. The secreted collagen forms networks, with the cells residing in spaces in the networks or on the surface of them. The collagen must therefore be able to chemically bond effectively to the cells. It also has a more mechanical role; collagen fibres have exceptional tensile strength, and for this reason are the principal component of skin, muscle and tendon and the principal organic component of bone. The individual collagen molecules must aggregate together in a very specific hierarchical structure [1] to produce these high-strength fibres.

In terms of composition, collagen is largely based on a Gly-X-Y repeat unit, where Gly is glycine. X and Y vary, with proline (Pro) being common. Collagen is unique in containing large quantities of hydroxyproline (Hyp) residues formed by post-translational modification of proline by the addition of a hydroxyl group to the γ -carbon. In vertebrates, only proline residues in the ‘Y’ position are hydroxylated [3]. The hydrogen bonding ability of these hydroxyl groups is thought to be important for the stability of the quarternary structure of the collagen chains [3]. The high proportion of proline residues imposes a conformational restraint on the molecules, allowing them to adopt the correct quarternary structure.

In its native state, the quarternary structure of collagen is triple-helical. The individual chains adopt the polyproline II conformation [4, 5]. The triple-helix is very important for the mechanical properties of the protein. Accurate structure determination of the collagen triple-helix is important if we are to be able to completely understand the health problems caused by diseased or mechanically-degraded collagen. We need to be able to assess the defects in the structure in these cases and be able to compare with healthy collagen. Even if collagen can be synthetically produced with the correct primary structure, for it to be biologically or medically useful we need to be sure it has exactly the same structure as native collagen.

How do we determine the structure accurately? X-ray diffraction is unsuitable for studying a material as heterogeneous as collagen. NMR is an attractive option; collagen is insoluble, so solid-state NMR is necessary. The isotropic chemical shift of a particular nucleus does depend on its environment, but more detailed information can in principle be obtained through analysis of full chemical shift tensors. These tensors depend on the geometry of the molecule, through their dependence on electronic structure. In practice the dependence on geometry is complex. In this work, the correlation of the shift tensors with molecular structure is investigated with model peptides — shorter triple-helical peptides whose structures can be determined using X-ray diffraction. No studies of this type have been reported before for collagen. The same studies are then performed with native collagen samples, and the triple-helical tensors compared to those for single-helical peptides with the polyproline II conformation and a polymeric peptide which forms both single- and triple-helical regions.

Ultimately this work should lead to the development of an NMR toolkit for the simple and fast identification of the triple-helical structure of collagen.

2 Recoupling of Anisotropic Interactions

Magic-angle spinning (MAS) is an indispensable technique in solid-state NMR for removing anisotropic interactions and allowing signals from inequivalent nuclei to be resolved. However, the chemical shift tensor¹ principal components² potentially contains a large amount of structural information. It is therefore desirable to have a technique for obtaining the anisotropic chemical shift information in a MAS experiment.

2.1 Chemical Shift and Shielding Tensors

It is important to distinguish the chemical *shielding* tensor, with components denoted σ_{ab} , from the chemical *shift* tensor, with components denoted δ_{ab} . The chemical shift tensor is an experimentally-determined quantity which relies on a reference compound [6]:

$$\delta_{ab} = \frac{\sigma_{ab}^{\text{ref}} - \sigma_{ab}}{1 - \sigma_{ab}}. \quad (1)$$

The chemical shielding tensor contains explicit information about the electronic structure around a nucleus, and can be denoted as an ellipsoid. The larger a component of σ (i.e. the more ‘shielded’ the component), the lower the corresponding component of δ .

¹Strictly, the chemical *shift* parameters cannot be described as a tensor — the chemical *shielding* is a tensor whilst chemical shift describes how the shielding tensor interacts with a magnetic field. However, reference to the ‘chemical shift tensor’ or ‘CSA tensor’ (where CSA stands for chemical shift anisotropy) is an intuitive concept and is used in this work.

²Determination of the ‘full’ shielding tensor would require determination of the directions in addition to the values of the principal components. The principal directions were not investigated in this work since it is highly non-trivial to do so.

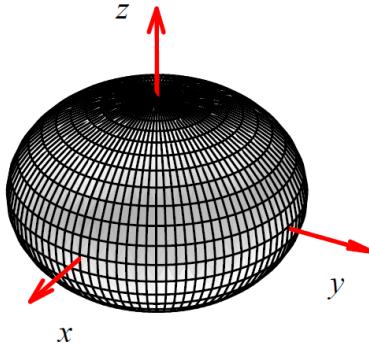


Figure 1: Ellipsoid representation of a chemical shielding tensor. The ellipsoid is associated with an axis frame that is fixed with respect to the molecular structure. Taken from Orr [7].

The shielding tensor has nine components:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

The tensor can be separated into symmetric and antisymmetric components. Only the symmetric part has an effect on the observed interaction with radiation [6], so in fact only the six components of the symmetric part are measurable:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}.$$

By rotation of the axis frame that defines the tensor, we can choose a ‘principal axis frame’ where the tensor is diagonal, with three ‘principal components’,

$$\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}.$$

There are several different notations for how these components are labelled [8]. The ‘Haeberlen’ notation labels the components as σ_{XX} , σ_{YY} and σ_{ZZ} , where

$$|\sigma_{ZZ} - \sigma_{iso}| > |\sigma_{XX} - \sigma_{iso}| > |\sigma_{YY} - \sigma_{iso}|,$$

i.e. the ZZ component is the furthest from the isotropic shielding:

$$\sigma_{iso} = \frac{1}{3}(\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}).$$

The XX , YY , ZZ refer to the principal x , y , z axes. This translates directly to the chemical shift tensor, so δ_{ZZ} is the furthest component from δ_{iso} . The IUPAC notation uses numerical subscripts:

$$\sigma_{11} < \sigma_{22} < \sigma_{33}.$$

In shift notation, this translates to

$$\delta_{11} > \delta_{22} > \delta_{33}.$$

Both notations are useful: the Haeberlen notation allows easier definition of parameters called the anisotropy (Δ) and the asymmetry (η), whilst the IUPAC notation allows easier comparison of structural components. Δ and η are normally quoted in terms of chemical *shift*:

$$\Delta = \delta_{ZZ} - \delta_{iso} \tag{2}$$

$$\eta = \frac{(\delta_{XX} - \delta_{YY})}{\Delta}. \tag{3}$$

2.2 Recoupling Experiments

We can in principle imagine a two-dimensional experiment in which the conventional MAS spectrum is recorded in one dimension, and a static spectrum is recorded in the other. However, in practice it is not feasible to stop the sample spinning and restart again — it takes a long time and the magnetisation will relax back to equilibrium. We instead use a pulse sequence to ‘recouple’ the anisotropic interactions. For simple molecules it is possible to determine the principal shift components from spinning sidebands in a 1D experiment; for complex molecules this is extremely difficult. Experiments exist that recouple anisotropic interactions by switching the spinning angle away from the magic angle for a portion of the experiment [9, 10], but these require special hardware and are not considered here.

2.2.1 Types of Recoupling Experiment

There are broadly speaking two types of two-dimensional CSA recoupling experiment — those that give a ‘quasi-static’ powder pattern in the indirect dimension and those that give sidebands in the indirect dimension. Advantages of the powder-pattern method include:

- The asymmetry parameter η is more reliably obtained from a powder pattern analysis [11].
- Powder pattern recoupling experiments can be performed at higher spinning rates, producing sharper lines in the direct dimension [12].
- For a perfect powder pattern, the principal components can be read off the spectrum. In reality the powder patterns obtained are not perfect and require simulation, but the fitting of powder patterns is less time-consuming than the fitting of sideband spectra [12].

Advantages of the sideband method include:

- Sideband spectra concentrate the signal into narrower components, with a corresponding increase in signal-to-noise ratio.
- The anisotropy parameter Δ is more reliably obtained from a sideband analysis [11], because of the improved signal-to-noise ratio.

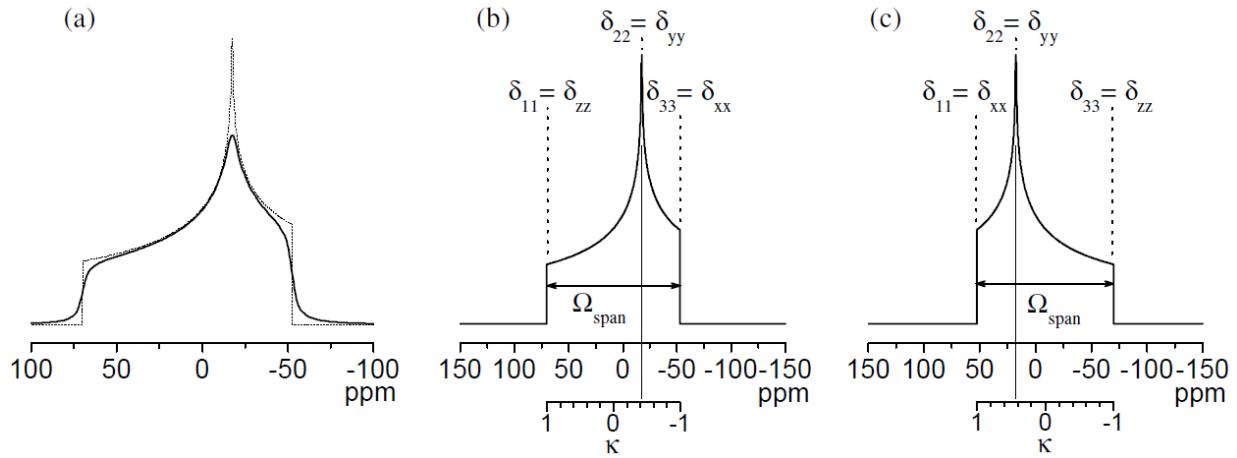


Figure 2: (a) An ideal powder-pattern (dotted line) and a typical powder-pattern (solid line). Relaxation during the recoupling sequence introduces line-broadening upon Fourier transformation — hence the observed patterns are not ideal. (b) and (c) Comparison of the IUPAC and Haeberlen notations. Note how the principal components can in principle be read straight from a powder-pattern. Taken from Orr [7].

3 The 2DCSA Experiment

In this work a variant of the experiment developed by Tycko et al [13] is used to recouple anisotropic interactions in the indirect dimension of a 2D experiment. The advantage of this experiment is that it is robust to small errors in pulse timings, and can be modified [7, 11] to produce very well-defined powder-patterns.

3.1 Theory

This section follows closely the argument in Tycko et al [13]. For a sample spinning at the magic angle at (angular) frequency ω_r , the instantaneous precession frequency $v(t)$ of a nucleus in the rotating frame is

$$\begin{aligned} v(t) = & \Delta v + C_1(\Omega) \cos(\omega_r t) + C_2(\Omega) \cos(2\omega_r t) \\ & + S_1(\Omega) \sin(\omega_r t) + S_2(\Omega) \sin(2\omega_r t) \end{aligned} \quad (4)$$

where Δv is the isotropic offset frequency. $C_1(\Omega)$, $C_2(\Omega)$, $S_1(\Omega)$ and $S_2(\Omega)$ depend on the molecular orientation Ω at time $t = 0$ — i.e. they arise from the anisotropic interaction we seek to recouple. Consider a static sample; $\omega_r = 0$ so the sine terms vanish and the cosine terms give

$$v(t) = \Delta v + C_1(\Omega) + C_2(\Omega). \quad (5)$$

The experiment of Tycko et al recouples powder-patterns using π -pulses symmetrically arranged around the centre of a rotor period. The original experiment used four π -pulses in one rotor period, but in general $2n$ π -pulses can be used in m rotor periods (where m is odd). For the following discussion, we assume that the π -pulses are perfect in the sense that they have no phase errors and are infinitesimally short. The effect of a π -pulse is to change the sign of $v(t)$; n pulses cause a sign change of $(-1)^n$. These are incorporated into a step function $f(t)$, which takes values ± 1 and changes sign after each π -pulse.

We define a ‘pulse sequence unit’ as running from $t = -\frac{m\tau}{2}$ to $t = \frac{m\tau}{2}$ (where τ is the rotor period and m is the number of rotor periods per pulse sequence

unit). The net precession angle ϕ in the t_1 dimension after the pulse sequence unit is given by

$$\phi = \int_{-\frac{m\tau}{2}}^{\frac{m\tau}{2}} v(t)f(t) dt . \quad (6)$$

The π -pulses are arranged symmetrically about the centre of the pulse sequence unit, so $f(t)$ is even; and thus the sine terms vanish in the integral in equation 6. For the pulse sequence to produce undistorted static-like powder patterns, we require C_1 and C_2 to be multiplied by the same factor, as in Equation 5, i.e.

$$\int_0^{\frac{m\tau}{2}} f(t) \cos(\omega_r t) dt = \int_0^{\frac{m\tau}{2}} f(t) \cos(2\omega_r t) dt . \quad (7)$$

The lower limit of the integral can be written as zero because $f(t)$ is even. The precession frequency is $\phi/m\tau$:

$$\frac{\phi}{m\tau} = \frac{1}{m\tau} \int_{-\frac{m\tau}{2}}^{\frac{m\tau}{2}} v(t)f(t) dt \quad (8)$$

$$= \frac{1}{m\tau} \int_{-\frac{m\tau}{2}}^{\frac{m\tau}{2}} (\Delta v + C_1(\Omega) \cos(\omega_r t) + C_2(\Omega) \cos(2\omega_r t)) f(t) dt \quad (9)$$

$$= \xi \Delta v + \chi [C_1(\Omega) + C_2(\Omega)] , \quad (10)$$

where

$$\chi = \frac{2}{\tau} \int_0^{\frac{m\tau}{2}} f(t) \cos(\omega_r t) dt \quad (11)$$

$$\xi = \frac{2}{\tau} \int_0^{\frac{m\tau}{2}} f(t) dt . \quad (12)$$

Equation 10 is of the same form as a static powder-pattern (Equation 5), with the isotropic shift scaled by ξ and the orientation-dependent components scaled by χ .

We now seek an expression which allows the timings of the π -pulses that will satisfy Equation 7 to be calculated. Consider Equation 7 in more detail for the specific case of six π -pulses given at times $t = \pm T_i$ in m rotor periods (Figure 3). $f(t)$ is negative at $t = 0$, and changes sign with each pulse. The LHS of Equation 7 can be expanded to:

$$\begin{aligned} \int_0^{\frac{m\tau}{2}} f(t) \cos(\omega_r t) dt &= - \int_0^{T_1} \cos(\omega_r t) dt + \int_{T_1}^{T_2} \cos(\omega_r t) dt \\ &\quad - \int_{T_2}^{T_3} \cos(\omega_r t) dt + \int_{T_3}^{\frac{m\tau}{2}} \cos(\omega_r t) dt , \end{aligned} \quad (13)$$

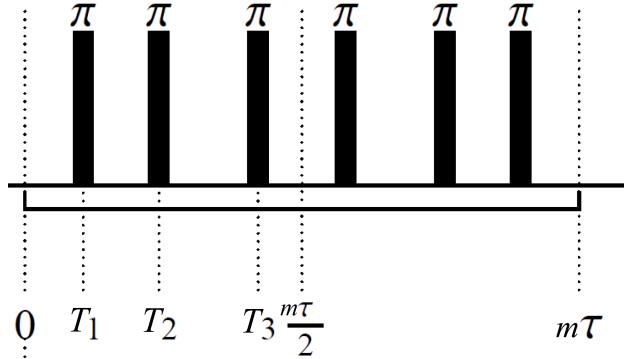


Figure 3: 2DCSA Pulse sequence unit consisting of six π -pulses. Adapted from Orr [7].

using the standard integral $\int \cos(x) dx = \sin(x)$ to get

$$\begin{aligned} \int_0^{\frac{m\tau}{2}} f(t) \cos(\omega_r t) dt &= - [\sin(\omega_r t)]_0^{T_1} + [\sin(\omega_r t)]_{T_1}^{T_2} \\ &\quad - [\sin(\omega_r t)]_{T_2}^{T_3} + [\sin(\omega_r t)]_{T_3}^{m\tau/2} \end{aligned} \quad (14)$$

$$\begin{aligned} &= - 2 \sin(\omega_r T_1) + 2 \sin(\omega_r T_2) \\ &\quad - 2 \sin(\omega_r T_3) + \sin(\omega_r m\tau/2). \end{aligned} \quad (15)$$

We can do the same for the RHS of Equation 7, using the identity $\int \cos(2x) dx = \cos(x) \sin(x)$ to get

$$\begin{aligned} \int_0^{\frac{m\tau}{2}} f(t) \cos(2\omega_r t) dt &= - \int_0^{T_1} \cos(2\omega_r t) dt + \int_{T_1}^{T_2} \cos(2\omega_r t) dt \\ &\quad - \int_{T_2}^{T_3} \cos(2\omega_r t) dt + \int_{T_3}^{\frac{m\tau}{2}} \cos(2\omega_r t) dt \end{aligned} \quad (16)$$

$$\begin{aligned} &= - [\sin(\omega_r t) \cos(\omega_r t)]_0^{T_1} + [\sin(\omega_r t) \cos(\omega_r t)]_{T_1}^{T_2} \\ &\quad - [\sin(\omega_r t) \cos(\omega_r t)]_{T_2}^{T_3} + [\sin(\omega_r t) \cos(\omega_r t)]_{T_3}^{m\tau/2} \end{aligned} \quad (17)$$

$$\begin{aligned} &= - 2 \sin(\omega_r T_1) \cos(\omega_r T_1) + 2 \sin(\omega_r T_2) \cos(\omega_r T_2) \\ &\quad - 2 \sin(\omega_r T_3) \cos(\omega_r T_3) + \sin(\omega_r m\tau/2) \cos(\omega_r m\tau/2). \end{aligned} \quad (18)$$

Equating Equations 15 and 18 gives:

$$\begin{aligned}
 & -2\sin(\omega_r T_1) + 2\sin(\omega_r T_2) - 2\sin(\omega_r T_3) + \sin(\omega_r m\tau/2) \\
 = & -2\sin(\omega_r T_1)\cos(\omega_r T_1) + 2\sin(\omega_r T_2)\cos(\omega_r T_2) - 2\sin(\omega_r T_3)\cos(\omega_r T_3) \\
 & + \sin(\omega_r m\tau/2)\cos(\omega_r m\tau/2)
 \end{aligned} \tag{19}$$

$$\begin{aligned}
 0 = & -2\sin(\omega_r T_1)[1 - \cos(\omega_r T_1)] + 2\sin(\omega_r T_2)[1 - \cos(\omega_r T_2)] \\
 & - 2\sin(\omega_r T_3)[1 - \cos(\omega_r T_3)] + \sin(\omega_r m\tau/2)[1 - \cos(\omega_r m\tau/2)].
 \end{aligned} \tag{20}$$

Recognising that $\tau = 2\pi/\omega_r$ allows us to eliminate the final term in Equation 20 (and cancel the factors of 2) to give

$$\begin{aligned}
 & -\sin(\omega_r T_1)[1 - \cos(\omega_r T_1)] + \sin(\omega_r T_2)[1 - \cos(\omega_r T_2)] \\
 & - \sin(\omega_r T_3)[1 - \cos(\omega_r T_3)] = 0,
 \end{aligned} \tag{21}$$

from which the T_i can be calculated. A more general form of Equation 21 for $2n$ π -pulses is [13]

$$\sum_{i=1}^n (-1)^i \sin(\omega_r T_i)[1 - \cos(\omega_r T_i)] = 0. \tag{22}$$

We seek solutions to Equation 22 where $\xi = 0$. This gives *only* anisotropic interactions in the indirect dimension. For six π pulses, there are two continua of timings for $\xi = 0$ that give $-0.393 \leq \chi \leq -0.201$ or $-0.121 \leq \chi \leq 0.363$ [13]. This allows us to choose a value of χ so the powder pattern fits the spectral width in the t_1 dimension, avoiding truncation of the lineshape.

3.1.1 Pulse sequence for idealised pulses

For idealised pulses — where there are no phase imperfections and the RF field created by the pulses is the same throughout the sample volume — the pulse sequence consists of three elements (Figure 4):

1. Excitation of magnetisation on ^1H , and cross polarisation³ to ^{13}C

³Cross-polarisation (CP) transfers magnetisation from one nucleus to another, most commonly from ^1H to the observed nucleus. It is used extensively in solid-state NMR.

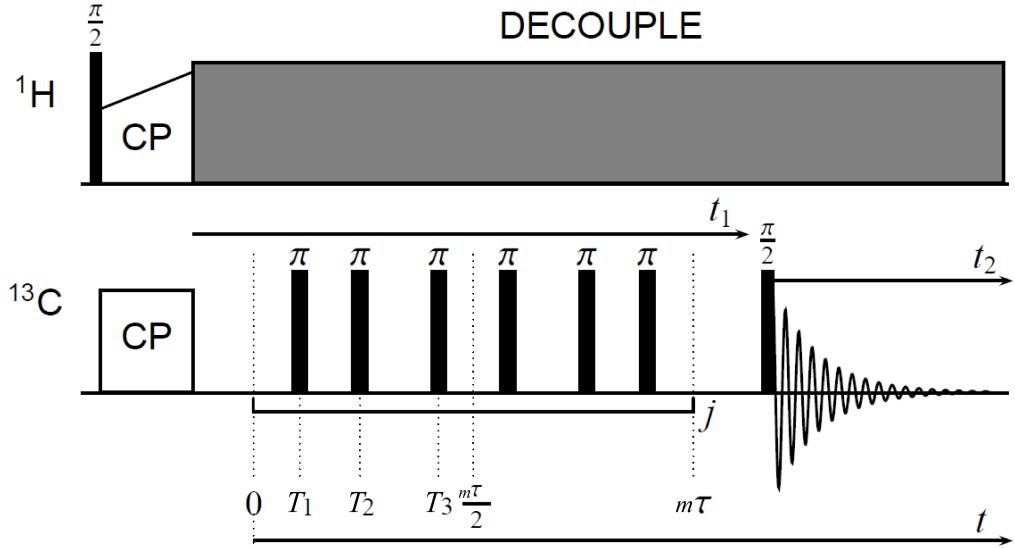


Figure 4: 2DSCA pulse sequence, adapted from Orr [7].

2. The recoupling sequence (t_1)
3. Acquisition (t_2)

3.2 Phase Cycling

Unfortunately, pulses are not perfect in a real experiment. We can minimise the effect of phase imperfections using *phase cycling*, where the experiment is repeated and pulse phases are changed systematically. Signal from the desired coherence order constructively interferes, while undesirable coherence cancels once the cycle is complete.

3.2.1 Nested Phase Cycling

Nested phase cycling involves changing the phase of each pulse in turn [14–16]. The number of steps for each pulse is determined by the coherence order change that the pulse produces. A π -pulse involves a change in coherence order from $+1$ to -1 or vice versa. An isolated spin-half nucleus such as ^{13}C can generate coherence states 0 and ± 1 , so an imperfect π -pulse can lead to any of these

three. To avoid unwanted coherence, the experiment can be repeated and the pulse phases cycled. For a π -pulse, a three-step cycle is required to select a particular coherence [16]. For n π -pulses, nested phase cycling requires a 3^n -step cycle. For the 2DSCA experiment the number of steps is very large; nested phase cycling is inappropriate.

3.2.2 Cogwheel Phase Cycling

Cogwheel phase cycling drastically reduces the number of steps required for a pulse sequence [16]. It achieves this by incrementing the phase of the pulses simultaneously. It has been applied to the 2DCSA experiment [7, 11] to give a phase-cycled 2DCSA experiment with a viably small number of steps.

3.3 Optimised and Constant-Time 2DCSA Experiments

Orr [7, 11] has developed the cogwheel phase-cycled 2DCSA experiment with ‘optimised’ and ‘constant-time’ versions.

3.3.1 Constant-Time Experiment

Rearrangement of Equation 10 gives the net precession angle ϕ over one pulse sequence unit:

$$\phi = m\tau(\zeta\Delta\nu) + m\tau\chi[C_1(\Omega) + C_2(\Omega)]. \quad (23)$$

ϕ can be incremented by increasing the number of pulse sequence units j (as in the conventional experiment) or by changing χ . The constant-time experiment uses the same number of pulse sequence units for each t_1 ‘increment’ (or t_1 ‘slice’), but the pulse timings are changed to give a different value of the anisotropic scaling parameter χ . The ‘effective’ value of t_1 is thus changed between experiments by changes in χ . The advantage of the constant-time experiment is that it reduces line-broadening in the t_1 dimension. Residual ^1H – ^{13}C dipolar coupling leads to dephasing of magnetisation during t_1 . The amount of signal-loss increases with t_1 ; thus incrementing t_1 by incrementing j leads to progressively worse signal-loss. Upon Fourier transformation, this leads to line-broadening in the t_1 dimension. If

j is constant, the signal-loss will be the same for each slice. No line-broadening will be seen upon Fourier transformation.

3.3.2 Optimised Experiment

The constant-time experiments leads to long pulse sequences with many π -pulses which tends to reduce signal-to-noise. The optimised experiment uses the minimum possible π -pulses for each slice, maximising signal-to-noise. χ and j (the number of pulse sequence units) are calculated from [11]

$$j(k) = \text{Ceil} \left(\frac{\nu_r(k-1)}{SW\chi_{\max}m} \right) \quad 1 \leq k \leq k_{\max} \quad (24)$$

$$\chi(k) = \frac{\nu_r/mSW(k-1)}{\text{Ceil} \left(\frac{\nu_r(k-1)}{SW\chi_{\max}m} \right)} \quad 1 \leq k \leq k_{\max}, \quad (25)$$

where k denotes the t_1 slice and k_{\max} is the total number of t_1 slices. SW is the spectral width in t_1 , $m = 3$, $\chi_{\max} = 0.242$ and $\nu_r = \omega_r/2\pi$. The optimised experiment is used in this work.

4 Experimental Details

4.1 Recording of Spectra

All of the solid-state experiments described in this work were carried out on one of two Bruker Avance 400 MHz spectrometers, using 4mm zirconia rotors with the volume not restricted to the centre of the rotor. Where possible, double-resonance probes were used for the 2DCSA experiments: the experiment only requires irradiation of ^1H and ^{13}C and using a double-resonance probe gives better sensitivity than using a triple-resonance probe in double-resonance mode. However, for some experiments only a triple-resonance probe was available. All experiments were carried out at a spinning speed of 10 kHz unless otherwise stated.

Referencing of the spectra was performed using the α -carbon of glycine, which has an isotropic shift of 43.1 ppm.

To produce spectra of appropriate resolution to obtain good CSA lineshapes, it is necessary to separate the carbonyl region (around 170 ppm) and the aliphatic region (0–80 ppm). If both regions were of interest, two spectra were recorded, one centered on the carbonyl region and one centered on the aliphatic region. This was achieved by setting the RF transmitter frequency according to the region of interest (approximately 175 ppm for the carbonyl region and 40 ppm for the aliphatic region). Because carbonyl CSA lineshapes are wider than those of aliphatic carbons (carbonyl groups are planar, so the difference between in chemical shift between in- and out-of-plane components is large) the spectral width of the indirect dimension was chosen to be wider for the carbonyl spectra (32 kHz compared to 16 kHz). This corresponds to using a different value of χ — the pulse timings (calculated by Orr [7, 11]) in the recoupling sequence differ.

In general, fewer scans were recorded per t_1 increment of the 2DCSA experiment than were recorded for the CP spectrum (due to time constraints), and the long recoupling sequence allows relaxation. Hence the 2DCSA spectra are often noisier than the CP spectra. The powder-patterns were obtained by taking a slice through the 2DCSA spectrum. The slices were chosen to have the maximum signal-to-noise in the indirect dimension. It was therefore often the case that

the isotropic shift in the direct dimension where the slice was taken was slightly different to the isotropic shift of the peak in the CP spectrum.

Many of the 2DCSA spectra were produced with the help of Wing Ying Chow. In particular, those of $(\text{Gly-Pro-Hyp})_{11}$, poly(Pro-Gly-Pro) and the aliphatic region of $(\text{Pro-Hyp-Gly})_{10}$ were performed exclusively by her. The solution state spectra in this work were performed by the Chemistry Department NMR Service.

4.2 Samples

The collagen model peptides $(\text{Gly-Pro-Hyp})_{11}$ (labelled at Gly 7) and $(\text{Gly-Pro-Pro})_{11}$ (labelled at the second Proline in the seventh triplet, i.e. $(\text{Gly-Pro-Pro})_6-(\text{Gly-Pro-Pro}^*)-(\text{Gly-Pro-Pro})_4$) were obtained from Dr Dominique Bihan, Department of Biochemistry. The model peptide $(\text{Pro-Hyp-Gly})_{10}$ was obtained from Peptides International and used without further purification. A sample of collagen from sheep osteoblast was obtained from Dr Rakesh Rajan, Addenbrookes. Osteoblast cells were cultured for seven days; the final three days with a source of ^{13}C and ^{15}N labelled proline. The polymeric peptide samples were obtained from Sigma-Aldrich and used without further purification. Poly(L-Proline) had molecular weight 1000–10000, corresponding to a degree of polymerisation of roughly 10–100. Poly(Pro-Gly-Pro) had molecular weight 2000–10000, corresponding to a degree of polymerisation of 8–40 (i.e. between 8 and 40 Pro-Gly-Pro triplets). Ala-Pro-Gly and bovine achilles tendon collagen were obtained from Sigma-Aldrich and used without further purification.

4.3 Fitting of 2DCSA Lineshapes

In this work, the fitting process used the Bruker NMR software *Topspin*. Versions 2.1 (28.02.2010) and 3.0 (07.04.2010) were used. Unless stated otherwise, it was assumed that only one powder pattern contributes to each lineshape. The procedure was as follows:

1. Start the ‘Solids Lineshape Analysis’ module in *Topspin*.

2. Set the ‘spinning speed’ to zero to simulate a static powder pattern.
3. Add a new nucleus.
4. Fit the lineshape as well as possible by eye — the fitting converges faster if the initial lineshape is similar to the experimental lineshape.
5. Apply an appropriate amount of line broadening — all of the experimental powder patterns appear broadened to some degree.
6. Start the optimisation.

The quality of the fit is measured as the overlap of the two curves, given as a percentage. The fitting module offers two algorithms, ‘simplex’ and ‘genetic’. Sometimes, a larger overlap is obtained by using both of the algorithms sequentially: the genetic algorithm gives a better ‘starting powder pattern’ than fitting by eye, then the simplex algorithm refines this. Every result quoted in this work used the ‘simplex’ algorithm as the *final* procedure — this gives larger overlap than using the genetic algorithm last.

The output of the procedure gives δ_{iso} , δ_{11} , δ_{22} and δ_{33} , as well as the anisotropy parameters Δ and η (defined in Equations 2 and 3). The principal components are centred around the isotropic shift in the *indirect* dimension (recall that this should be zero for an experiment with no pulse imperfections). For comparison between samples the components need to be centred around the isotropic shift in the *direct* dimension (as they would be if an isolated static powder pattern could be recorded). The principal components are thus adjusted by addition of $(\delta_{iso}^{\text{direct}} - \delta_{iso}^{\text{indirect}})$. Δ and η are invariant under this transformation.

4.3.1 Fitting of Lineshapes with Multiple Components

It is sometimes likely that the observed lineshape is the sum of two or more powder patterns. *Topspin* allows the optimisation process to be performed with multiple lineshapes. It also allows the choice of which parameters are allowed to vary in the fitting procedure. Fitting multiple lineshapes was performed in two ways in this work:

- Simulation of a two-component lineshape using two appropriate sets of initial components (chosen as the measured components from compounds structurally analogous to those thought to comprise the pattern to be fitted), with Δ and η *not* allowed to vary. The two lineshapes stay the same, with their relative intensities and isotropic shifts varying. This gives an idea of whether an observed pattern can feasibly be made up of the two initial tensors.
- Simulation of a two-component lineshape using the same starting components as before, but with Δ and η allowed to vary. This tells us whether there is a significantly better fit with different components to those we have assumed.

Neither of the two methods is vastly superior to the other, and useful conclusions can be drawn from both.

4.4 Error Analysis

In theory, the 2DCSA experiment produces undistorted static-like powder patterns, which should very reliably give the principal shift components. The fitting procedure is fairly robust in the sense that it tends to converge on the same principal values each time if run multiple times on the same experimental lineshape (with different initial lineshapes). However, this does not mean that the values obtained are necessarily accurate. Uncertainty in the fitting comes from:

- Pulse imperfections leading to line broadening in the indirect dimension. The larger the degree of line broadening, the less effect a small change in one of the principal components has on the simulated lineshape. We can attempt to gauge the extent of pulse imperfections by looking at δ_{iso} in the indirect dimension. In theory $\delta_{iso} = 0$ for an isotropic scaling factor $\xi = 0$. If the measured average $|\delta_{iso}|$ is larger than the error in δ_{iso} this tells us there are at least some pulse errors.
- Poor signal-to-noise ratio. The inherently broad lines in a 2DCSA spectrum

lead to poor signal-to-noise ratio, so even after several repetitions of the cogwheel cycle the ratio can be poor. For N scans, the signal scales with N and the noise with \sqrt{N} , so repeating the experiment to get very good signal-to-noise is often impractical. Lineshapes from carbonyl ^{13}C are wider than those from aliphatic ^{13}C (Δ is larger), and so suffer from lower signal-to-noise than aliphatic spectra.

To obtain an estimate of the uncertainties in the fitting procedure, powder patterns were re-simulated with the values obtained from the fitting, then the values of δ_{iso} , Δ and η systematically increased (or decreased) in turn until the simulated pattern was judged to no longer fit the experimental lineshape. These tolerable differences in fitting were used to provide approximate error bounds for the data obtained. *Topspin* does not give the option of altering the principal components directly. Because the powder patterns for carbonyl signals are wider than those for aliphatic carbons, the absolute errors are correspondingly larger. The procedure was as follows:

1. Begin the ‘simulation’, starting with the recorded CSA fitting lineshape parameters.
2. Leaving all other parameters the same, change δ_{iso} to the *smallest* tolerable value as judged by eye.
3. Start an ‘optimisation’ with zero steps; this outputs a lineshape with the chosen δ_{iso} and gives the values of the principal components.
4. Repeat for the *largest* tolerable value for δ_{iso} .
5. Repeat the above steps for Δ and η .

This gives many tolerable values for the principal components. An estimate of the error in each component is made by taking the average value of the *absolute* difference between the tolerable values and the initial values:

$$\text{Estimate of Error} = \sum_n^1 |\delta_{AA}^{\text{tolerable}} - \delta_{AA}^{\text{initial}}| \quad (26)$$

where δ_{AA} is a principal component.

It is important to note that the uncertainty in δ_{iso} is *not* the uncertainty in the shift measured in the direct dimension — for a properly referenced spectrum this error is negligible in comparison — it tells us the uncertainty in the fitting process for δ_{iso}

5 NMR Studies of Poly(L-Proline)

The native state of collagen is triple-helical. As such, much of this work involves studies of triple-helical peptides to develop a diagnostic for a triple helix, which is important for characterising synthetic samples. The individual chains of the triple helix are of the polyproline II type [4, 5], so it is necessary to have representative ^{13}C CSA tensors for a single-helical polyproline II helix for comparison.

The polyproline II helix is common in globular proteins [17]. It is a left-handed helix consisting of proline residues with dihedral angles of approximately $\phi = -75^\circ$, $\psi = 145^\circ$ and *trans* peptide bonds [17–19]. There is a degree of tolerance to non-proline residues in the structure [18, 19]; in nature, the helices are very rarely formed solely from proline. There is another type of polyproline helix — the polyproline I helix is right-handed and has *cis* peptide bonds. In the absence of solvent effects, the polyproline II helix is lower in energy [19], so a pure proline polymer is expected to adopt this conformation in the solid state. Polyproline is hence a good sample to study the ^{13}C CSA tensors of a single helix.

5.1 ^{13}C Cross-Polarisation (CP) Spectrum of Polyproline

The ^{13}C CP spectrum of polyproline (Figure 6) shows two interesting features:

- The carbonyl carbon has a very low isotropic shift of 170.2 ppm. This is much lower than the ‘random coil’ proline carbonyl shift of approximately 174 ppm⁴ [20], and lower than the shifts in triple helices (section 7).
- The γ -carbon signal at 24 ppm is broad, and appears to be split into at least three separate signals.

5.2 γ -Carbon Splitting

Further ^{13}C solid-state NMR experiments were performed to determine whether the splitting of the γ -carbon could be better resolved at a faster spinning rate.

⁴Shifts for the residue X measured on pentapeptide Gly-Gly-X-L-Ala-OH in aqueous solution

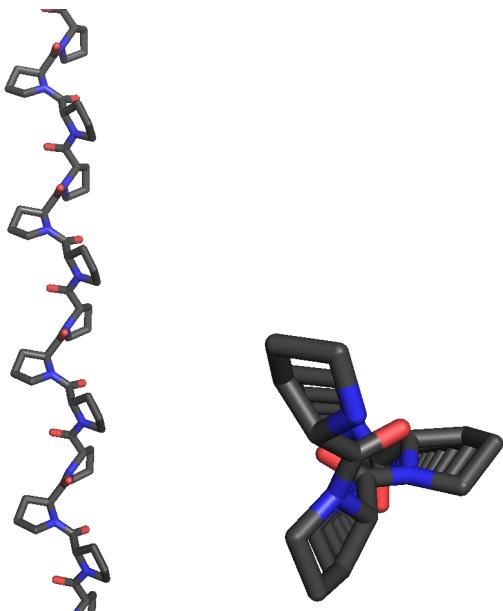


Figure 5: Molecular model of the polyproline II helix, viewed from the top and from the side. Note the three-fold symmetry. Images produced using the PyMOL program, www.pymol.org.

Spectra at MAS rates of 10 kHz and 14 kHz were performed (Figure 7)⁵. The 14 kHz spectrum does not have noticeably narrower signals than the 10 kHz spectrum, but both seem to show a four-fold splitting of the peak.

The pyrrolidine rings in polyproline II exist in the ‘half-chair’ conformation [21]; the γ -carbon is the most mobile vertex of the ring. There are two possible conformations of the ring. From a model of the helix (Figure 8), it is clear why ring-flip should cause a twofold splitting in the γ -carbon — in one conformation the γ -carbon points *into* the helix and in the other it points *out*. Why this should cause the observed four-fold splitting is not obvious; the following are possibilities:

- The γ -carbon could interact with the pyrrolidine ring of the next residue

⁵The spectra were run on a different spectrometer than was used for the spectrum in Figure 6, so the 10 kHz spectrum was repeated. 14 kHz was chosen as it is 1 kHz within the 15 kHz operating range of the rotor.

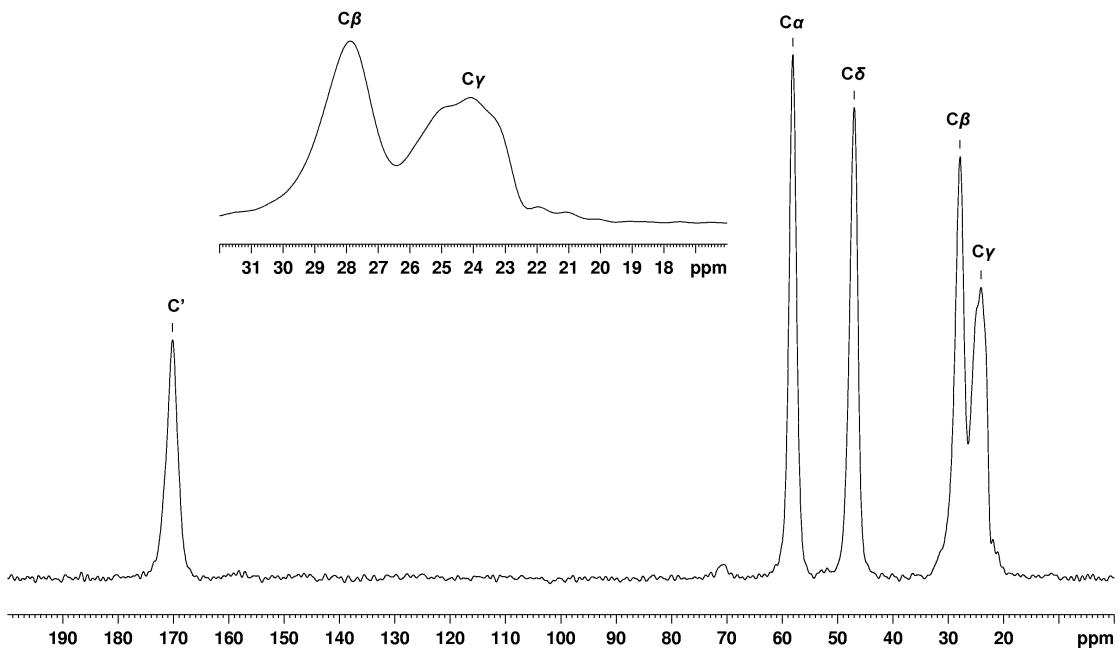


Figure 6: ¹³C CP spectrum of polyproline with an expansion of the region containing the β - and γ -carbon shifts. Note the apparent splitting in the γ -peak.

in the chain and be affected by its conformation — in the ‘up’ position interacting with the next residue ‘up’ the chain, and in the ‘down’ position interacting with the next residue ‘down’ the chain. However the geometry of the molecule (Figure 8) suggests this is unlikely.

- Packing effects — the γ -carbon is the most exposed atom in the chain; this is expected to be the most vulnerable to the effects of other molecules in the crystal.
- End effects — we expect the very end of the helix to be different to the central portion. Thus the shift may differ for residues near the end; the ‘splitting’ could be in fact multiple environments. However it is extremely unlikely that end effects will be significant for a polymer of this length, and we would expect more than one environment to be affected.

Rigorously testing these hypotheses is outside the scope of this work. However if the four-fold splitting was still present in solution would rule out packing effects.

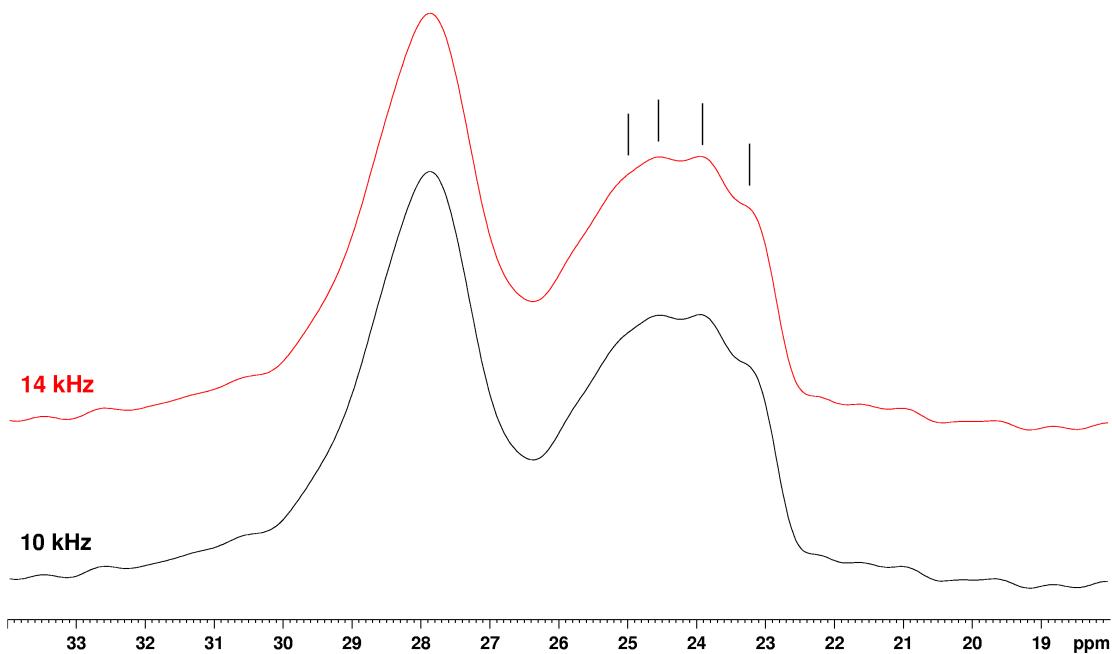


Figure 7: The β - and γ -carbon region of the ^{13}C CP spectrum of polyproline, recorded at spinning speeds of 10 khz and 14 kHz. The possible four-fold splitting is indicated by the vertical lines.

The solution state spectrum (Figure 9) does not show a four-fold splitting, so packing effects cannot be ruled out. Importantly, the *lack* of a four-fold splitting in solution does not *confirm* the existence of packing effects in the solid state.

5.2.1 Implications for Collagen

The ring-flip between the two conformations is slow on the NMR timescale (the signals do not coalesce) at room temperature, indicating a high activation energy barrier. If the same motion occurs in collagen, we believe it could contribute to absorbing energy of impacts, accounting for some of the mechanical strength of collagen.

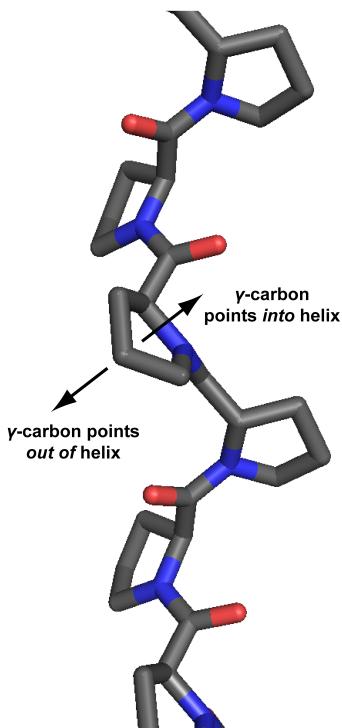


Figure 8: Side view of polyproline, showing the effect of the differences in ring conformation on the γ -carbon

5.3 ^{13}C 2DCSA of Polyproline

The proline carbonyl ^{13}C isotropic shift in polyproline is very low (170.2 ppm). Fitting of the 2DCSA lineshape (Table 1) gives a very low δ_{22} component of 173.8 ppm — the difference in isotropic shift between polyproline and the triple-helical peptides (section 7) is entirely due to this component. The δ_{22} component points along the carbon-oxygen bond vector [22]. Charge rotation of the oxygen lone pairs into the C–O π^* orbital (the largest coefficient of which is on the carbon) contributes to the paramagnetic deshielding of this component. The size of the paramagnetic effect depends on $1/r^3$ and $1/\Delta E$, where r is the average distance of the electrons from the nucleus and ΔE is the energy separation of the rotating orbitals. If the lone pairs are involved in hydrogen bonding:

- They are on average further away from the carbon atom.

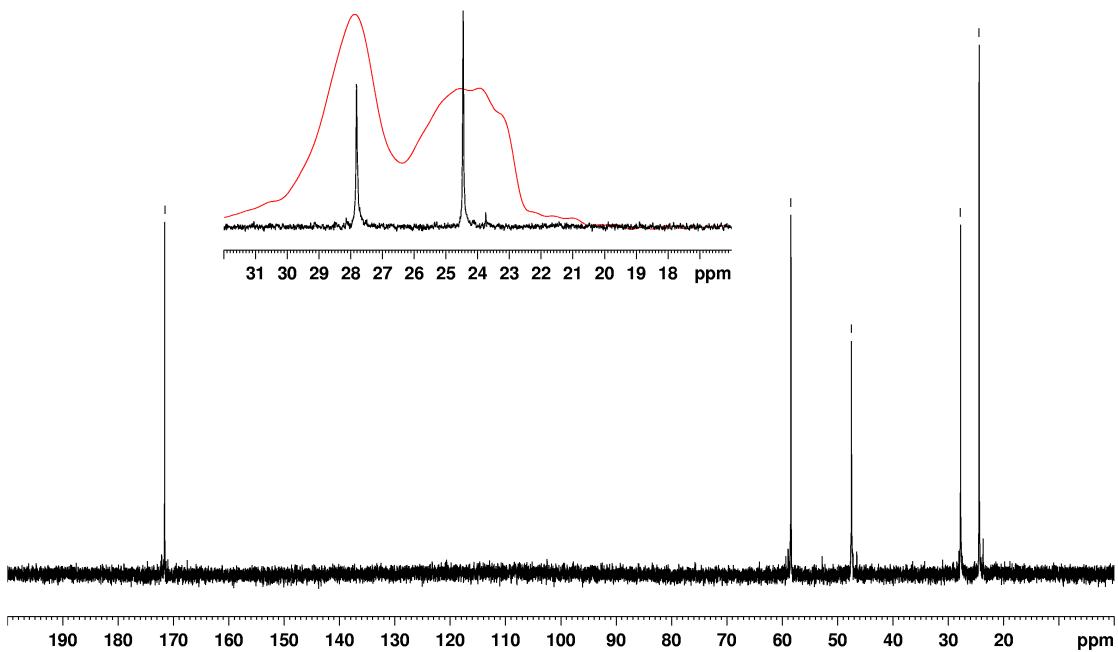


Figure 9: Solution state ^{13}C spectrum of polyproline in D_2O , with an expansion of the β and γ carbon peaks. The solid state spectrum is overlaid in red. The lack of four-fold splitting in solution means we cannot rule out packing effects.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
170.02	Pro C'	243	174	94	-76	0.90	1204	92.51
58.07	Pro C α	80	59	35	-23	0.88	725	95.91
47.02	Pro C δ	75	49	17	-30	0.84	555	-
27.90	Pro C β	47	24	13	19	0.57	671	96.01
24.16	Pro C γ	47	15	11	23	0.21	611	93.59

Table 1: ^{13}C 2DCSA fitting values for polyproline. For uncertainties, see Error Analysis, section 7.4.

- Their energy is lowered; ΔE increases.

These effects reduce the paramagnetic deshielding and lower the shift of the δ_{22} component. The carbonyl groups point to the outside of the helix; it is very likely that they are involved in hydrogen bonding with water of crystallisation.

6 NMR Studies of Poly(Pro-Gly-Pro)

Poly(Pro-Gly-Pro) is similar to collagen in its primary structure, and a monodisperse sample is expected to form stable triple-helices. Indeed, a crystal structure of $[(\text{Pro-Pro-Gly})_{10}]_3$ is available [23] (PDB: 1K6F⁶). Poly(Pro-Gly-Pro) has been shown to form triple-helical structures [4, 5]. A polydisperse sample is likely to have triple-helical and single-strand sections because the lengths of the chain sections that aggregate will not match exactly. The sample studied here was $(\text{Pro-Gly-Pro})_n$ where $n = 8 - 40$. Studies of Poly(Pro-Gly-Pro) allow observation of collagen-like chains in both the triple- and single-helical forms⁷.

6.1 ^{13}C CP Spectrum of Poly(Pro-Gly-Pro)

Of particular interest in the ^{13}C CP spectrum (Figure 10) is the carbonyl region; the proline signal has shoulders at 173.4 ppm and 172.4 ppm in addition to the main peak at 171.2 ppm. By comparison to the fully triple-helical peptides presented in section 7 — which show proline carbonyl shifts at approximately 172 ppm and 173.5 ppm — we see that there are indeed triple-helical regions present. It appears that the single-helical environment at 171.2 ppm is present in larger quantities. Importantly, it is clear that it is possible to tell the difference between triple- and single-helical structures using solid-state NMR.

6.2 Solution-State Studies of Poly(Pro-Gly-Pro)

To try to clarify the extent to which polydispersity in the sample creates different NMR environments, solution-state studies were performed in D_2O . Of interest was the ^{13}C spectrum, which showed many more peaks than we would expect for a structurally-homogeneous sample (c.f. polyproline, Figure 9). It also showed the presence of impurities.

A ^1H spectrum (Figure 12) was recorded along with the ^{13}C , which showed signals in the region 7.5 ppm and 9 ppm. It was thought that these could be

⁶Protein data bank, www.pdb.org

⁷It is assumed that the single-strand adopts a helical conformation similar to polyproline II

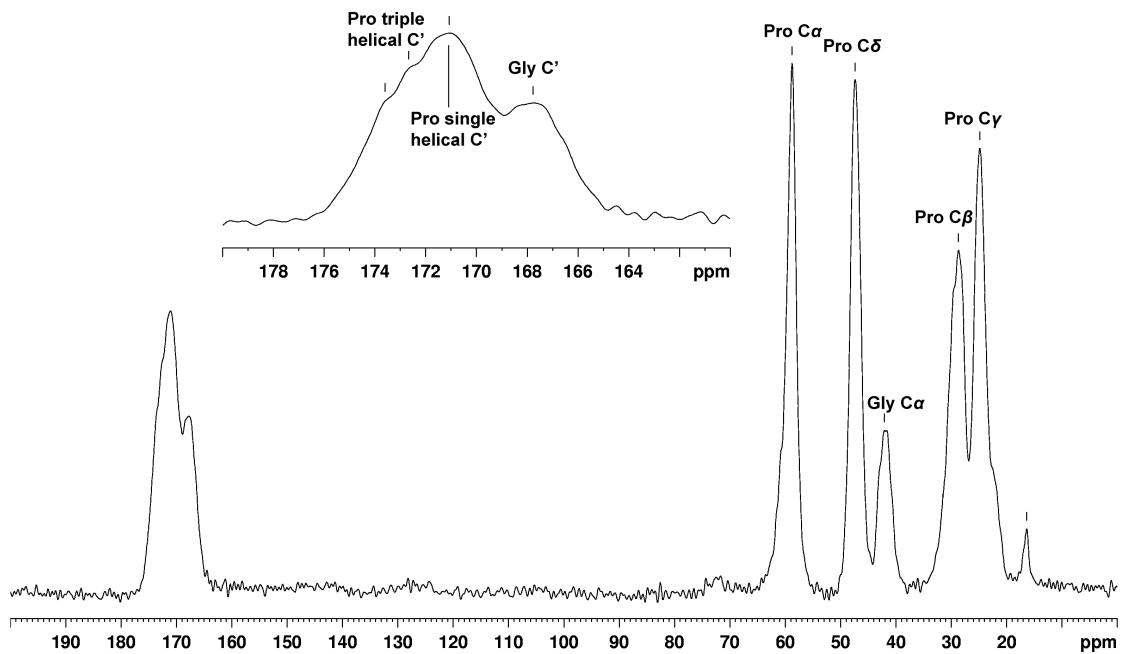


Figure 10: ¹³C CP spectrum of Poly(Pro-Gly-Pro), with expansion of the carbonyl region.

Note the shoulders on the proline carbonyl peak at 173.60 ppm and 172.67 ppm, corresponding to triple-helical regions. The small peak at 16.42 ppm is assumed to be an impurity.

glycine amide protons of the triple-helix — exchanging very slowly with D₂O due to strong hydrogen-bonding across the triple-helix. To test whether this could be the case, or if the signals were impurities, further spectra were recorded: after leaving for 24 hours, and then after heating with a heat gun (Figure 13).

Neither leaving for 24 hours nor heating appreciably changed the signals in the 7.5–9 ppm region. Strong heating disturbs the triple-helix so even strongly hydrogen-bonded protons would exchange; the peaks in the 7.5–9 ppm region are hence assigned as impurities. It is possible that the impurities in both the ¹³C and ¹H spectra can be accounted for if the Fmoc protecting group (Figure 14) was used in the synthesis of the poly(Pro-Gly-Pro) sample.

To see where the amide protons do appear, spectra in H₂O were recorded (using strong solvent suppression) at varying temperatures (Figure 15). The spec-

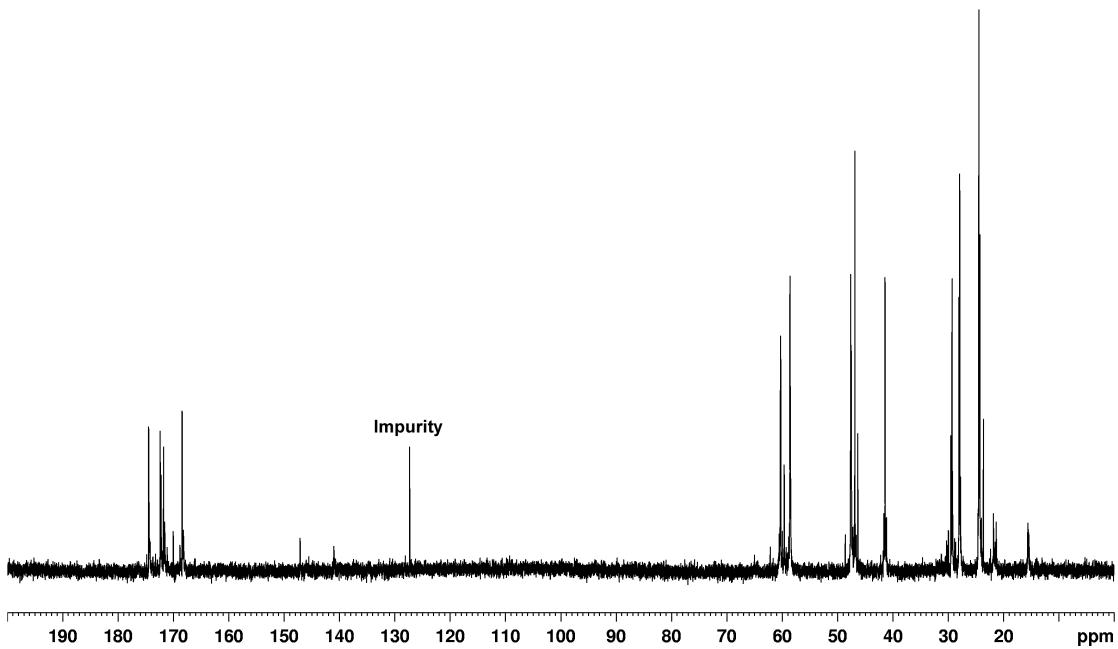


Figure 11: Solution-state ^{13}C spectrum of Poly(Pro-Gly-Pro) in D_2O . Note the large numbers of peaks indicating high structural inhomogeneity. The signal at 127.35 ppm does not correspond to any of the peptide residues and is labelled ‘impurity’.

tra were referenced to DSS⁸. New signals appeared, corresponding to amide protons. The chemical shift of around 8 ppm and the temperature dependence $\Delta\delta = -8.5 \text{ ppb K}^{-1}$ both suggest these are amide protons from single-strand regions [24]. This supports the conclusion from the solid-state spectrum that the majority of the peptide is single-stranded.

6.3 ^{13}C 2DCSA of Poly(Pro-Gly-Pro)

Slices of the ^{13}C 2DCSA spectrum of poly(Pro-Gly-Pro) suggest that there are indeed triple-helical regions present — the ^{13}C principal chemical shift components of the signals at 173.4 ppm and 172.4 ppm resemble closely those of the other triple helical peptides studied (section 7). The principal components of the single-

⁸DSS is a water-soluble reference compound similar to TMS. One of the four methyl groups present in TMS is replaced with $(\text{CH}_2)_3\text{SO}_3\text{H}$, the sulfonic acid group conferring water-solubility. The chemical shift of the TMS-like protons is defined as zero.

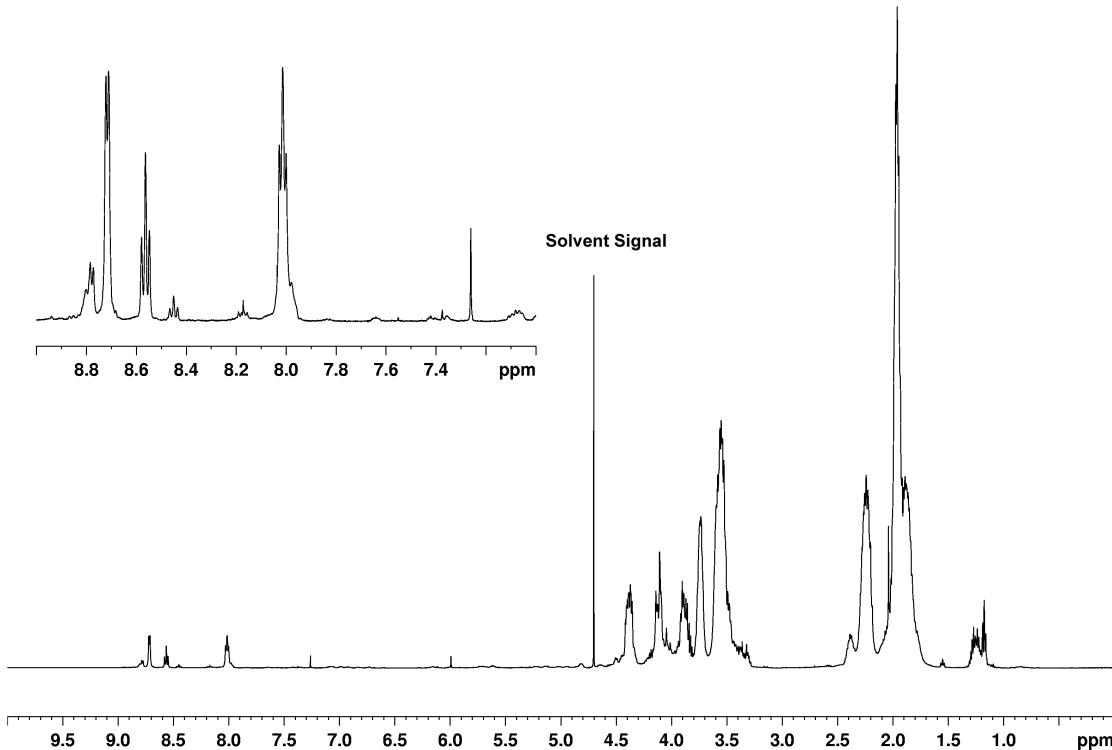


Figure 12: Solution-state ¹H spectrum of Poly(Pro-Gly-Pro) in D₂O. The solvent signal at around 4.7 ppm is H₂O produced from proton exchange with the solute. An expansion of the region with the possible amide peaks is plotted.

helical peak is similar to that of polyproline — it is not expected to be identical because of the presence of the glycine residues.

6.3.1 Fitting of 2DCSA Spectrum with Multiple Components

The fact that the signals corresponding to proline residues are not fully resolved from one another means that each powder lineshape in the indirect dimension will in principle contain elements from all three proline environments. To a first approximation it is assumed that the principal ‘impurity’ component in the triple-helical lineshapes is from the single-helical signal. The lineshapes were fitted to two powder patterns as described in section 4.3.1. The principal components of

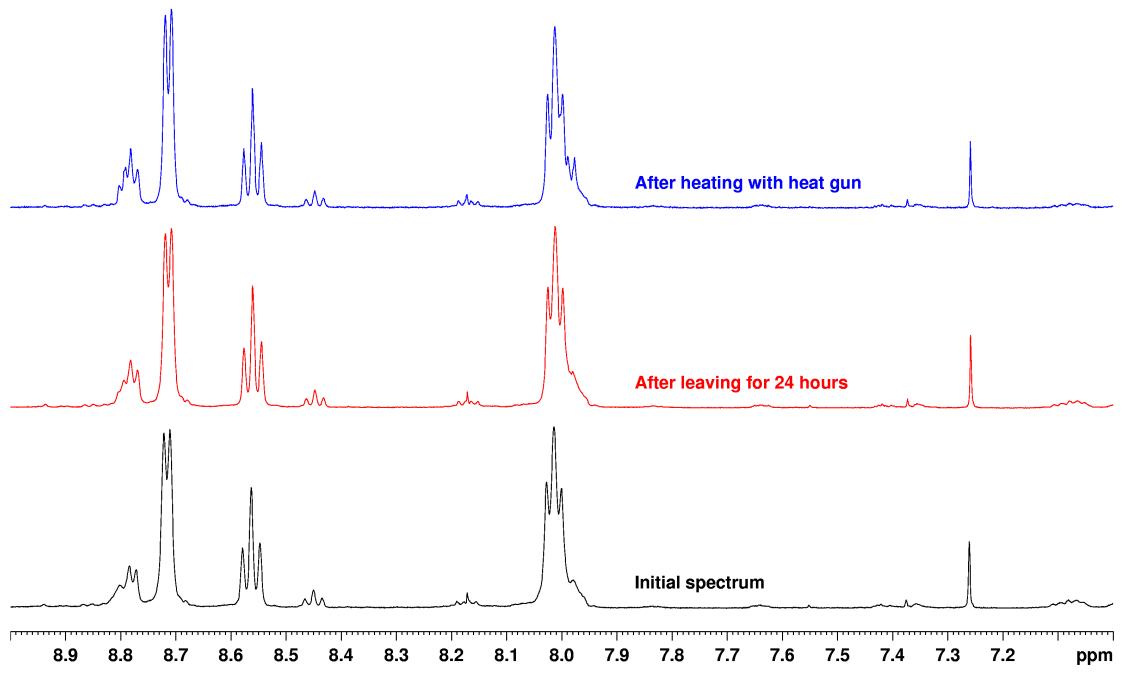


Figure 13: Solution-state ^1H spectra of Poly(Pro-Gly-Pro) in D_2O after 24 hours and after heating, showing the region between 7 and 9 ppm. The signals change very little, so are assumed to be impurities.

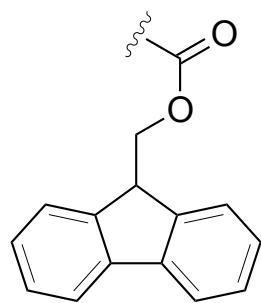


Figure 14: The FMOC protecting group. The aromatic carbon atoms could account for the signal at 127.35 ppm in the ^{13}C spectrum, and the aromatic protons could account for the signals around 8 ppm in the ^1H spectrum.

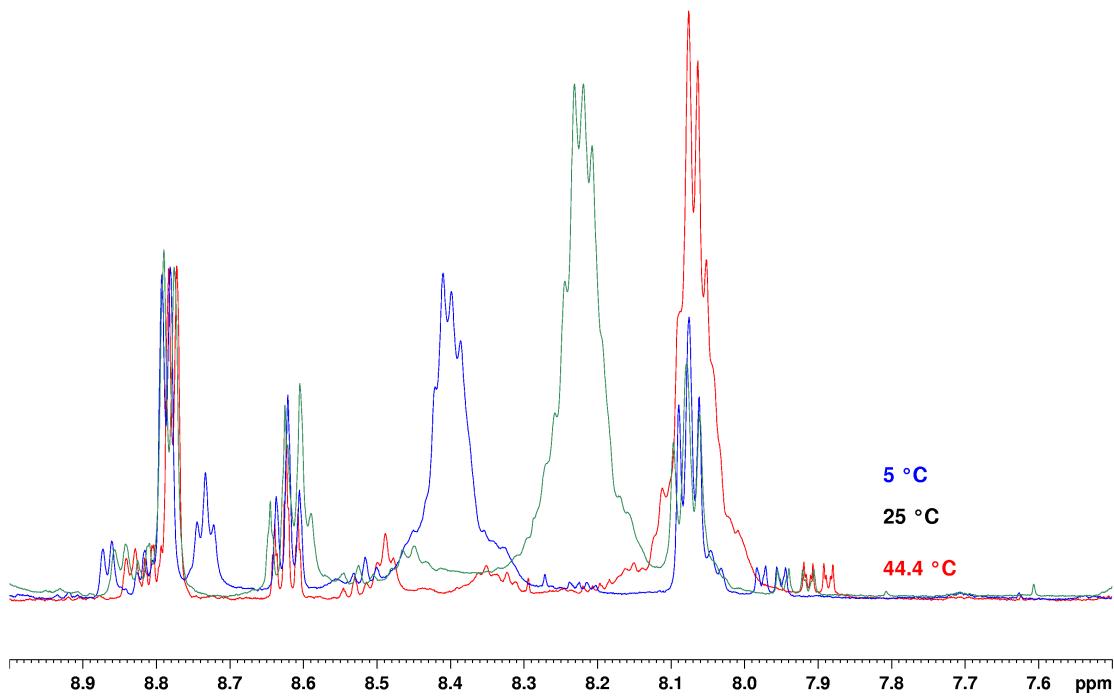


Figure 15: Variable temperature solution state spectra of Poly(Pro-Gly-Pro) in H_2O . The intensities are normalised to the peak at 8.8 ppm. The change in shift of the amide protons is clearly seen.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.43	Pro C'	244	181	95	-79	0.80	883	93.67
172.39	Pro C'	238	183	96	-76	0.72	930	94.82
171.17	Pro C'	237	180	96	-75	0.77	854	97.12
167.98	Gly C'	238	172	95	-73	0.90	1106	94.70
58.75	Pro C α	81	60	35	-24	0.86	1291	93.24
47.35	Pro C δ	72	48	22	-25	0.98	1174	94.27
41.94	Gly C α	72	40	14	30	0.88	729	93.18
28.62	Pro C β	48	26	12	19	0.73	1291	93.32
24.71	Pro C γ	45	21	7	21	0.68	1169	93.61

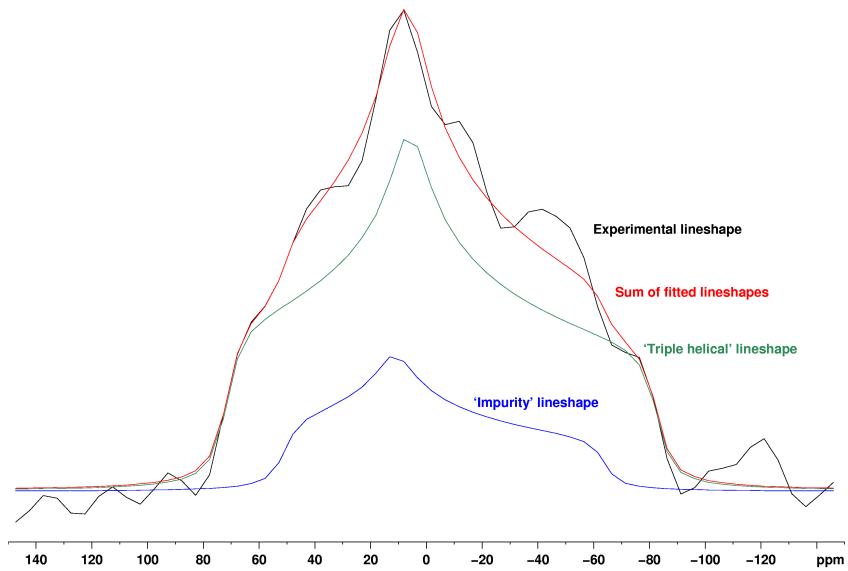
Table 2: ^{13}C 2DCSA fitting values for poly(Pro-Gly-Pro). For uncertainties, see Error Analysis, section 7.4.

the relevant ^{13}C environments in $(\text{Gly-Pro-Pro})_{11}$ and $(\text{Pro-Hyp-Gly})_{10}$ were used as starting values for the lineshapes at 173.4 and 172.4 ppm respectively. Polyproline was used for starting values for the other component — the lineshape at 171.1 ppm is unsuitable as it contains contributions from the triple-helical environments; polyproline was the only other single-helical environment where principal components were measured.

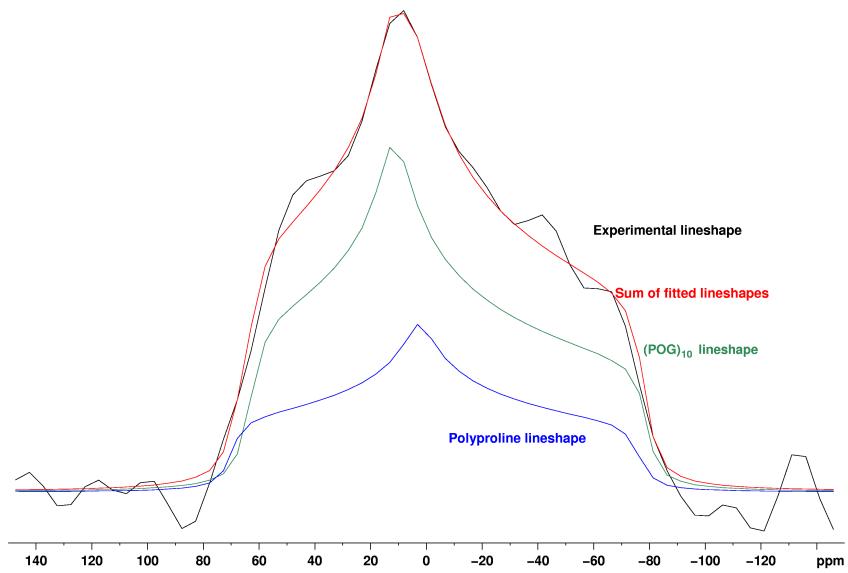
Lineshapes were simulated using both the fixed and variable methods described previously (Figure 16). For the 173.4 ppm signal, a more believable fit was obtained using the variable method; the converse was true for the 172.4 ppm signal. These are compared to the single-environment fits in Figure 17.

- For the 173.4 ppm signal, the ‘impurity’ is very different to polyproline; also the ‘improved’ fitting seems to simply fit the noise of the spectrum better.
- For the 172.4 ppm signal, the lineshape is fitted very well by a combination of $(\text{Pro-Hyp-Gly})_{10}$ and polyproline, as expected since $(\text{Pro-Hyp-Gly})_{10}$ is very similar to triple-helical poly(Pro-Gly-Pro). However, the overlap of the multi-component fit exceeds that of the single-component by only 0.2 %. This was not considered significant, especially with the approximation that the single-helical ‘impurity’ has the same CSA principal components as polyproline.

The results of the single-lineshape fittings are therefore used in later sections for comparison to other compounds.

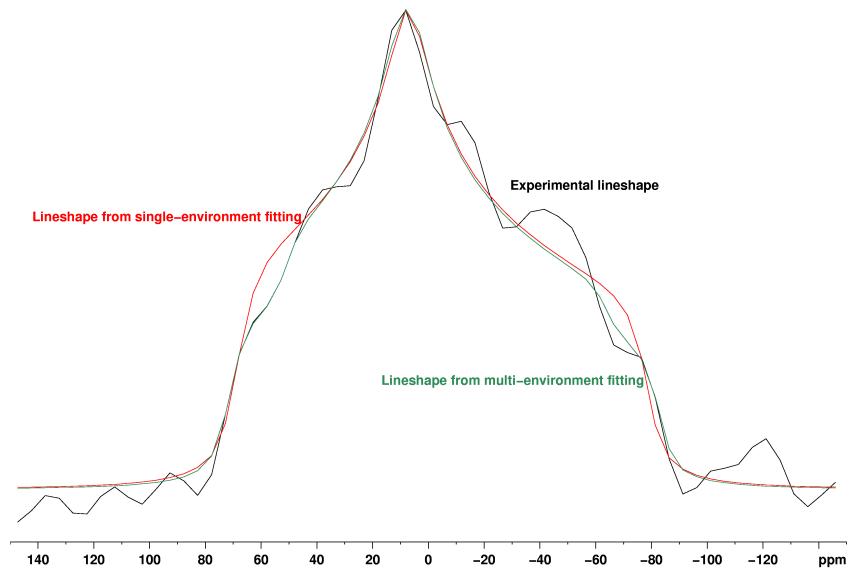


(a) Environment at 173.4 ppm

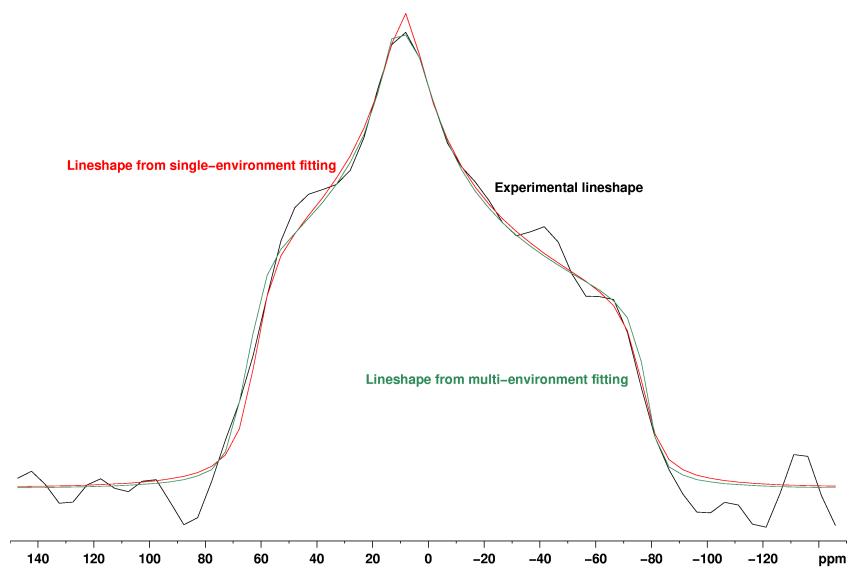


(b) Environment at 172.4 ppm

Figure 16: Multiple lineshape fitting for slices of the ^{13}C 2DCSA spectrum of poly(Pro-Gly-Pro).



(a) Environment at 173.4 ppm



(b) Environment at 172.4 ppm

Figure 17: Comparison of single- and multiple lineshape fitting for slices of the ^{13}C 2DCSA spectrum of poly(Pro-Gly-Pro).

7 ^{13}C 2DCSA Studies of Triple-Helical Peptides

7.1 Collagen Model Peptides

Crystal structures for $(\text{Pro}-\text{Hyp}-\text{Gly})_{10}$ and $(\text{Pro}-\text{Pro}-\text{Gly})_{10}$ are available [23, 25] (PDB: 1V6Q, 1K6F) and confirm the structures as triple-helical. Monodisperse samples of the peptides $(\text{Gly}-\text{Pro}-\text{Hyp})_{11}$ and $(\text{Gly}-\text{Pro}-\text{Pro})_{11}$ are thus assumed triple-helical. ^{13}C 2DCSA spectra of

- $(\text{Pro}-\text{Hyp}-\text{Gly})_{10}$
- $(\text{Gly}-\text{Pro}-\text{Hyp})_{11}$ (labelled at glycine 7)
- $(\text{Gly}-\text{Pro}-\text{Pro})_{11}$ (labelled at the second proline in triplet 7)

were recorded. Assignments of spectra refer to [26–29].

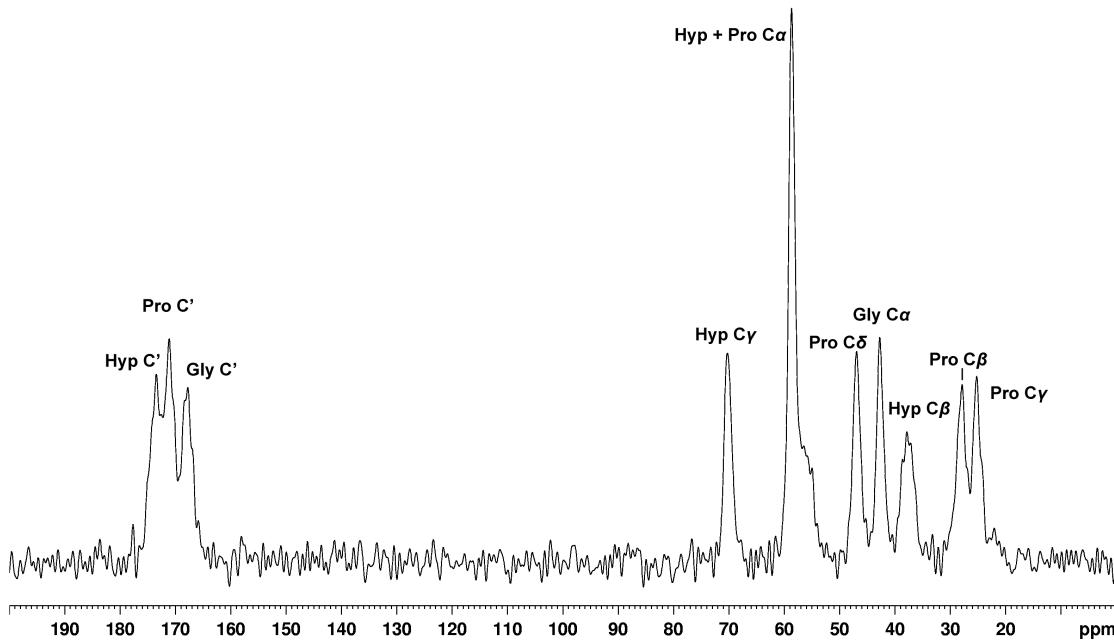


Figure 18: ^{13}C CP spectrum of $(\text{Pro}-\text{Hyp}-\text{Gly})_{10}$. The signal at 173.45 ppm is assigned to Hyp by comparison with the spectrum of $(\text{Gly}-\text{Pro}-\text{Pro})_{11}$. The labelled proline in the $(\text{Gly}-\text{Pro}-\text{Pro})_{11}$ sample is at the same position in the triplet (i.e. preceded by proline, succeeded by glycine) as the hydroxyprolines in $(\text{Pro}-\text{Hyp}-\text{Gly})_{10}$.

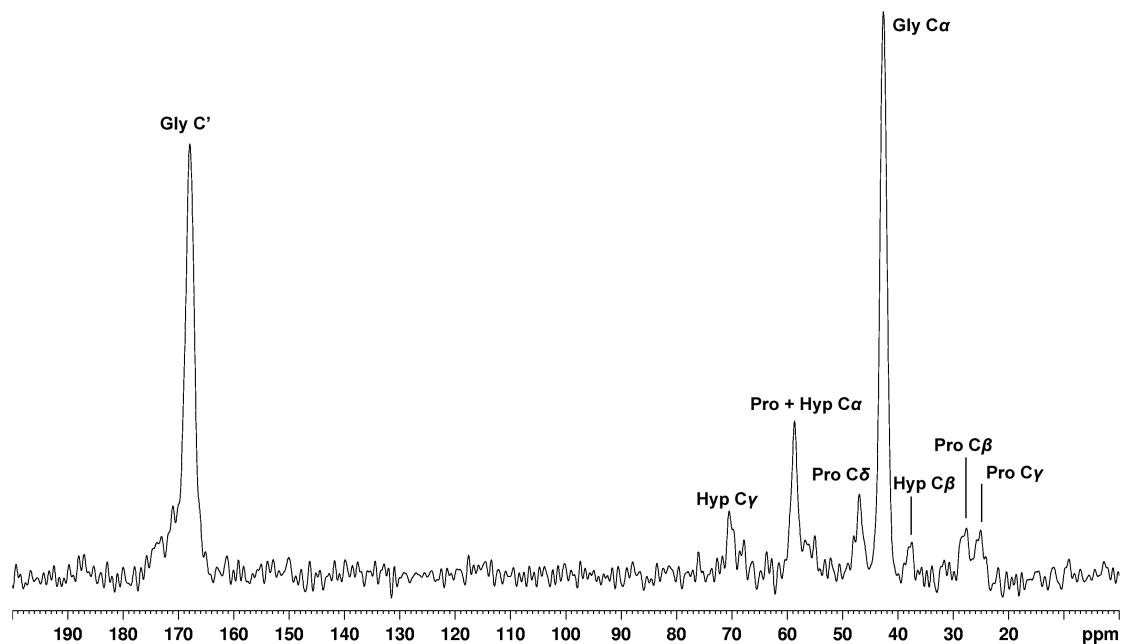


Figure 19: ^{13}C CP spectrum of $(\text{Gly-Pro-Hyp})_{11}$ labelled at glycine 7

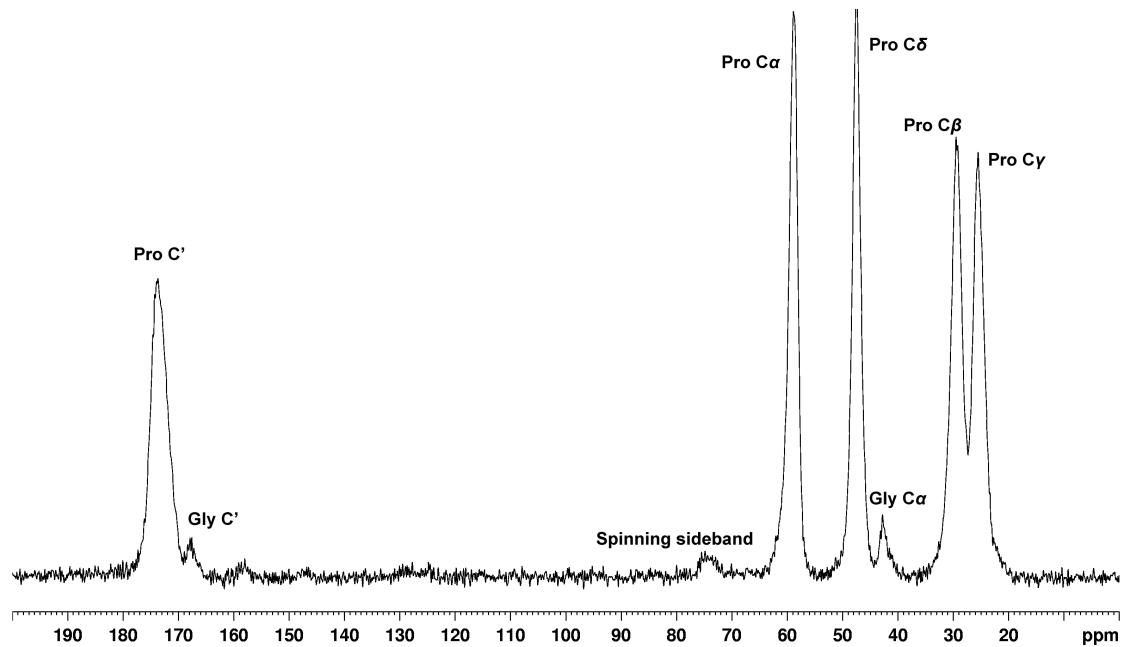


Figure 20: ^{13}C CP spectrum of $(\text{Gly-Pro-Pro})_{11}$ labelled at the second proline residue in triplet 7

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.59	Hyp C'	241	184	96	-78	0.74	1339	93.89
171.48	Pro C'	237	184	93	-79	0.67	806	95.13
168.25	Gly C'	236	175	93	-75	0.81	868	93.27
70.22	Hyp C γ	90	64	56	20	0.39	494	94.16
58.64	Hyp + Pro C α	79	53	44	20	0.48	484	91.44
46.88	Pro C δ	66	49	26	-21	0.77	840	90.88
42.73	Gly C α	63	45	20	-23	0.79	630	90.53
37.81	Hyp C β	63	34	15	26	0.74	689	91.42
27.86	Pro C β	48	31	5	-23	0.71	891	95.41
25.25	Pro C γ	42	28	6	-19	0.77	720	93.61

Table 3: ^{13}C 2DCSA fitting values for $(\text{Pro-Hyp-Gly})_{10}$. For uncertainties, see Error Analysis, section 7.4.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
168.01	Gly C'	237	172	95	-73	0.88	1111	94.89
42.60	Gly C α	63	46	19	-24	0.78	845	93.59

Table 4: ^{13}C 2DCSA fitting values for $(\text{Gly-Pro-Hyp})_{11}$ labelled at glycine 7. For uncertainties, see Error Analysis, section 7.4.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.39	Pro C'	247	177	96	-77	0.90	1377	93.46
58.72	Pro C α	78	62	37	-22	0.72	1190	94.28
47.30	Pro C δ	73	45	24	25	0.82	982	93.13
29.19	Pro C β	50	25	13	21	0.57	1082	94.25
25.52	Pro C γ	45	19	12	20	0.32	980	93.43

Table 5: 2DCSA fitting values for $(\text{Gly-Pro-Pro})_{11}$ labelled at the second proline of triplet 7. For uncertainties, see Error Analysis, section 7.4.

For $(\text{Pro-Hyp-Gly})_{10}$, the signal at 59 ppm contains contributions from proline and hydroxyproline α carbons. In principle, it is possible to fit the lineshape with representative tensors for proline $C\alpha$ and hydroxyproline $C\alpha$. However, whilst a ‘pure’ proline $C\alpha$ tensor has been measured for $(\text{Gly-Pro-Pro})_{11}$, a sample was not available where hydroxyproline was present in the absence of proline.

7.2 Other Triple-Helical Peptides

Samples with no available crystal structure but that are assumed to be triple-helical were also studied:

- Sheep osteoblast collagen
- Poly(Pro-Gly-Pro)
- Aluminium (III) tanned leather.

7.2.1 Sheep Osteoblast Collagen

Osteoblasts secrete collagen as part of the bone formation process [30]. The sheep osteoblast collagen sample described in section 4.2 was prepared with ^{13}C and ^{15}N labelled proline. The signals in the NMR spectrum are from proline and hydroxyproline residues only (labelled proline will be post-translationally modified to give labelled hydroxyproline).

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.72	Pro C'	245	188	88	-85	0.67	667	91.18
70.33	Hyp C γ	88	72	50	-20	0.80	368	94.79
58.87	Hyp + Pro C α	80	62	35	-24	0.75	710	94.64
47.51	Pro C δ	72	47	23	24	1.00	936	95.11
37.04	Hyp C β	53	36	22	16	0.84	1014	91.93
30.02	Pro C β	44	29	17	14	0.88	707	93.20
24.98	Pro C γ	43	22	10	18	0.70	692	95.56

Table 6: ^{13}C 2DCSA fitting values for sheep osteoblast collagen. For uncertainties, see Error Analysis, section 7.4.

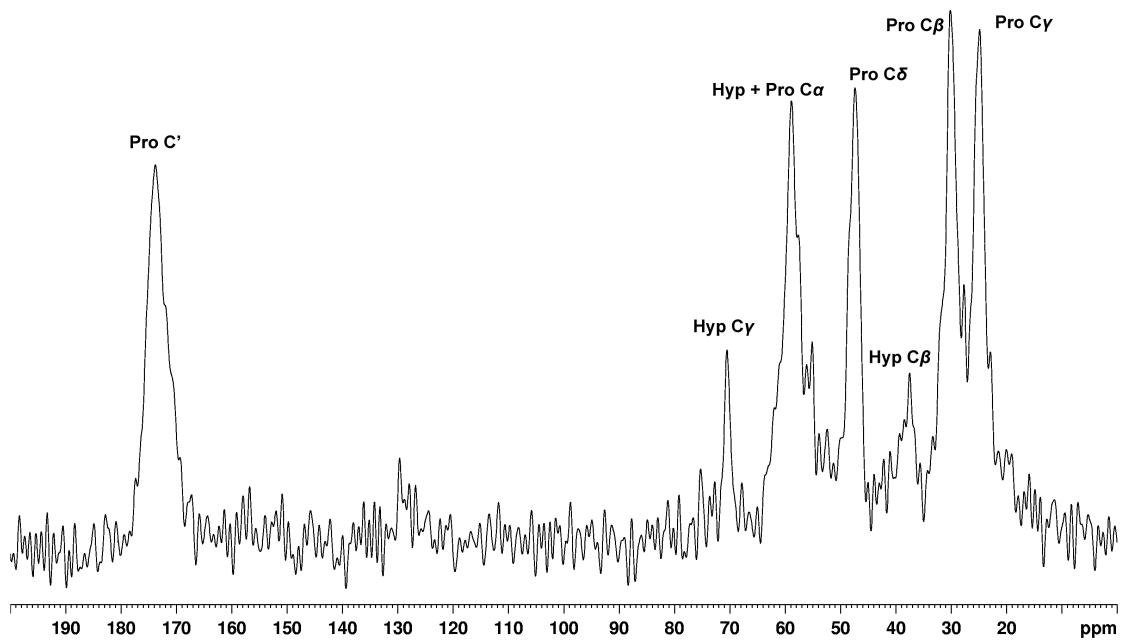


Figure 21: ^{13}C CP spectrum of sheep osteoblast collagen cultured with ^{13}C and ^{15}N labelled proline

7.2.2 Poly(Pro-Gly-Pro)

Poly(Pro-Gly-Pro) was studied in section 6, the ^{13}C 2DCSA results are repeated in Table 7.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.43	Pro C'	244	181	95	-79	0.80	883	93.67
172.39	Pro C'	238	183	96	-76	0.72	930	94.82
171.17	Pro C'	237	180	96	-75	0.77	854	97.12
167.98	Gly C'	238	172	95	-73	0.90	1106	94.70
58.75	Pro C α	81	60	35	-24	0.86	1291	93.24
47.35	Pro C δ	72	48	22	-25	0.98	1174	94.27
41.94	Gly C α	72	40	14	30	0.88	729	93.18
28.62	Pro C β	48	26	12	19	0.73	1291	93.32
24.71	Pro C γ	45	21	7	21	0.68	1169	93.61

Table 7: ^{13}C 2DCSA fitting values for poly(Pro-Gly-Pro). For uncertainties, see Error Analysis, section 7.4.

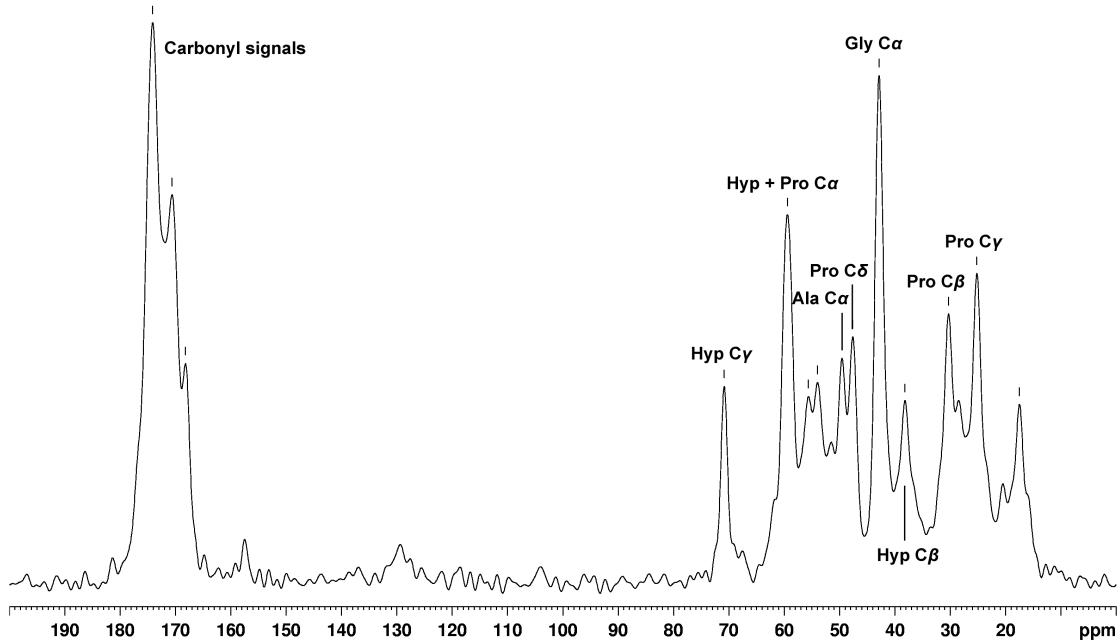


Figure 22: ^{13}C CP spectrum of Aluminium-tanned leather recorded by Dr David G. Reid.
The spinning rate was 14 kHz.

7.2.3 Aluminium Tanned Leather

It was noticed [29] that leather tanned with Aluminium (III) produces sharper ^{13}C signals than untanned leather, without changing the isotropic shifts, indicating a higher degree of structural homogeneity. Leather is made from skin, which is primarily collagen. ^{13}C 2DCSA experiments were recorded to ascertain whether tanning changes the CSA tensors of the residues. The ^{13}C CP spectrum of the leather sample is shown in Figure 22; the ^{13}C CSA tensors are shown in Table 8.

7.3 Ala-Pro-Gly

Ala-Pro-Gly is *not* triple-helical (a crystal structure is available [31]), it was studied to provide reference proline CSA components for a random coil-type structure⁹.

⁹The compound was sold as Ala-Pro-Gly, but the ^{13}C spectrum did not show a signal for a carboxylate. It is assumed that the sample is actually Ala-Pro-Gly-NH₂.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.46	Pro C'	238	187	95	-79	0.65	2026	95.89
171.65	Pro C'	245	184	86	-86	0.72	1642	93.45
168.44	Gly C'	243	182	80	-88	0.70	1520	91.82
58.75	Hyp + Pro C α	88	59	31	-29	0.98	508	91.22
47.35	Pro C δ	76	51	16	-31	0.80	0	84.69
41.94	Gly C α	76	38	13	34	0.75	681	91.98
28.62	Pro C β	48	41	2	-28	0.24	1055	89.96
24.71	Pro C γ	53	30	-7	-32	0.69	350	87.40

Table 8: ^{13}C 2DCSA fitting values for Aluminium-tanned leather. For uncertainties, see Error Analysis, section 7.4.

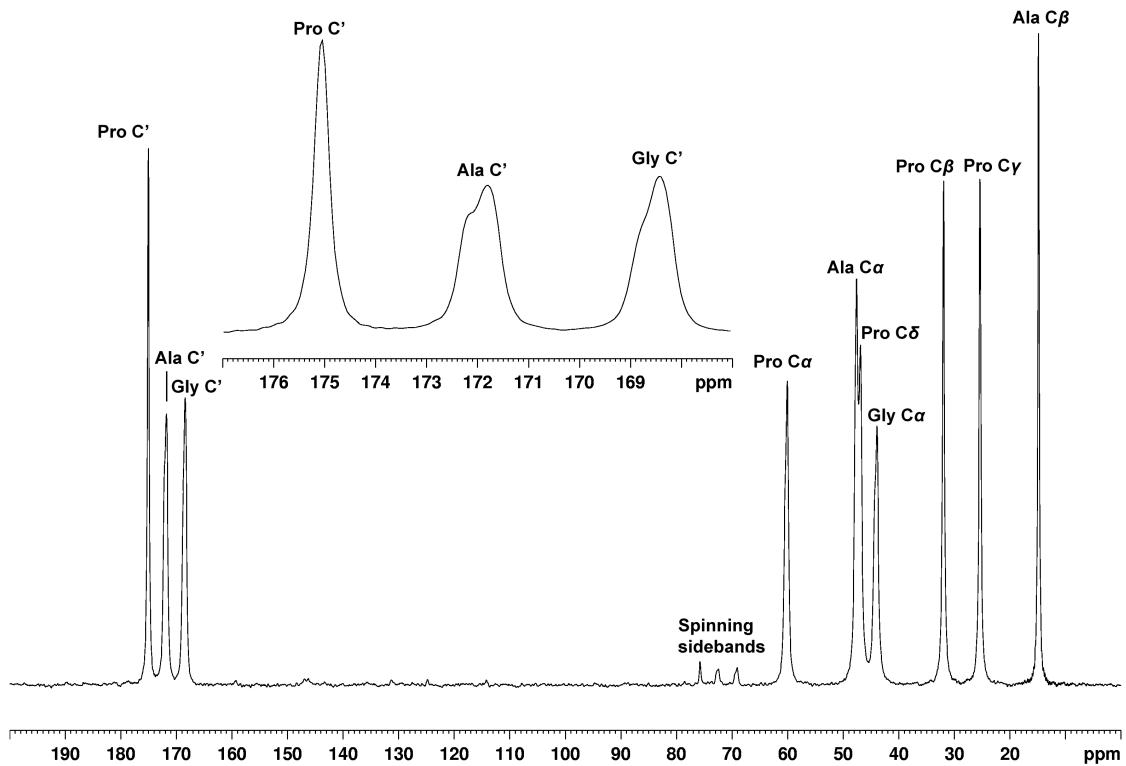


Figure 23: ^{13}C CP spectrum of Ala-Pro-Gly-NH₂, with expansion of the carbonyl region. The glycine and alanine signals show splittings due to ^{14}N quadrupolar coupling. The splitting is not seen for the proline signal, which appears to have a smaller quadrupolar coupling constant.

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
174.81	Pro C'	239	180	106	-69	0.86	830	98.05
172.00	Ala C'	242	181	93	-79	0.77	1100	97.52
168.60	Gly C'	238	175	93	-76	0.82	840	97.56
60.08	Pro C α	80	63	37	-23	0.74	871	94.75
47.47	Ala C α	65	50	28	-19	0.78	820	94.75
46.76	Pro C δ	70	45	26	23	0.82	1390	95.29
43.97	Gly C α	65	47	20	-24	0.75	821	94.50
31.98	Pro C β	54	27	14	22	0.58	718	96.85
25.39	Pro C γ	46	18	12	20	0.30	685	96.68
14.79	Ala C β	3	-3	-21	-14	0.42	706	96.38

Table 9: ^{13}C 2DCSA fitting values for Ala-Pro-Gly. For uncertainties, see Error Analysis, section 7.4.

7.4 Error Analysis

Error analysis as described in section 4.4 was preformed on a representative sample of carbonyl and aliphatic data: aluminium tanned leather, sheep osteoblast and polyproline. The results are shown in Table 10.

Environment	Average	Errors			
	δ_{iso} / ppm	δ_{iso} / ppm	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm
Aliphatic	3.1	1.0	1.3	1.1	1.5
Amide	-4.2	2.6	3.9	2.8	2.3

Table 10: Error estimates for the 2DCSA fitting procedure, and average measured δ_{iso} values from the fitting process.

The average measured values for δ_{iso} (which should be zero — section 4.4) are larger than the error in δ_{iso} indicating the presence of pulse imperfections in the experiment, which can lead to lineshape distortions. However, the lineshapes produced do not seem to be distorted, so these effects are not considered further. The differing values of the errors in the principal components are due to the different contributions that each principal component has to Δ and η (recall that Δ and η are varied in the error analysis rather than the principal components themselves).

7.5 Chemical Shift Prediction Programs

Programs designed to predict chemical shifts from protein crystal structures have been reported [32–37]. One such program, SHIFTX2, was used to predict isotropic shifts for (Pro-Gly-Gly)₁₀. Predicted isotropic shifts in the carbonyl region were much higher than experimentally measured. The programs are ‘trained’ on neural network databases that do not contain collagen. This suggests that the structural effects required to produce the specific low shifts seen experimentally are rare outside the triple-helical environment. Crystal structures are not available for polyproline or poly(Pro-Gly-Pro), but it is expected that the software would overestimate these shifts for the same reason.

Residue	Average Predicted δ_{iso} / ppm
Gly- Pro -Pro	175.4
Gly- Pro - Pro	175.6
Gly	173.1

Table 11: Results of a SHIFTX2 prediction for ^{13}C carbonyl shifts in (Pro-Gly-Gly)₁₀. The predicted isotropic shifts are higher than those measured by solid-state NMR. The central six triplets were chosen to eliminate end-effects.

7.6 Discussion

The most striking aspects of the data are in the carbonyl region. The isotropic shifts for glycine and proline are lower than expected for a ‘random coil’ (174 ppm for Proline, 172 ppm for Glycine ¹⁰ [20]), and there are two distinct proline shifts, at 172 ppm and 173.5 ppm. The low isotropic shifts seem to be a factor of the helix, because they are not well predicted by software.

7.6.1 Glycine Carbonyl Chemical Shifts

It seems that the low isotropic shifts seen for glycine are evidence for a helical structure, however the poly(Pro-Gly-Pro) spectrum does not show separate sig-

¹⁰Shifts for the residue X measured on pentapeptide Gly-Gly-X-L-Ala-OH in aqueous solution

nals for triple- and single-helical environments. The glycine shift is therefore unsuitable as a triple-helix fingerprint. It is possible to distinguish between triple- and single-helical proline shifts in poly(Pro-Gly-Pro), so proline shift is the best candidate as a triple-helical fingerprint.

7.6.2 Proline Carbonyl Chemical Shifts

Two isotropic shifts are seen for proline in the triple-helical environment, around 172 ppm and 173.5 ppm. The $(\text{Gly-Pro-Pro})_{11}$ sample — which is labelled as $(\text{Gly-Pro-Pro})_6-(\text{Gly-Pro-Pro}^*)-(\text{Gly-Pro-Pro})_4$ — shows only the higher of the two signals. The assumption is therefore made that the proline *before* glycine in the peptide chain produces this signal in all of the samples studied and the proline *after* glycine produces the lower of the signals.

Both of these shifts are lower than those in a random-coil (on average 174 ppm [20]), and are not predicted by software. They are rationalised using the crystal structure of $(\text{Pro-Hyp-Gly})_{10}$. It is assumed that the γ -hydroxy group of hydroxyproline does not affect the carbonyl group; the argument is equally valid for $(\text{Pro-Pro-Gly})_{10}$. The carbonyl groups of glycine and the proline residue immediately after it in the chain are eclipsed by the pyrrolidine ring of the adjacent proline or hydroxyproline residue (Figure 24). We propose that this accounts for the low chemical shifts of those environments.

We can rationalise the differences in principal components between the two environments (Table 12). The δ_{22} component is orientated along the C=O bond. The environment at 173.5 ppm has carbonyl groups that point tangentially around the helix. They are likely to be involved in strong hydrogen-bonding with water of crystallisation, and hence have a lower δ_{22} component, as described in section 5.3. The δ_{11} component is orientated perpendicular to the plane of the peptide bond, i.e. it points into the π -bond. The difference in δ_{11} components is likely to be due eclipsing (or lack thereof) of the carbonyl groups with the pyrrolidine ring of the adjacent proline residue (as described for the differences in isotropic shifts for the two environments). The δ_{33} components are similar for both environments — δ_{33} depends largely on the σ -bonding framework and is not greatly affected by

hydrogen-bonding. The ^{13}C 2DCSA experiments hence show that the differences in ^{13}C isotropic shift in the proline carbonyl groups arises from the orientation of the carbonyl groups in relation to the peptide chain, and the extent of their hydrogen-bonding.

$\delta_{iso} / \text{ppm}$	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm
173.5	244.0	181.0	95.5
172.0	237.5	184.0	94.5

Table 12: The average ^{13}C principal components of the two proline carbonyl environments.

Figures 25 and 26 show that the leather has appreciably different principal components to the ‘untreated’ samples. This is unsurprising — the aluminium is likely to interact with the carbonyl groups and hence affect the δ_{11} (pointing into the π -system) and δ_{22} (along the C=O bond) components . Indeed it is remarkable that even the isotropic shift is preserved.

7.6.3 Aliphatic Tensors

No trends were observed for the ^{13}C principal components of aliphatic signals, and the non-helical Ala-Pro-Gly was within the observed range for triple-helical samples. Aliphatic components therefore do not seem to be useful in assigning triple-helicity.

7.7 Hypothesis Testing

The data suggest that the presence of proline shifts at around 172 ppm and 173.5 ppm are indicative of a triple-helix. In an unlabelled collagen sample, other residues contribute to the carbonyl region; sharp signals are not seen. The 2DCSA experiment overcomes this lack of resolution — slices taken at approximately 172 ppm and 173.5 ppm and the tensors can be compared to those measured for triple-helical peptides. To test this method on an unlabelled collagen sample a ^{13}C 2DCSA spectrum of bovine achilles tendon collagen was recorded.

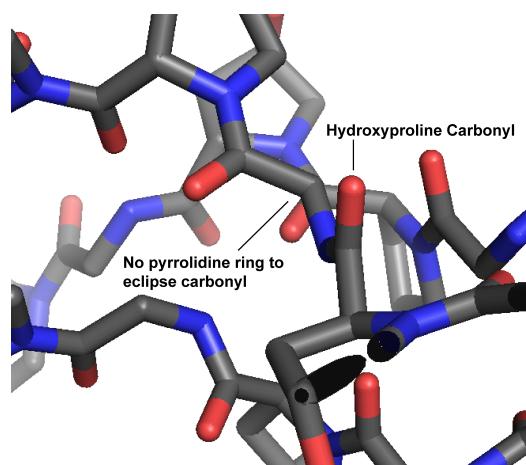
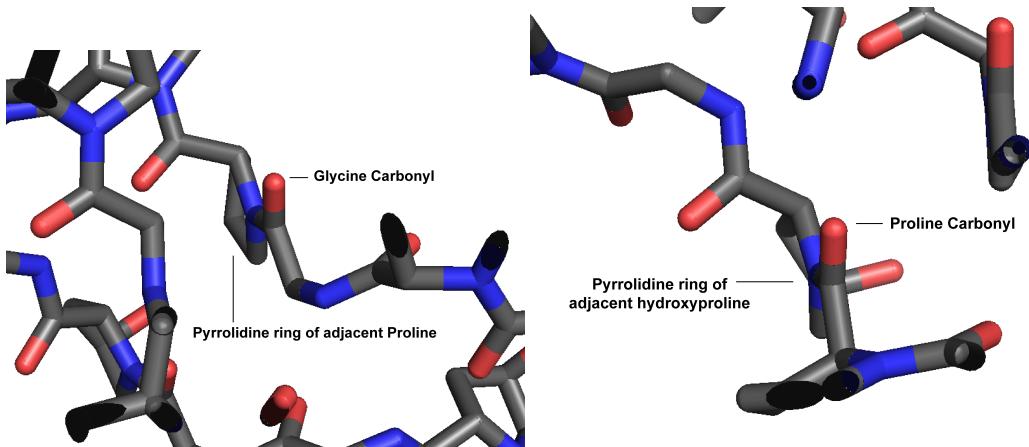


Figure 24: Carbonyl groups of Pro, Hyp and Gly residues in $(\text{Pro-Hyp-Gly})_{10}$, showing the carbonyl groups of Pro and Gly eclipsed with the pyrrolidine rings of adjacent Hyp and Pro respectively. Images produced using the PyMOL program, www.pymol.org

$\delta_{iso} / \text{ppm}$	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.62	Pro C'	241	185	95	-79	0.71	818	93.65
172.84	Pro C'	243	182	93	-79	0.76	704	93.93

Table 13: ^{13}C 2DCSA fitting values for bovine achilles tendon collagen. For uncertainties, see Error Analysis, section 7.4.

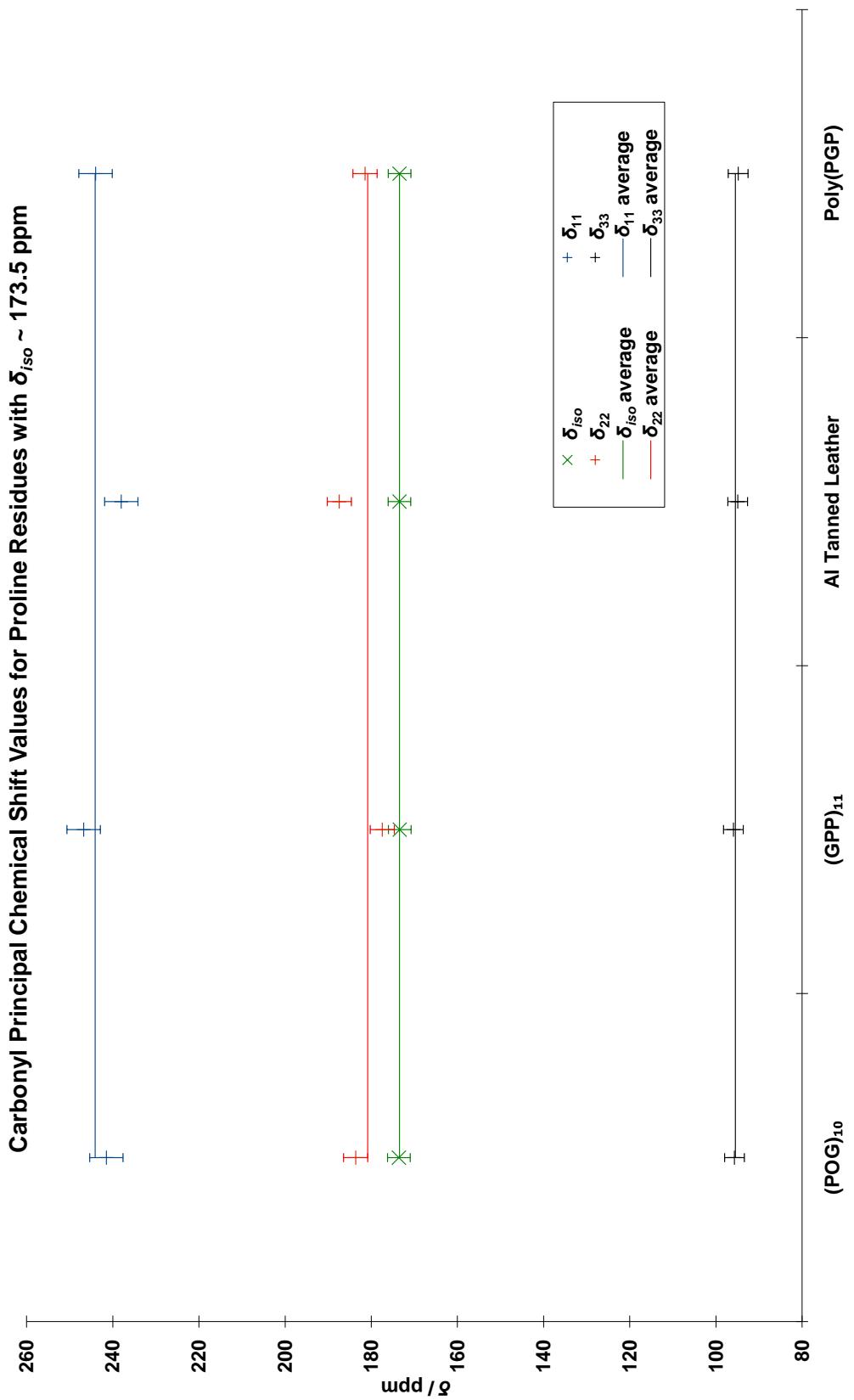


Figure 25: Plot of principal components for proline carbonyls with $\delta_{iso} 173.5$ ppm. The leather sample has different values of δ_{11} and δ_{22} to the untreated samples and is not included in the ‘average’ lines.

Carbonyl Principal Chemical Shift Values for Proline Residues with $\delta_{iso} \sim 172$ ppm

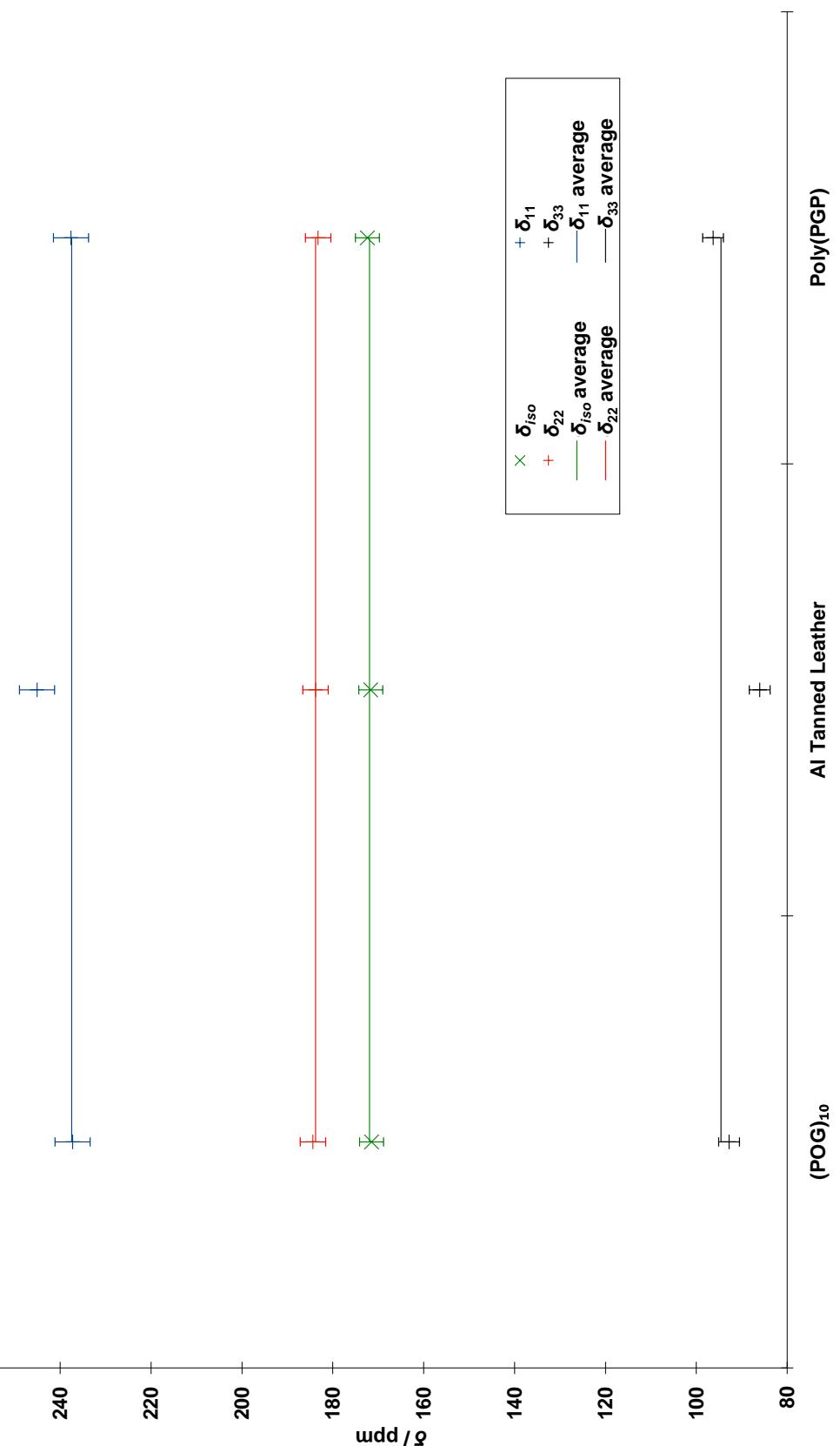


Figure 26: Plot of principal components for proline carbonyls with δ_{iso} 172 ppm. The leather sample has different values of δ_{11} and δ_{22} to the untreated samples and is not included in the ‘average’ lines.

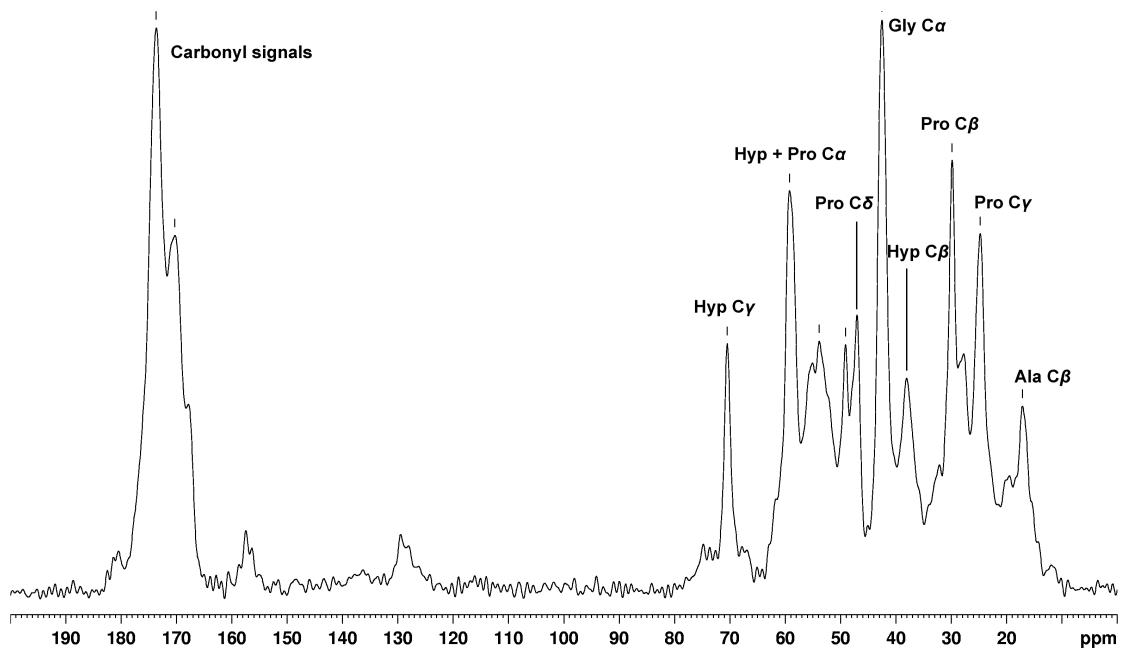


Figure 27: ^{13}C CP spectrum of bovine achilles tendon collagen. Note the broad peak in the carbonyl region; it is impossible to assign triple-helicity on the basis of a 1D spectrum.

Slices at 173.6 ppm and 172.8 ppm fit the measured proline carbonyl principal components for the other triple-helical peptides within error; the test appears robust (Figures 28 and 29). Fitting of the lineshapes with multiple components (as in section 6.3.1) was performed for collagen. No evidence of a single-helical component was seen — a native collagen sample is not expected to have single-helical regions.

7.8 ^{13}C 2DCSA Measurements of Gelatine

Gelatine is denatured collagen. It is proposed that gelatine retains the triple-helical structure on the basis of its proline carbonyl ^{13}C principal components (Table 14, Figures 28 and 29).

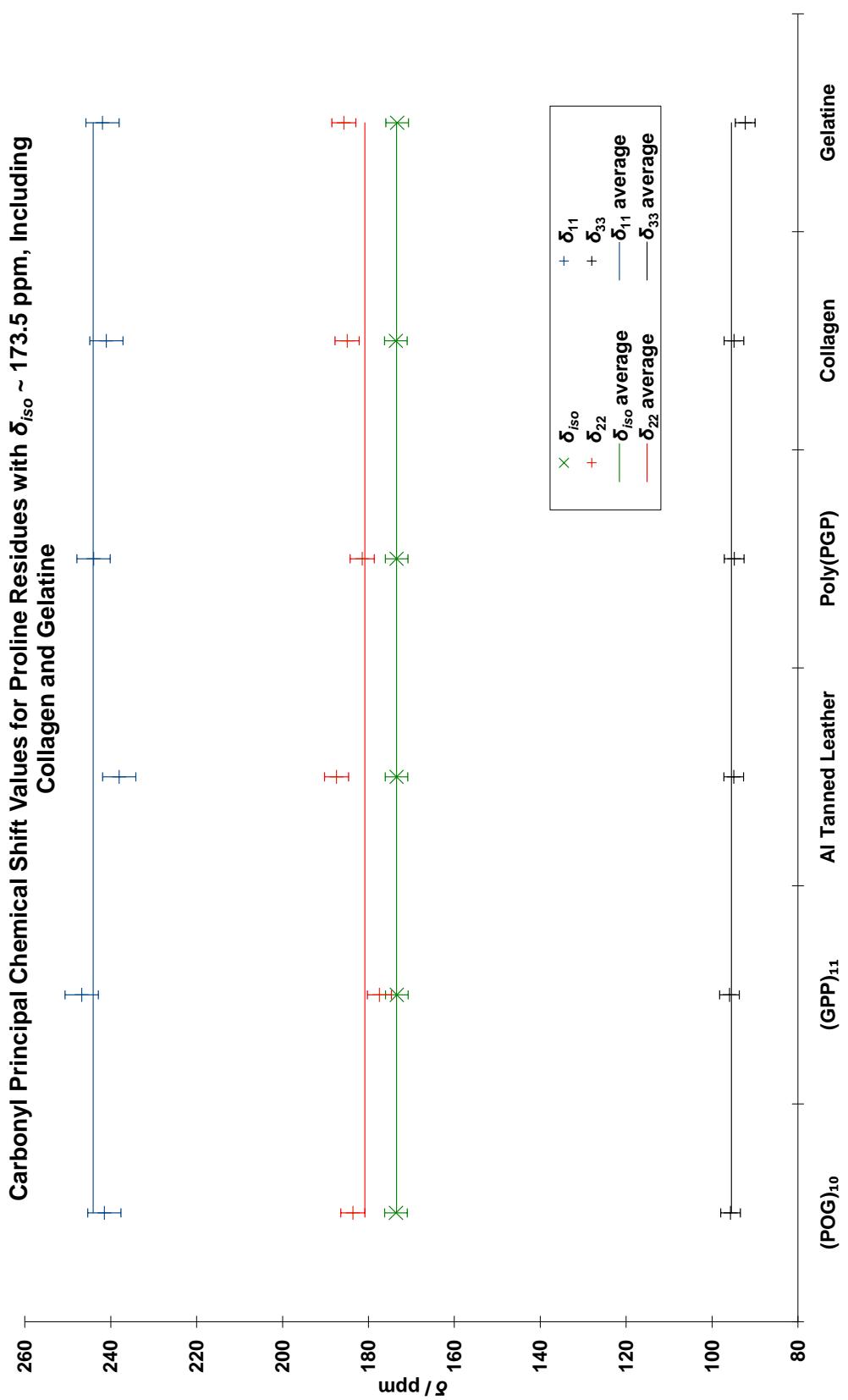


Figure 28: Plot of principal components for proline carbonyls with δ_{iso} 173.5 ppm, including collagen and gelatine. The leather sample has different values of δ_{11} and δ_{22} to the untreated samples and is not included in the ‘average’ lines.

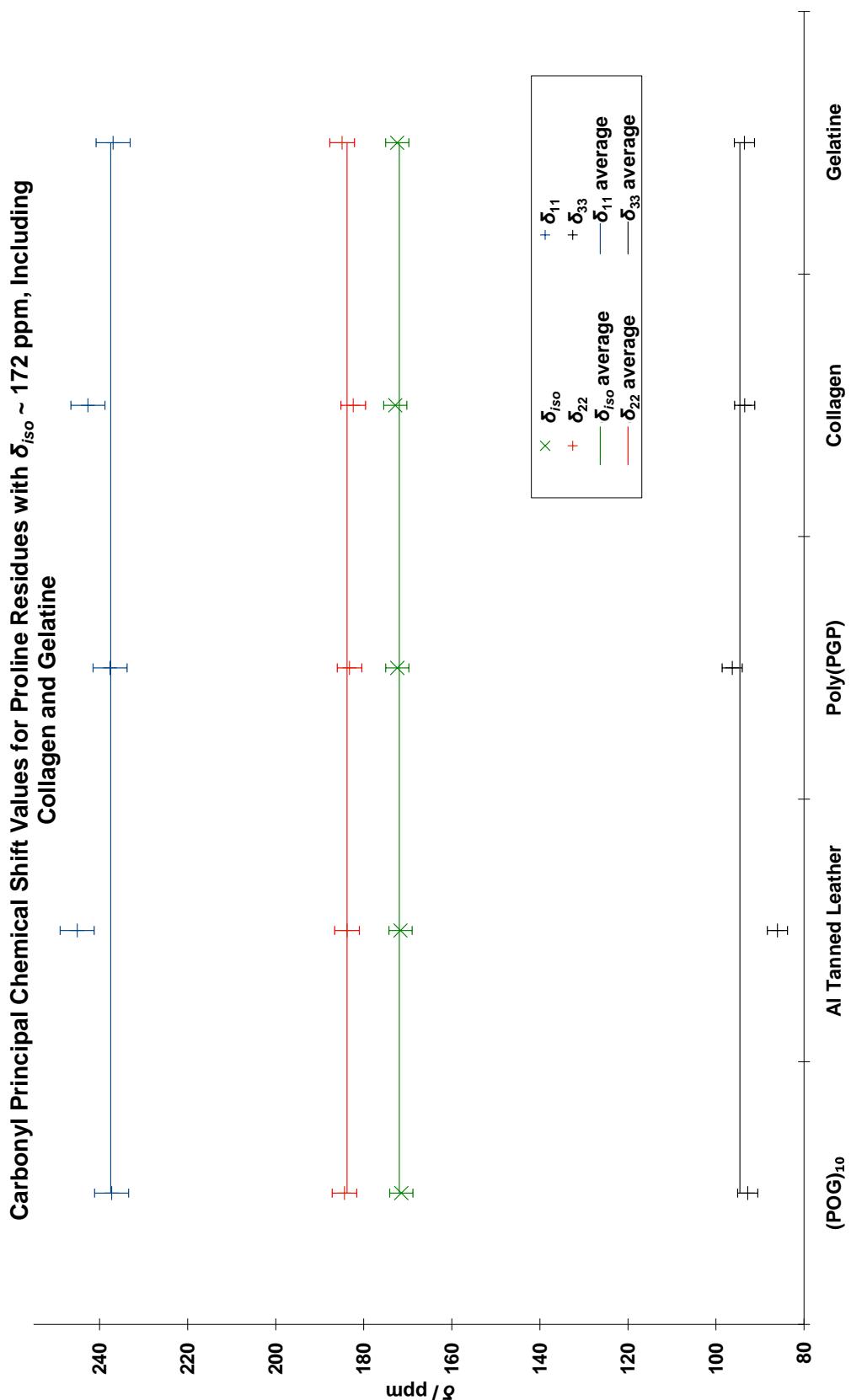


Figure 29: Plot of principal components for proline carbonyls with $\delta_{iso} 172$ ppm, including collagen and gelatine. The leather sample has different values of δ_{11} and δ_{22} to the untreated samples and is not included in the ‘average’ lines.

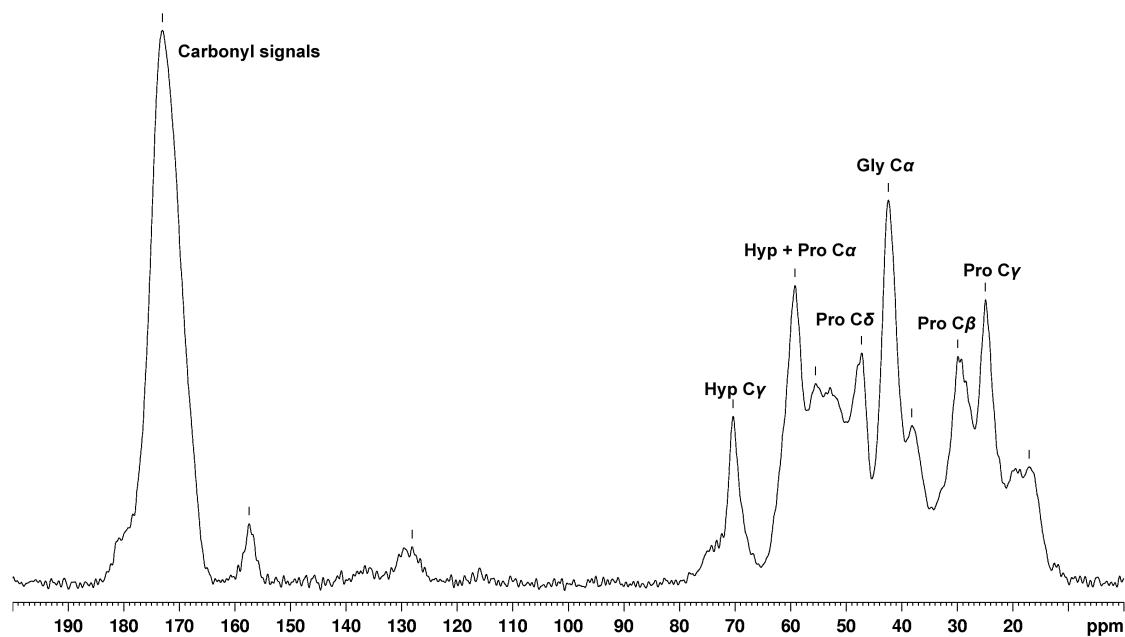


Figure 30: ^{13}C CP spectrum of gelatine

δ_{iso} / ppm	Assignment	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm	Δ / ppm	η	Broadening / Hz	Overlap / %
173.3	Pro C'	242	186	92	-81	0.69	1296	96.37
172.4	Pro C'	237	185	95	-77	0.68	1439	95.32

Table 14: ^{13}C 2DCSA fitting values for gelatine

8 Conclusions and Further Work

This work demonstrates that it is possible to use solid-state NMR to ascertain whether a collagen-like peptide is triple-helical by fitting slices of a ^{13}C 2DCSA experiment taken at approximately 172 ppm and 173.5 ppm. The triple-helix is the minimum requirement for any synthetically-produced collagen to possess. Further work should measure CSA tensors of more triple-helical peptides to improve the reliability of the test. Recording the 2DCSA experiments at higher field or for longer would increase signal-to-noise ratio and increase reliability. 2DCSA experiments could be conducted using ^{15}N NMR; CSA tensors of ^{15}N nuclei potentially contain a large amount of structural information. Cross-referencing of the results against another tensor-measuring experiment should also be performed; a sideband-based experiment would be a suitable complement to 2DCSA [38].

8.1 Other Model Compounds

Z-Gly-Pro-Leu-Gly-Pro (where Z is a protecting group) is available and its crystal structure [39] shows a beta-turn. The conformation of proline in this compound is similar to that in a triple-helix. It would be interesting to study the CSA tensors and compare to triple-helical peptides. The Z-group may be important in the beta-turn; removal of the group and repetition of the 2DCSA experiment could give an interesting comparison.

Ramipril is a drug that contains a proline-like fragment. The ring is bicyclic in Ramipril, and will be less liable to conformational change than the ring in proline. The effects of this on the ^{15}N CSA tensor could be investigated, with comparison to Ala-Pro-Gly. Ultimately this could lead to an improved understanding of the pyrrolidine rings in collagen.

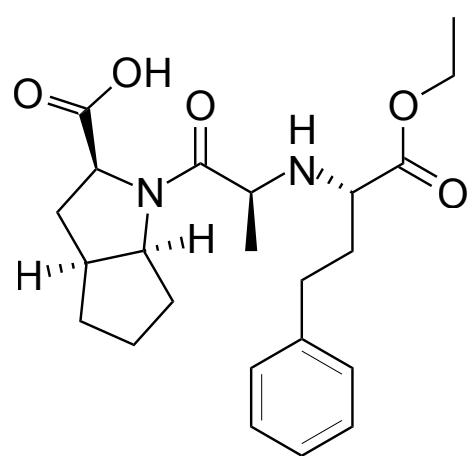


Figure 31: Ramipril

References

- [1] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter, *Molecular Biology of the Cell*, chapter 19, pages 1096–1102, fourth edition, Garland Science, 2002.
- [2] Detlef Reichert, Ovidiu Pascui, Eduardo R. deAzevedo, Tito J. Bonagamba, Klaus Arnold and Daniel Huster, *Magnetic Resonance in Chemistry*, 2004, **42**, 276–284.
- [3] Alberto Lessari, Emilio J. Cocinero, Juan C. López and José L. Alonso, *J. Am. Chem. Soc.*, 2005, **127**, 2572–2579.
- [4] W. Traub and A. Yonath, *Journal of Molecular Biology*, 1966, **16**, 404–414.
- [5] J Engel, J. Kurtz, E. Katchalski and A. Berger, *Journal of Molecular Biology*, 1966, **17**, 255–272.
- [6] Melinda J. Duer, *Introduction to Solid-State NMR Spectroscopy*, Blackwell, 2004.
- [7] Robin M. Orr, *Measuring Anisotropic Interactions Using Solid-State NMR Spectroscopy*, Ph.D. thesis, University of Cambridge, 2006.
- [8] Robin K. Harris, Edwin D. Becker, Sonia M. Cabral de Menezes, Pierre Granger, Roy E. Hoffman and Kurt W. Zilm, *Pure and Applied Chemistry*, 2008, **80** (1), 59–84.
- [9] Ad Bax, Nikolaus M. Szeverenyi and Gary E. Maciel, *Journal of Magnetic Resonance*, 1983, **55**, 494–497.
- [10] T. Terao, T. Fujii, T. Onodera and A Saika, *Chemical Physics Letters*, 1984, **107** (2), 145–148.
- [11] Robin M. Orr and Melinda J. Duer, *Journal Of Magnetic Resonance*, 2006, **181**, 1–8.

- [12] S-F. Liu, J-D. Mao and K. Schmidt-Rohr, *Journal of Magnetic Resonance*, 2002, **155**, 15–28.
- [13] Robert Tycko, Gary Dabbagh and Peter A. Mirau, *Journal Of Magnetic Resonance*, 1989, **85**, 265–274.
- [14] Geoffrey Bodenhausen, Herbert Kogler and R. R. Ernst, *Journal of Magnetic Resonance*, 1984, **58** (3), 370–388.
- [15] Alex D. Bain, *Journal of Magnetic Resonance*, 1984, **56** (3), 418–427.
- [16] Malcolm H. Levitt, P. K. Madhu and Colan E. Hughes, *Journal of Magnetic Resonance*, 2002, **155**, 300–306.
- [17] Benjamin J. Stapley and Trevor P. Creamer, *Protein Science*, 1999, **8**, 587–595.
- [18] Trevor P. Creamer, *Proteins: Structure, Function, and Bioinformatics*, 1998, **33** (2), 218–226.
- [19] Jorge A. Vila, Héctor A. Baldoni, Daniel R. Ripoll, Avijit Ghosh and Harold A. Scheraga, *Biophysical Journal*, 2004, **86** (2), 731–742.
- [20] ‘Biological Magnetic Resonance Bank, www.bmrb.wisc.edu’, .
- [21] Christiane Garbay-Jaureguiberry, Bernadette Arnoux, Thierry Prangé, Suzanne Wehri-Altenburger, Claudine Pascard and Bernard P. Roques, *J. Am. Chem. Soc*, 1980, **102** (6), 1827–1837.
- [22] Karine Loth, Philippe Pelupessy and Geoffrey Bodenhausen, *J. Am. Chem. Soc.*, 2005, **127**, 6062–6068.
- [23] R. Berisio, L. Vitagliano, L. Mazzarella and A. Zagari, *Protein Science*, 2002, **11**, 262–276.
- [24] David A. Slatter, Christopher A. Miles and Allen J. Bailey, *Journal of Molecular Biology*, 2003, **329**, 175–183.

- [25] K. Okuyama, C. Hongo, R. Fukushima, G. Wu, H. Narita, K. Noguchi, Y. Tanaka and N. Nishino, *Biopolymers*, 2004, **76**, 367–377.
- [26] Erica R. Wise, *Solid state NMR studies of mineralised skeletal tissues*, Ph.D. thesis, University of Cambridge, 2008.
- [27] Abil E. Aliev, *Biopolymers*, 2005, **77** (4), 230–245.
- [28] Akira Shoji, Takuo Ozaki, Teruaki Fujito, Kenzo Deguchi, Isao Ando and Jun Magoshi, *Journal of Molecular Structure*, 1998, **441**, 251–266.
- [29] Frederik H. Romer, Andrew P. Underwood, Nadine D. Senekal, Susan L. Bonnet, Melinda J. Duer, David G. Reid and Jan H. van der Westhuizen, *Molecules*, 2011, **16**, 1240–1252.
- [30] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter, *Molecular Biology of the Cell*, chapter 22, pages 1305–1308, fourth edition, Garland Science, 2002.
- [31] S. Wu, J.-P. Declercq and M. van Meerssche, *Bulletin des Sociétés Chimiques Belges*, 1987, **96** (7), 515–520.
- [32] Yang Shen and Ad Bax, *Journal of Biomolecular NMR*, 2007, **38** (4), 289–302.
- [33] Yang Shen and Ad Bax, *Journal of Biomolecular NMR*, 2010, **48** (1), 13–22.
- [34] S. Neal, A. M. Nip, H. Zhang and D. S. Wishart, *Journal of Biomolecular NMR*, 2003, **26** (3), 215–240.
- [35] Beomsoo Han, Yifeng Liu, Simon W. Ginzinger and David S. Wishart, ‘SHIFTX2: Significantly Improved Protein Chemical Shift Prediction’, , 2010, Unpublished Work.
- [36] Kai J. Kohlhoff, Paul Robustelli, Andrea Cavalli, Xavier Salvatella and Michele Vendruscolo, *J. Am. Chem. Soc.*, 2009, **131** (39), 13894–13895.

- [37] Karsten Seidel, Manuel Etzkorn, Robert Schneider, Christian Ader and Marc Baldus, *Solid State NMR*, 2009, **35**, 235–242.
- [38] Robin M. Orr and Melinda J. Duer, *Solid State Nuclear Magnetic Resonance*, 2006, **30** (1), 1–8.
- [39] S. Bando, N. Tanaka, T. Ashida and M. Kakudo, *Acta Crystallographica, Section B*, 1978, **34**, 3447–3449.