Elucidating Structure and Dynamics of Extracellular Matrix Collagen Using Solid-State NMR

Ieva Goldberga

Darwin College

September 2019

A thesis submitted to the University of Cambridge for the degree of

Doctor of Philosophy

Declaration

I hereby declare that this thesis entitled *Elucidating Structure and Dynamics of Extracellular Matrix Collagen Using Solid-State NMR*, is the result of work that I have undertaken in the University Chemical Laboratory at the University of Cambridge between October 2015 and September 2019. Credit for results obtained through collaboration with other parties is referenced directly in the text and the Formal Acknowledgements and Project Contributions section. Except where stated to the contrary, this thesis is my own work. Figures that have been reproduced or adapted from other sources are indicated as such in their corresponding caption references. Aspects of the results presented herein have been published in peer-reviewed journals; original publications are referenced in the main text and a list of publications is included. This thesis has not been, nor is currently being, submitted for any other degree, diploma or other academic qualification at this or any other university. It does not exceed 60,000 words in length.

Ieva Goldberga
September 2019

Abstract

In recent years, solid-state Nuclear Magnetic Resonance (NMR) has emerged as an established spectroscopic method to afford detailed structural information on native cellular and extracellular components at atomic-scale resolution. Fibrillar collagens are the most common component of the extracellular matrix (ECM), comprising up to 20% by weight of the human body and is found in most of the tissues. Due to their diverse structures and compositions, collagens serve many functions, providing structural and mechanical support for surrounding cells, and playing important roles in cell-to-cell communication. Nonetheless, despite being at first glance a simple protein formed by three homologous polypeptide chains of repeating three-amino-acid triads trimerised into a triple helix, it is a highly versatile and complex system. Due to the complexity and size of the triple helix, the scientific community still lacks understanding of collagen structure, flexibility and dynamics at the atomic level, in spite of today's advances in technology. The combination of ^{13}C, ^{15}N-labelled amino acid enrichment of in-vitro or in-vivo materials with two-dimensional solid-state NMR spectroscopy potentially provides a more detailed understanding of the complex collagen structure and dynamics at atomic resolution. Furthermore, our knowledge of undesirable structural changes within the extracellular matrix, such as non-enzymatic glycation reactions with reducing sugars, is limited. Glycation-modified extracellular matrix (ECM) leads to abnormal cell behaviour and widespread cell necrosis, potentially causing numerous health complications, e.g. in diabetic patients. Solid-state NMR is a powerful probe to study these structural changes.

The work presented in this thesis demonstrates how solid-state NMR can be used to study the effects of genetic and glycation chemistry on the molecular structure and dynamics of the collagen. We employed a selection of synthetic model peptides that contain a variation of the native sequence representing normal and defected collagen triple-helical compositions to assess the backbone motions via the use of the ^{15}N T_1 relaxation. Further, we use U-^{13}C,^{15}N-isotopically enriched collagen ECM samples to investigate the conformational and dynamic changes after glycation of the hydrophilic and hydrophobic regions of the collagen fibrils. Finally, we propose a methodology that can be employed to probe different sites (gap and overlap zones) of the collagen fibrils in their native state which can be exploited to detect less abundant species found in the collagen protein.

Publications and Copyright Information

Parts of this thesis have been published in the following peer-reviewed journals:

1. Ieva Goldberga, Rui Li, Melinda J. Duer, *Acc. Chem. Res.*, **2018**, *51*, 1621-1629

2. Wing Ying Chow, Rui Li, Ieva Goldberga, David G. Reid, Rakesh Rajan, Jonathan Clark, Hartmut Oschkinat, Melinda J. Duer, Robert Hayward, Catherine M. Shanahan, *Chem. Commun.*, **2018**, *54*, 12570-12573

3. Ieva Goldberga, Rui Li, Wing Ying Chow, David G. Reid, Ulyana Bashtanova, Rakesh Rajan, Anna Puszkarska, Hartmut Oschkinat, Melinda J. Duer, *RSC Adv.*, **2019**, *9*, 26686-26690

Any published work included in this thesis has been reproduced herein with permission. Additional work that was carried out during the course of this PhD project, but not included in this thesis, has been published in the following peer-reviewed journals:

4. Jake L. Greenfield, Felix J. Rizzuto, Ieva Goldberga, Prof. Jonathan R. Nitschke, *Angew. Chem. Int. Ed.*, **2017**, *56*, 7541-7545

5. Karin H. Müller, Robert Hayward, Rakesh Rajan, Meredith Whitehead, Andrew M. Cobb, Sadia Ahmad, Mengxi Sun, Ieva Goldberga, Rui Li, Uliana Bashtanova, Anna M. Puszkarska, David G. Reid, Roger A. Brooks, Jeremy N. Skepper, Jayanta Bordoloi, Wing Ying Chow, Hartmut Oschkinat, Alex Groombridge, Oren A. Scherman, James A. Harrison, Anja Verhulst, Patrick C. D'Haese, Ellen Neven, Lisa-Maria Needham, Steven F. Lee, Catherine M. Shanahan, Melinda J Duer, *Cell Rep.* **2019**, *27*, 3124-3138

Formal Acknowledgements and Project Contributions

Aspects of the work presented in this thesis were undertaken as collaborative projects with other scientists working under the supervision of Professor Melinda J. Duer (Department of Chemistry, University of Cambridge). All contributions to these projects are explained in detail below and are highlighted at the point of their occurrence within each chapter. Unless stated otherwise, the work presented herein is that of the author, Ieva Goldberga.

The project presented in Chapter 6 was conceived and carried out by Ieva Goldberga, [U-^{13}C,^{15}N]-isotopically enriched peptides were prepared by Dr Dominique Bihan (Department of Biochemistry, University of Cambridge). Amino acid analysis presented in Chapter 5 were carried out by Dr Peter Sharratt (Department of Biochemistry, University of Cambridge). The DNP NMR were carried out with assistance from Dr Wing Ying Chow (Leibniz-Institut für Molekulare Pharmakologie, Campus Berlin-Buch). The TEM imaging were performed with assistance from Dr Karin Müller (Cambridge Advanced Imaging Centre, Univeristy of Cambridge).

Personal Acknowledgements

This particular section is the hardest part to write of my thesis, there are not enough words to help describe how lucky I have been through all these four years in the Duer group! Office 251 forever will stay in my heart!

I would like to thank Prof. Melinda Duer for taking me on as a PhD student, for all her support and advice through all these four years of research. I feel incredibly lucky to work with such unique, patient and very, very enthusiastic supervisor! Thank you for being a 'human being', I could not ask for a better mentor!

My sincere thanks goes to Dave for devoting time and patience to deal with all my questions, however trivial, and sparing much of his time to help me with improving my writing. Rakesh, for introducing me to the cell culturing and Uliana for helping me better understand biology, however hard and frustrating it can be. Rui for being my 'PhD buddy' and friend trough all these four years working together and trying trouble-shooting either of NMR machines, travelling together half of the world for conferences, and lifting up my spirits when needed, without you it would have been pretty boring. Anna for being not only a good colleague but also a friend and lighting up my mood! Karin, thank you for your help with the TEM imaging and entertaining stories during the imaging sessions. Thank you, Adrian, for keeping me and the rest of the group well-fed on your baked goods! And of course, I would like to thank Thomas, Sham, Tom, Holly and other students that have joined our group throughout the years that I had an opportunity to teach.

Outside of work, I would like to thank my family, sister and Jake and his family, for believing in me and encouraging me, even when I am not so sure and confident about myself. And friends who kept me sane.

And thanks goes to Shams coffee machine for providing refreshing coffee beverages and keeping me up when needed!

Table of Contents

Table of Abbreviations

Abbreviation	Definition
AFM	Atomic Force Microscopy
AGEs	Advanced Glycation End Products
BF	Bright Field
CD	Circular Dichroism
CP	Cross Polarization
Crh	Corticotropin-Release Hormone
CSA	Chemical Shift Anisotropy
DC	Direct Current
DMEM	Dulbecco's Modified Eagle Medium
DNP	Dynamic Nuclear Polarization
DOPA	3,4-Dihydroxylphenylalanine
DQ	Double-Quantum
DTPA	Diethylenetriaminepentaacetic Acid
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
EFG	Electric Field Gradient
FACITs	Fibril Associated Collagens with Interrupted Triple helices
FCS	Foetal Calf Serum
FID	Free Induction Decay
FSLG	Frequency-Switched Lee-Goldberg
FSOB	Foetal Sheep Osteoblasts
FT	Fourier Transform
HETCOR	Heteronuclear Correlation Spectroscopy
HPLC	High Performance Liquid Chromatography
ILT	Inverse Laplace Transform
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
LC-MS	Liquid Chromatography - Mass Spectrometry
LTMAS	Low-Temperature Magic Angle Spinning
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation - Time of Flight
MAS	Magic Angle Spinning
MD	Molecular Dynamics
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OI	Osteogenesis Imperfecta
PBS	Phosphate-Buffered Saline
PDB	Protein Data Bank
PDSD	Proton-Driven Spin Diffusion
PMMA	Poly(methyl methacrylate)
POST-C7	Permutationally Offset Stabilised C7
PREs	Paramagnetic Relaxation Enhancement

Table of Abbreviations

Abbreviation	Definition
PSC	Paramagnetic Chemical Shift
PSD	Proton Spin Diffusion
PTFE	Polytetrafluoroethylene
R5P	Ribose-5-Phosphate
RAGE	Receptor for Advanced Glycation End Products
RER	Rough Endoplasmic Reticulum
RF	Radio Frequency
S/N	Signal-to-noise
SEM	Scanning Electron Microscopy
SPINAL	Small Phase Incremental Alteration
SQ	Single-Quantum
SQ-DQ	Single-Quantum Double-Quantum
SSB	Spinning Sideband
ssNMR	Solid-State Nuclear Magnetic Resonance
TEM	Transmission Electron Microscopy
TEMPO	2,2,6,6-Tetramethyl-1-piperidinyloxy, Free Radical
TPPM	Two-Pulse Phase Modulation
TRITYL	Triphenylmethyl
TSP	Trimethylsilylpropanoic Acid
UA	Uranyl Acetate
UV-Vis	Ultraviolet - Visible
WB	Wide Bore
XRD	X-Ray Diffraction

Table of Amino Acid Abbreviations

Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1

Introduction

Type-I collagen is the most abundant structural protein in mammals. Collagen fibrils are found in nearly all structural tissues and play an essential role in mechanical and biological processes. Since the 1950s, when the collagen triple helix structure was first proposed,[1–4] many experimental techniques have been used to further study its structure and function in order to create a molecular model of collagen. Many sophisticated animal,[5,6] bacterial[7–9] and plant[10] models have been proposed to produce triple-helical collagens; and even further used for generating biomimetic models for medical applications.[11–13] Alterations of collagen at the molecular level, in vivo, through disease or ageing changes its physical and biological functions which has a significant effect on the proper cell function. However, the scientific community still lacks understanding of the collagen structure and dynamics at the atomic level in its native environment; in particular, understanding how collagen is changed through ageing and disease is deficient.

Multiple techniques have been used throughout the years to look at the hierarchal structure of collagen proteins. X-ray diffraction studies of rat tail tendon fibres determined a triple helical structure. Further microscopical investigations with TEM, AFM and SEM provided insight on the overall periodic morphology of collagen type-I fibrils and their further ordering within the tissues. All these studies have further driven people's curiosity to understand better the residue arrangement that results in the molecular ordering giving rise to the fibrillar structure that we know today. The synthetic approach of mimetic collagen peptides enabled the study of sequence effects in greater detail than what can currently be achieved in the study of the native collagen protein. This synthetic approach helped to identify some of the binding sites on collagen molecules, generating novel insights into the extracellular matrix protein interactions that are essential for the integrity of the extracellular matrix. With advances in today's technology, computational studies are becoming more advantageous as they can be employed in combination with MD simulations to acquire new understanding into the triple-helical folding and further towards fibril formation. All of these techniques can probe the structure of collagen at different length scales; however, structural studies of native collagen at an atomic scale resolution are still challenging. Further, none of these methods are suited towards the study of molecular dynamics at the time scales of the biological processes that collagen fibrils are involved in generating knowledge of molecular dynamics at relevant timescales, e.g. the timescale of collagen-ligand binding, would help us better understand physical and biological properties. Exceptions are AFM and MD simulations, but AFM tests mechanical properties and so only indirectly probes molecular dynamics, and MD simulations require experimental validation. There is thus a need for techniques that could further cover a broad scope of dynamics at different time-scales and at the same time provide insights into the atomic scale structure of collagen within its native form, i.e. in fibrillar arrangement within native ECM. Solid-state NMR spectroscopy can be used to complement and enrich our understanding of the structural details and dynamics of collagen at this much needed resolution. In this thesis, we use solid-state NMR spectroscopy to study the molecular structure of collagen within collagen fibrils and explore experiments to gain insight into the molecular motions of the collagen protein, again in native-like settings. An NMR relaxation-based signal separation method is first tested and investigated on small synthetic collagen peptides, and then on the native isotopically enriched collagen produced in the in vitro cell culture. The latter is further modified by glycation chemistry as a model of collagen dysfunction to gain a better understanding of the possible range structural changes in collagen pathologies and the impact on molecular motions.

Solution-state NMR spectroscopy has been used for decades to help solve atomic structures of many biomolecules. In addition, due to time-dependent phenomena in the NMR technique,

it is possible to study molecular dynamics of these biomolecules, such as protein folding, protein-protein interactions and reaction kinetics. Solution-state NMR spectroscopy is often used together with X-ray diffraction studies. The X-ray diffraction analysis provides a snap-shot of the molecular structural while NMR spectroscopy helps to validate it within a more native-like setting, i.e. aqueous solution; and further gives insight into structural variations possible from conformational dynamics. Furthermore, NMR spectroscopy has been invaluable where proteins cannot be crystallised. In the Solid-state, NMR spectroscopy has gleaned new insight into the structures of insoluble or non-crystalline and much bigger proteins. There are, of course, limitations to NMR spectroscopy as with any other technique. Depending of the atomic nuclei under study, NMR spectroscopy can be an inherently insensitive technique. For a better sensitivity, isotopic enrichment, e.g. of ^{13}C, ^{15}N and ^{2}H nuclei are required for NMR studies of low natural abundance NMR nuclei. Isotopic enrichment also enables use of 2D NMR techniques which can improve signal resolution and that can be beneficial in three-dimensional structural studies.

As it was mentioned before, the protein-ligand interactions depend on the collagen dynamics; further structural changes of collagen due to the pathological processes will change these interactions. In this thesis, we first focus on investigating what type of molecular motion information on the collagen can be obtained from the solid-state NMR relaxation studies. The first goal of this work is to better understand the collagen backbone dynamics. Therefore, we focus on a selection of synthetic collagen peptides, where we use ^{15}N to look at the collagen backbone. These mimetic peptides represent normal and pathological collagen sequences, namely mutated and disrupted sequences. We used these structurally different peptides to help us establish an ^{15}N relaxation scale of backbone motions possible within the collagen triple helix backbone. We then looked at [U-^{13}C,^{15}N]-isotopically enriched collagen itself with isotopically enriched hydrophilic residues and residues that are involved in glycation chemistry to assess the scale of structural and molecular changes possible in ageing and disease. Glycation is a non-enzymatic reaction between protein side-chain amine groups and reducing sugars. It is a pathological process and formation of the advanced glycation end (AGEs) products are thought to contribute to multiple diseases.[14,15] This chemistry leads to stiffening of the collagen fibrils, which is widely thought to be a result of molecular cross-linking. However, solid-state NMR studies have shown that the glycation chemistry leads to side-chain modifications rather than intermolecular cross-link formation. This further results in a molecular misalignment of the fibrils, suggesting that the effects of glycation on collagen molecular properties are rather more complex than previously thought.[16,17]

Collagen fibrils have a periodic structure consisting of alternating ‘gap’ and ‘overlap’ zones, where the ‘overlap’ zone is a more tightly packed region and has higher molecular ordering compared to the ‘gap’ zone. These two zones are thought to have different mechanical properties that have been extensively studied by AFM.[18–22] However, this method is only capable of probing the surface of the collagen fibrils, and little or no information on the interior can be obtained. NMR can yield more detailed information on the atomic structure of the collagen fibrils for all these zones. However, in practice, the resulting NMR signals corresponding to the different collagen regions are overlapping and cannot be resolved further. Therefore, we test a selection of paramagnetic tags that could potentially help to differentiate between the gap and overlap, or interior and exterior of the collagen fibrils. These paramagnetic complexes will influence the surrounding nuclei that will result in an observable change in the signal intensity or even chemical shift change.[23]

In this thesis, we show that solid-state NMR spectroscopy can be used to probe structural and molecular dynamics of the collagen protein in its native state and can provide an insight into the biological and disease processes on an atomic length scale, as well as dynamic processes at different time-scales. Solid-state NMR spectroscopy on its own cannot provide all the answers to the outstanding biological questions surrounding collagen proteins. Nevertheless, we believe that this technique can complement and enrich our understanding by providing a different perspective to the existing established techniques that are used to study the extracellular matrix.

2

Collagen and the Extracellular Matrix

Tissues in our bodies are not only comprised of cells; the bulk space around them is filled with extracellular matrix (ECM). The ECM is much more than just a scaffold that holds the cells in place, it is a complex dynamic network consisting of many extracellular macromolecules that are crucial for biological function of the cells. The detailed molecular structure of the ECM defines the environment of a tissues cells and their properties, including morphology, differentiation, and their metabolic status. Changes of the ECM structure through disease or ageing can alter the cell function and protein expression resulting in cellular dysfunction and further modification of the extracellular space. While all proteins and their interactions within the extracellular space are essential, the collagen proteins are the primary component of the ECM.

This chapter covers the biosynthesis of the fibrillar collagen proteins – the main protein of interest in this thesis - from triple helical molecular units to fibrils. The chapter ends with an overview of the most commonly used techniques to characterise and study these proteins.

2.1. Overview of Collagen

Collagen makes up to a fifth by weight of the human body and is found in most tissues. Collagen proteins have a diverse nature and composition, because of which they can serve many functions, such as providing structural and mechanical support for the surrounding cells, and roles in cell-to-cell communication.[24] The characteristic feature of collagen proteins is their unique triple-helical structure. This triple-helical structure is formed of three parallel polypeptide strands, each displaying in a left-handed helical twist, with a one-residue stagger-alignment affording a right-handed helix. The triple helix itself is held together by internal hydrogen bonds (Figure 2.1.).

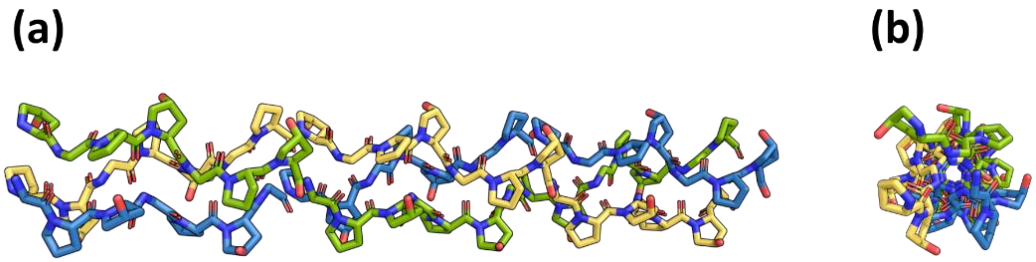

Figure 2.1. Crystal structure of a collagen triple-helix structure for a model peptide formed from three $(Gly\text{-}Pro\text{-}Hyp)_{10}$ repeat units of (a) side view and (b) top view. Each polypeptide chain is coloured differently, showing the left-handed twisting of each polypeptide chain, and right-handed supercoiling of the triple helix. The structure obtained from the Protein Data Bank (PDB) entry 1V4F.[25] Structure was rendered using PyMOL.[26]

Collagen triple helices can assemble into a variety of higher order structures such as fibrils, sheets, and networks. Each type of collagen is genetically different and can be classified according to the polymeric structures they form. Overall there are twenty-eight known types of collagen which can be divided into six classes based on their higher-order structural features (Figure 2.2.): (a) fibrils (types I, II, III, V and XI), (b) network-like structures (types IV, VIII and X), (c) beaded filaments (type-VI), (d) collagens associated with fibril surfaces (FACITs) (types IX, XII, XIV, XVI and XIX), (e) anchoring fibrils for basement membranes (type-VII), and (f) collagens with transmembrane domains (types XIII and XVII). Types XV and XVIII only have been recently identified; thus, they have been only partially characterised.[27]

The most abundant is type-I, a fibril-forming collagen which accounts, for instance, for 90% of bone matrix protein.[28]

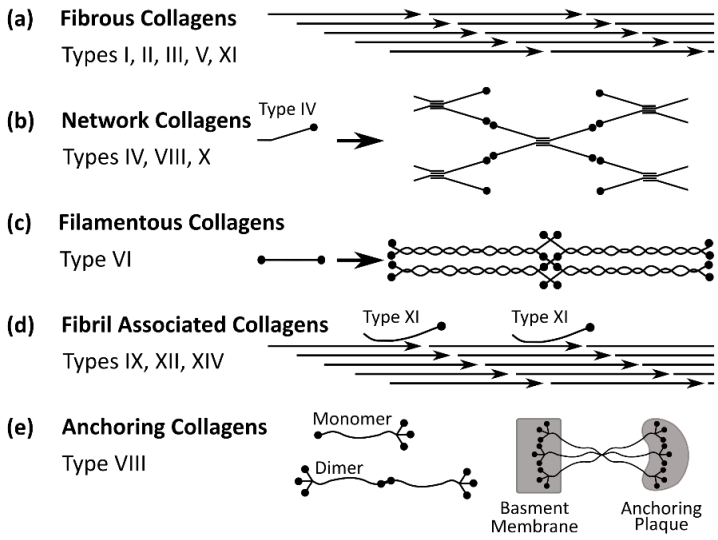

Figure 2.2. Classes of the collagen superfamily according to their higher-order structures. Figures adapted from references [27,29].

Each polypeptide chain of the collagen triple helix is known as an α-chain, and each collagen type has its own different polypeptide α-chain composition. The length of an α-chain varies from 662 to 3152 amino acids.[30] The self-assembled triple helices can be either homotrimers, where all three polypeptide α-chains are identical, or heterotrimers, where at least one of the α-chains is different.[31,32] Collagen type-I, the subject of this work, is a heterotrimer compromising two α1(I)-chains and one α(2)(I)-chain. The collagen triple helices of collagen type-I can further self-assemble into microfibrils, that further assemble into collagen fibrils in vivo that can be up to a few microns in length. Collagen type-I is an integral part of all structural tissues, such as tendon, vascular tissue, the calcified tissues, bone, teeth and even in the eye. The following sections will provide a more in-depth discussion of the structure of the collagen triple helix, its biosynthesis and modifications that develop its biological functionality.

2.2. Collagen Triple Helix and Fibrillar Structure

Folding of the collagen triple helix requires glycine to be present as every third amino residue due to the steric constraints. This repetition enforces so-called repeating units GXY on the primary structure, where G is glycine (Gly, G), and X and Y are other amino acids, and therefore collagen sequences can be represented as $(GXY)_n$ where *n* is a number of the tripeptide units between 337 and 343 in the fibrillar collagens. The most common amino acid

in the X-position is (2S)-proline (Pro, P) and that in the Y-position is (2S,4R)-hydroxyproline (Hyp, O). Gly-Pro-Hyp is the most frequently occurring triplet in native collagen type-I and makes up to 10.5% of the sequence.[1,3,33] Each collagen type-I molecule has ca. 3000 amino acids, resulting in a helical length of ~300 nm and diameter between 1 – 2 nm. Besides imino rich amino acids (Pro and Hyp), there is a still significant proportion of other amino acids that occupy the X and Y positions, except cysteine and tryptophan which are absent in the triple-helical region of collagen molecules.

Early fibre X-ray diffraction data showed that the triple helix has a 10/3 helical symmetry (sometimes denoted as 10_7) with a 28.6 Å axial repeat, which indicates that each polypeptide chain has 10 amino acids for 3 turns of the helix,[4] meaning that the triple-helical structure has 30 amino acids per turn. However, X-ray diffraction studies on synthetic collagen-like peptides have shown a 20.0 Å axial repeat in these synthetic molecules, which results from 7 amino acids in 2 helix turns.[34,35] This tighter conformation is denoted 7/2 symmetry, or 7_5. The synthetic collagen-like peptides are short triple-helical segments, typically consisting of around 30 amino acids per polypeptide chain. They consist of $(GPO)_n$ repeats at the beginning and end of the mimetic sequences to ensure the trimerisation into triple helical structures and to help maintain their structural stability in solution. Different helical pitches have been found within these model synthetic triple helical peptides, varying between 10/3 and 7/2 and are due to the variations in the sequences studied. The model peptides have more of the imino rich amino acids present in the sequence than native collagen and the helical pitch varies depending on the number of the Pro and Hyp residues present in the particular section of the triple helix.[36] It is clear then that Pro and Hyp are central to defining the triple helical structure; these residues will therefore be used in this thesis as monitors of collagen backbone conformation.

The structure of the collagen triple helix is stabilised by the inter-α–chain hydrogen bonding between all three α–chains. There is one hydrogen bond per GXY triplet that is formed between the carbonyl group on the X position residue and the amide (NH) on the Gly residue on the opposite α–chain: $C{=}O_X$---H-N_G (Figure 2.3.).[4] Substitution of Gly with any other amino acid is expected to inhibit hydrogen bonding formation resulting in a change in the local structure of the collagen triple helix. Such changes can further affect the collagen fibril assembly, stability and major ligand binding with other extracellular proteins.[37,38] These changes are pathological and are associated with various diseases, like osteogenesis imperfecta (OI), Ehlers-Danlos syndrome IV and chondrodysplasia.[39–42] In this thesis, the effect of Gly substitution will be explored in Chapter 6 in order to investigate the structural and dynamic changes on the collagen back bone via ^{15}N T_1 relaxation studies.

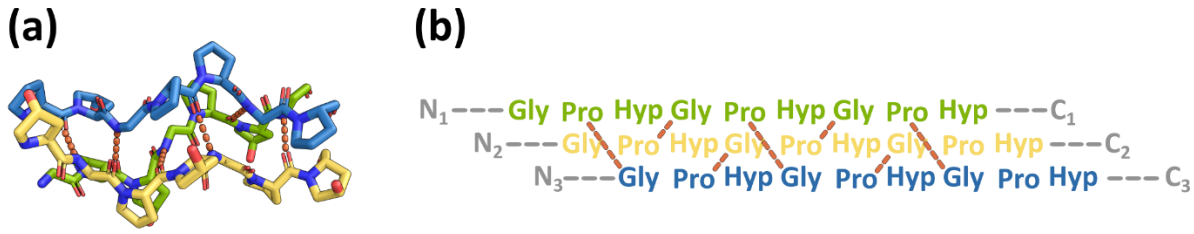

Figure 2.3. (a) Crystal structure of a collagen triple helix showing only three GPO repeats with inter-chain hydrogen bonding highlighted with red dashes. PDB entry 1V4F.[25] Structure was visualised using PyMOL.[26] (b) Schematic representation of the sequence shown in (a) with hydrogen bonding highlighted with orange dashes, where Gly is glycine, Pro is proline and Hyp is hydroxyproline.

Current understanding of the collagen proteins structure comes from the X-ray diffraction of the rat tail tendon.[43] The given resolution of the diffraction data is not enough to gain an insight at atomic scale resolution. Nevertheless, it gives a spatial arrangement of the triple helices and unit cell parameters, which are for the axial and lateral resolution of 5.16 and 11.1 Å, respectively (Figure 2.4.).[43] Five staggered tropocollagen units form a microfibril, which is arranged in a periodic way known as 'D-periodicity' (Figure 2.4. (a)). Each D-period is 67 nm in length, consisting of two regions known as 'gap' and 'overlap' zones, where gap zone is known to be wider with the length of 0.54 D. Further, these microfibrils are compressed in a quasi-hexagonal array forming collagen fibrils (Figure 2.4. (b)). The supercoiled structure is further supported by enzymatic cross-link formation at the N- and C-terminal telopeptides (see Section 2.3.2.). The fibrillar collagen composition can be varied, type-I fibrils can also contain other types of collagens, like type-III which is located near the fibril surface, and type-XI which is thought to help with initial nucleation of the fibril formation. The size of the collagen fibrils and the level of cross-linking will determine the physical and mechanical properties of the tissue. It has been observed that with ageing, the diameter increases.[44] Fibrillar growth is suggested to be controlled by various extracellular components.[45] Non-collagenous proteoglycans are known to bind to the collagen surface at specific locations. It has been suggested that proteoglycans are involved in the lateral growth of the collagen as well as the development of their form.[46] On the other hand, it has been observed that the fibrillar size varies with the concentration of N-propetide, where higher concentrations result in a smaller diameter of the collagen fibrils.[47]

The fully formed collagen fibrils can be recognised in atomic force microscopy (AFM) and transmission electron (TEM) images by distinctive 67 nm periodic banding patterns that arise from the stagger-like arrangement of collagen molecules within the fibrils (Figure 2.5).

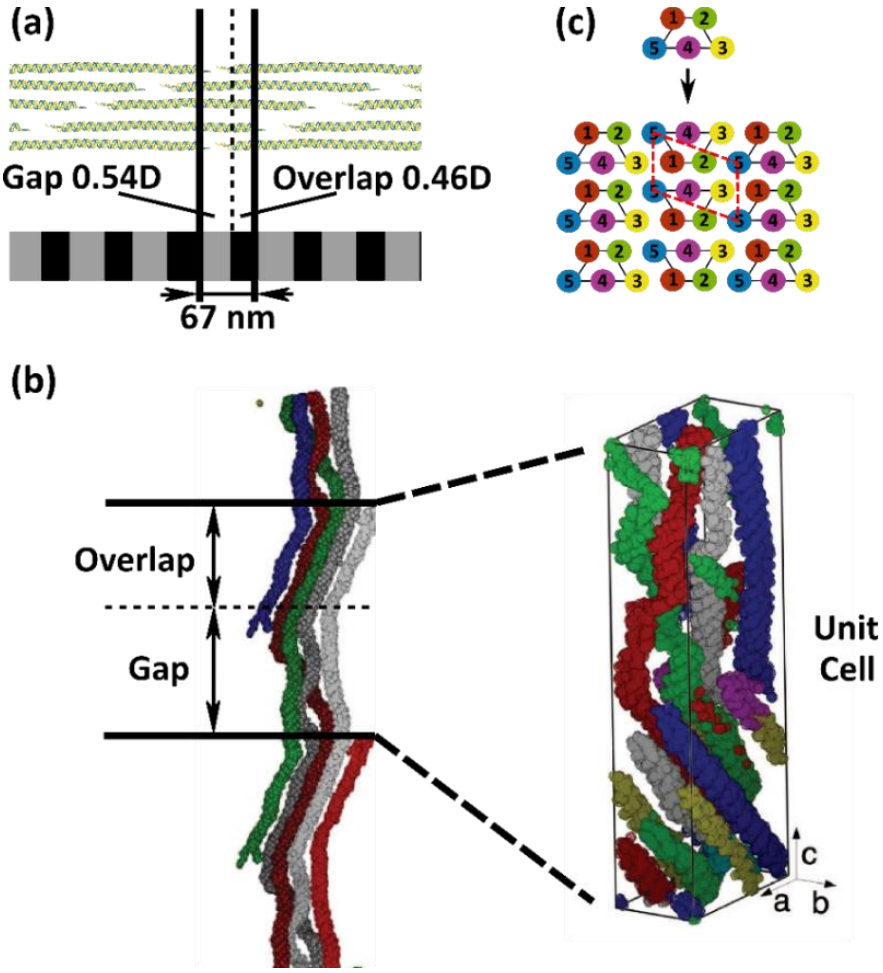

Figure 2.4. (a) Five staggered tropocollagen units arranged in a staggered D-periodic array shown, with D = 67 nm. The D-period consists of a region where tropocollagens are all overlapping known as 'overlap' zone, and reign where gaps are observed between the units also known as 'gap' zone. Gap and overlap zones are 0.54 D and 0.46 D, respectively. The tropocollagen units here are shown in a straight arrangement in relation to one another for clarity; they are twisting around each other in a supercoiled structure shown in (b); packing in a unit cell is shown as well on the right. Structures shown in (b) were adapted from a reference [43]. (c) On the top is shown a quasi-hexagonal arrangement of the five tropocollagen units with respect to one another. On the bottom is represented a packing of the collagen fibrils with unit cell highlighted with a rectangle. This figure was adapted and derived from references [43,48].

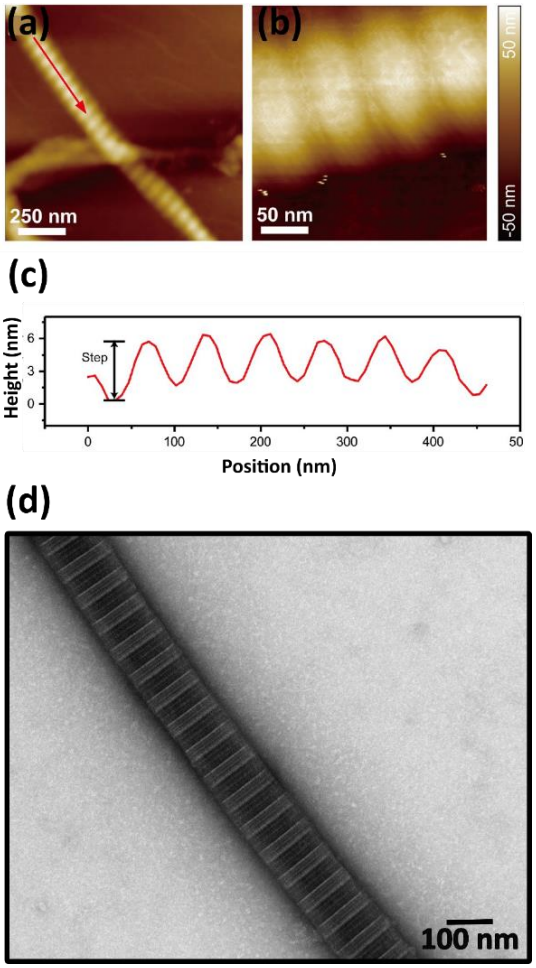

Figure 2.5. (a) and (b) Atomic force microscopy (AFM) height images of collagen type-I fibrils. Red arrow indicates position from which the height profile shown in (c) is taken. (c) The height profile from (a) showing a periodicity of 67 nm. The atomic force microscopy (AFM) images and the height profile was taken from a reference [49]. (d) Bright-field transmission electron microscopy (BF-TEM) image of a negatively-stained collagen fibril using uranyl acetate to enhance the contrast between overlap and gap zones. In the TEM image, gap zones are observed as darker regions due to the staining technique used, where the electron-dense uranyl acetate is accumulated in the atomically-less dense gap regions. The characteristic 67 nm periodic banding pattern characteristic of collagen type-I fibrils is observed in all three images.

2.3. Biosynthesis of the Collagen from α-chains to Fibrils

Biosynthesis of collagen is a highly complex, multistep process carried out within the cell and extracellularly space.[24,31,32] This section describes briefly the biosynthesis of the collagen into the fibrils, its structural features, such as posttranslational modifications of hydroxylation and glycosylation, that are important for the fibrillar structure and stability.

The transcription of the collagen genes from DNA to mRNA occurs in the cell first, where each α-chain is encoded by its specific gene forming prepro-α chains (Figure 2.6.). Then the collagen mRNA migrates to the rough endoplasmic reticulum (RER), where translation from mRNA to the individual α-chains occurs. After the translation process, the chains undergo post-translational modifications, where some of the lysine and proline residues are hydroxylated by the hydroxylase enzymes followed by further glycosylation on some of the hydroxylated residues. More details about hydroxylation and glycosylation are covered in Section 2.3.1.

At this stage, the polypeptide chains consist of triple and non-triple helical sections. The non-triple helical sections are known as N- and C-propeptides, which help to prevent the aggregation and misfolding of the triple-helical strands within the RER and the cell (Figure 2.6.). Formation of the collagen triple helix starts with nucleation at the C-terminal by forming a globular complex that pre-organises the three α-chains for the trimerisation process. Once the procollagen triple-helical unit has formed, it is transported by the vesicles from RER to the Golgi apparatus followed by secretion in the extracellular space.

In the extracellular space, the propeptides are removed by proteolytic enzymes to enable the formation of the stable collagen fibrils. Each end is cleaved by an appropriate enzyme, known as N- and C-proteinase. The tropocollagen units consists of triple-helical region and non-triple helical the N- and C-telopeptides (Figure 2.6.). Formation of the fibrils happens via aggregation of the telopeptides, direction and polarity of this process can be controlled by the cells.

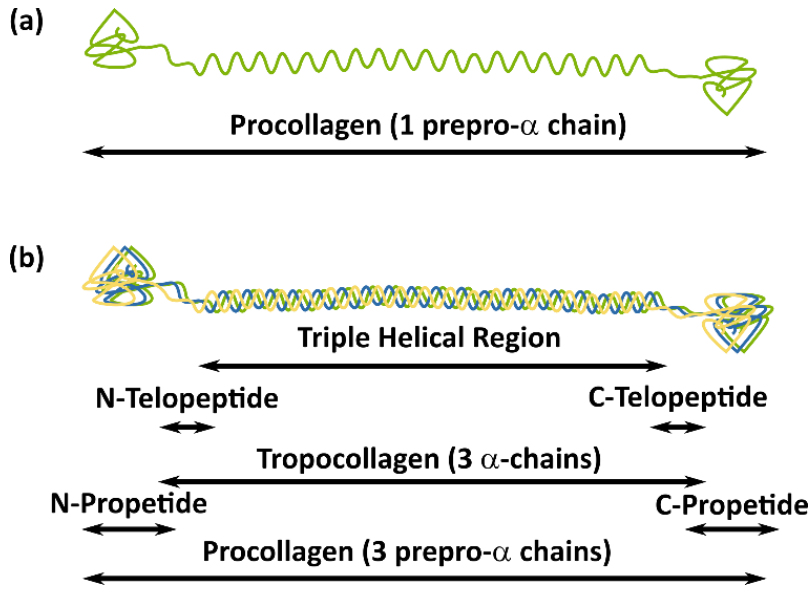

Figure 2.6. The representation of the (a) prepro-α chain and (b) the final procollagen unit that is secreted by the cell in the extracellular space. The procollagen unit consists of N- and C-propetides that are cleaved by proteolytic enzymes leaving N- and C-telopeptides forming a tropocollagen unit.

2.3.1. Post-translational Modifications

There are two type post-translational modifications, hydroxylation and glycosylation, where both of these modifications are enzymatic processes.[31,32] Post-translational hydroxylation of the majority of Y-position prolines is essential for the proper formation of the collagen triple helices. The Hyp ring adopts an exo ring conformation that is essential for pre-organising the alpha chain backbone for the further folding of the three peptide chains into a triple helix (additional details in Section 2.5.4.).[50–54] Furthermore, the Hyp increases the thermal stability of the collagen triple helices.[55] From mimetic peptide studies used in this thesis, it was found that in the peptide containing only GPO repeats, the melting temperature was higher by 20 °C compared to the peptide containing only GPP triplets (measured on peptides using circular dichroism (CD) spectroscopy, see Sections 5.7 and 6.4). Variations in the hydroxylation of the Pro depends on the tissue type, as well as species. For mammalian collagen type-I, there are about 90 Hyp residues per 1000 amino acids in collagen type-I.[30]

Hydroxylation of the lysine is the other important post-translational modification of collagen type-I and occurs more rarely than proline hydroxylation. Variations in the number of Hyl residues in tissues are even greater compared to the Hyp residues. In type-I collagen, there are around 6 – 17 Hyl residues per 1000 amino acids in the polypeptide chain, depending on the tissue and the age of the animal.[30] Hyl is important in providing sites for collagen glycosylation and enzymatic cross-linking. Glycosylation occurs at specific sites in a

conserved sequence of $GK^{H}X$, where K^{H} is Hyl.[56] The levels of collagen glycosylation are low, for example, in human bone there are 0.6 – 2 Hyl glycosylated per one collagen molecule.[57] Besides enzymatic reactions between the collagen molecules and the sugars, collagen can undergo a non-enzymatic glycation process with ambient sugars present in the extracellular space (see Section 2.4.). Structures of hydroxylated Pro, Lys residues and glycosylated residues is represented in Figure 2.7.

The position of collagen glycosylations until now have been studied via use of the mass spectroscopy.[58] Recently, our group has shown that these species can be detected in [U-^{13}C,^{15}N]-Lys enriched intact ECM of mouse skin and from in vitro cell culture of bovine smooth muscle cells using solid-state dynamic nuclear polarisation (DNP) NMR.[59] This is the only technique, to date, that is capable of providing access to the structural consequences of the collagen modifications of intact tissues. However, the paramagnetic 'doping' might enable one to detect these modifications as well and study them at biologically relevant temperatures (see Section 8.4.), as the DNP NMR technique is limited to cryogenic temperatures.

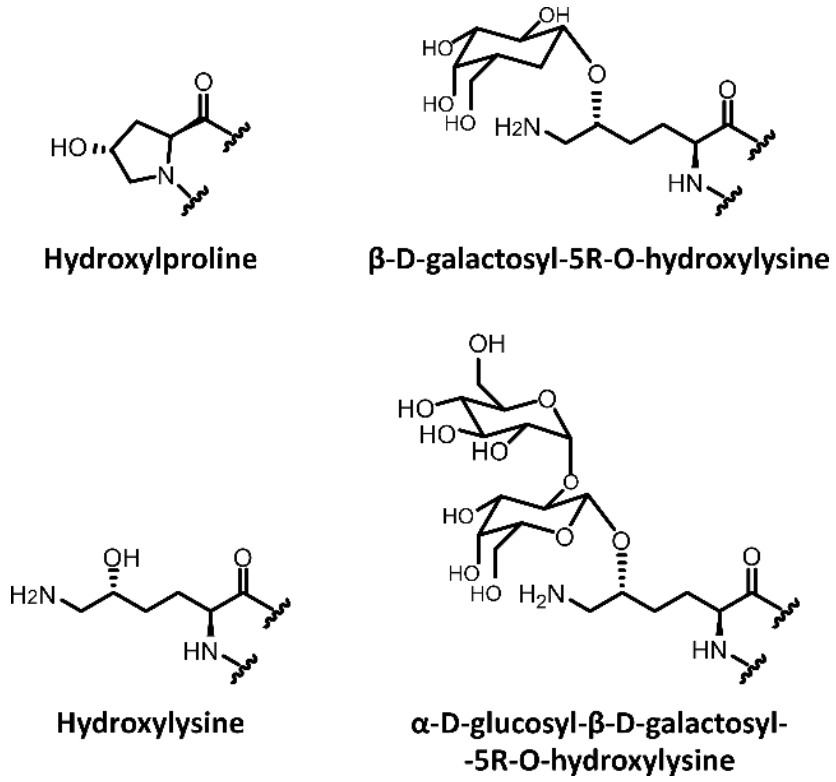

Figure 2.7. Structures of the hydroxyproline, hydroxylysine, and O-linked glycosylation of the hydroxylysine.

2.3.2. Enzymatic Cross-linking from Lysine and Hydroxylysine

The lysine and hydroxylysine residues are essential for the cross-link formation, which occurs at specific residues. Four specific cross-linking sites have been found in collagen type-I: one in each N- and C-terminal telopeptides, and another two in the triple-helical domains.[60] Cross-linking is catalysed by an enzyme called copper-dependent lysil oxidase.[61] The Lys and Hyl side-chain amino group undergoes deamination and forms reactive aldehydes. The further condensation reaction between an aldehyde and the Lys or Hyl can occur, forming divalent cross-links, also known as immature cross-links (Figure 2.8.). Further enzymatic maturation of cross-links can occur forming trivalent structures.[62] Divalent cross-links contain a Schiff base or a ketoamine group, where trivalent cross-links in the collagen type-I are shown to contain pyrrole and pyridinoline functionalities. Chemical structures of the cross-links are shown in Figure 2.8. The cross-linking of the collagen provides the fibrils with tensile strength and mechanical stability. The level of the cross-linking will depend on the tissue type. However, the overall amounts of the enzymatic cross-links, in general, are low in the collagen, and thus the detection even via solid-state NMR is challenging.

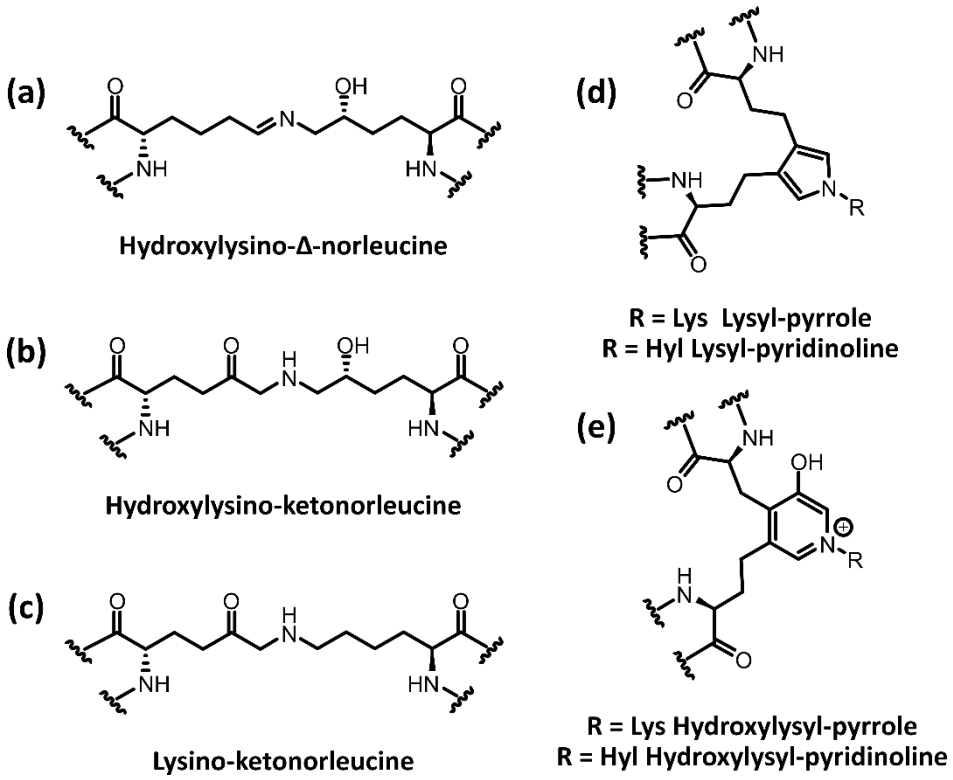

Figure 2.8. Structures of initial (a) – (c) divalent and (d) – (e) trivalent mature cross-links shown.[62,63]

2.4. Glycation

Glycation is defined as a non-enzymatic reaction between amine groups of proteins and ambient reducing sugars in the extracellular space. In the collagen free amino groups of Lys, Hyl and Arg residues are susceptible towards glycation that eventually results in the formation of advanced glycation end products (AGEs) which can lead to undesirable addition of sugar adducts to the affected collagen residues and cross-linking of collagen molecules and fibrils.[29] This reaction is a pathological process occurring with age. Importantly, it is more problematic in diabetic individuals who have high levels of glucose in the extracellular space.

During the glycation process, characteristic 'browning' of the proteins occurs, that is termed the Maillard reaction. Glycation reaction starts with the free amino group reacting with the free aldehyde group of the acyclic form of the sugar in a nucleophilic attack. The reaction itself is fast; however, the rate of the initial glycation step depends on the rate of sugar ring opening. Glucose, in general, has one of the most stable ring forms of all sugars; therefore, of all the aldose sugars, it is the least reactive.[64] Other sugars like ribose and ribose-5-phosphate (R5P) are more reactive as they have less stable sugar ring forms. Both these sugars are important cellular metabolites, for example, ribose phosphate is a component in RNA synthesis and is known to quickly react with cellular proteins.[65,66] The next step in the glycation reaction is the initially-formed Schiff base undergoing irreversible Amadori rearrangement forming early glycation intermediate Amadori products. These intermediates undergo further rearrangements, and irreversible oxidation and dehydration reactions finally forming AGEs. Some common AGEs formed in glycation chemistry are shown in Figure 2.9.

During ageing, the concentration of AGEs increases, as they cannot be broken down by enzymes within the cell, and the chemistry that led to their formation is irreversible; in return this changes the mechanical properties and can slow down remodelling. For example, bone becomes more brittle and stiffer after incubation with ribose.[67] In highly glycated ECM, cells change their behaviour which affects protein expression, cell differentiation and can cause apoptosis.[68] Receptors for the detection of advanced glycation end products (RAGE) located on the surface of cell membranes, can trigger inflammatory responses in cells.[69] It has been shown that the exposure to AGEs in the in vitro cell culture of mouse pre-osteoblasts (MC3T3-E1) inhibits mineralisation.[70]

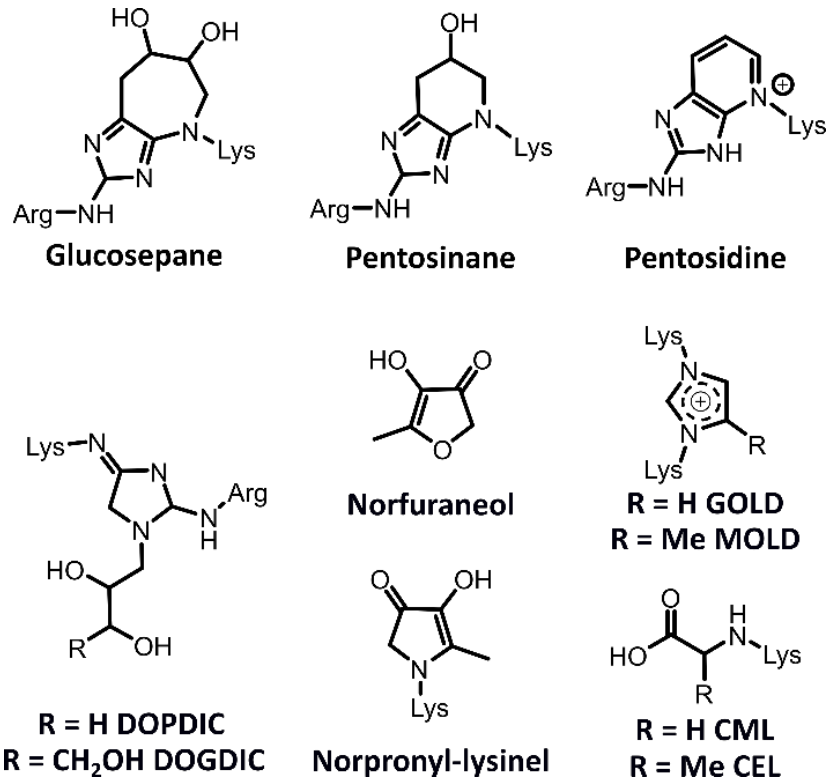

Figure 2.9. Chemical structures of some of the advanced end glycation products (AGEs) formed in the glycation process.

Glycation chemistry of collagen in the ECM is a highly complex topic. Products generated during the glycation chemistry causes structural changes of the collagen fibrils resulting in stiffness, that is thought to be due to the increase of the AGE cross-linking. In recent studies of extracellular matrix that was glycated using ^{13}C isotopically enriched sugars (ribose and R5P) and examined by solid-state NMR spectroscopy has revealed that the vast majority of the reaction products are side-chain modifications and not cross-links. This is presumably due to the lack of a suitable cross-linking partner within the spatial constraints forced by the collagen fibril structure.[16,17] In this thesis, glycated collagen with R5P was used as model of dysfunctional changes to collagen molecular structure to understand the scope that such changes can have on the molecular dynamics of collagen.

2.5. Techniques for Studying the Structure of Collagen

The collagen proteins have been studied by a wide variety of different characterisation techniques at different length-scales. In this section, the most commonly used characterisation techniques will be considered and discussed. Microscopy techniques such as TEM, AFM and SEM offer resolution of collagen nanostructures and aggregates at nanometre length scales. X-ray diffraction (XRD) studies have been used to obtain crystal structures of many biological

macromolecules, allowing determination of their conformation and location of specific residues. However, in the case of the native fibrillar collagen, the visualisation at this atomic resolution is not possible due to the size and the disorder of the collagen monomer unit. Nevertheless, XRD has provided structural information on the molecular packing within type-I collagen fibrils.[43,71] None of these structural techniques however offer dynamic information. NMR spectroscopy can give more detail on the local structural information at the atomic length scale and via use of specific experiments give insights into molecular dynamics. Furthermore, computational studies have been used to analyse sequence of the collagen proteins and study folding dynamics of the collagen triple helix of normal and mutated sequences.[72] It is important to verify the results of computational simulation with experimental data. This verification can be complicated as no single experimental technique is capable of covering all structural length scales and the femto-nanosecond timescales of molecular motion that is studied using molecular dynamic simulations.

2.5.1. Scanning and Transmission Electron Microscopies

Scanning electron microscopy (SEM) is a characterisation technique that uses a focused electron beam which enables one to study the surface topology and composition of the collagen fibrils. SEM in collagen studies is primarily used to observe the fibrillar structure in various tissues and help to map fibrillar organisation (Figure 2.10.),[73,74] gathering insights on the structural framework that can be further used for preparing collagen-based scaffolds.[11–13] Transmission electron microscopy (TEM) also uses a focussed electron beam to visualise the substrate. This microscopy uses a smaller wavelength of electrons therefore it can offer a higher resolution of the collagen fibrils by providing very detailed information of the molecular organisation and morphology of the fibrils (Figure 2.5. (d)).[75,76] TEM is more commonly used to assess fibrillar shape and diameter.[77] TEM microscopy enables more accurate measurement of the fibril dimeter than SEM studies. [78]

The key drawback to both of these techniques are associated with the sample preparation procedures which require the samples to be dehydrated. This dehydration step can disturb the native state of the collagen fibrils. As a result, the measured D-banding of the collagen fibrils by SEM can be less than 67 nm.[78] For TEM studies, very thin sections of the samples can be investigated (50 – 70 nm) and heavy atom dyes like uranyl acetate are used to improve contrast in the images. All of these procedures can introduce changes on the native collagen structure and therefore might not be an accurate representation of fibrils in a native state.

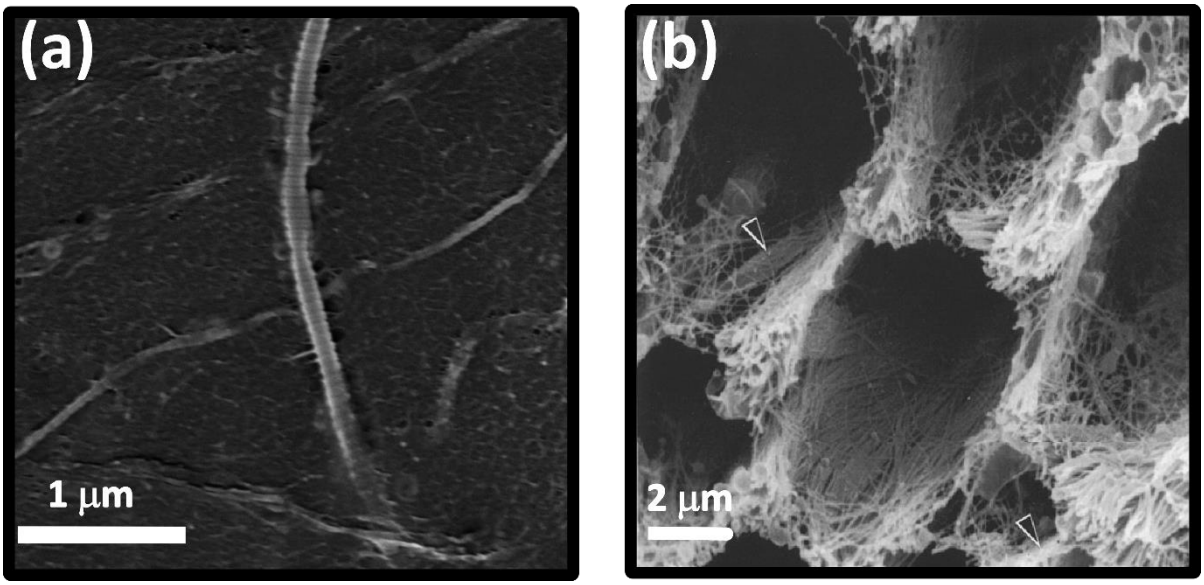

Figure 2.10. Example SEM images of collagen materials shown. (a) SEM image of reconstituted collagen fibril from the bovine Achilles tendon with banding pattern shown. SEM image was recorded by Sham Al Zahabi (Chemistry Department, University of Cambridge). (b) SEM image of endoneurial collagen fibrils from cylindrical tubes housing individual nerve fibres (removed), network of collagen fibrils shown. Image was adapted from reference [73].

In this thesis, TEM imaging was used to help assess where heavy paramagnetic dyes are concentrated on collagen fibrils. The paramagnetic dyes are employed as NMR spectral editors, and we hypothesised that determining their spatial distribution by TEM would allow us to assign NMR signals to specific regions of collagen fibrils (Chapter 8.). TEM was also used in this thesis to assess the overall morphology of in vitro cell culture produced collagen fibrils after the rest of the ECM was digested to ensure that the fibrils are not denatured by the procedure (Sections 5.2.4.).

2.5.2. Atomic Force Microscopy Studies of Collagen Fibrils

Atomic force microscopy (AFM) uses an atomically sharp tip that is appended to the end of a cantilever. The tip is moved across the specimen in grid like fashion and the forces exerted on the cantilever by the interaction with the substrate surface via feedback circuit is measured, thus gleaning information on the surface topology of the specimen under study.[79] Hence, AFM can give an image of the overall topology of the collagen fibrils.[80,81] Examples of AFM images of collagen fibrils are shown in Figure 2.5.

There are three types of operating modes in AFM – contact, non-contact and tapping modes. The most commonly used mode for the collagen studies is the tapping mode, where an oscillating cantilever at specific frequency is used. Besides visualising the surface of the

sample AFM can be used to measure mechanical properties such as stiffness and elasticity of the collagen fibrils at nano- and micromechanical scale.[22,79,82] AFM studies have revealed that the gap and overlap zones of the collagen fibrils have different mechanical properties, where the overlap zones are much stiffer along the axial direction.[22] At the macroscopic scale, AFM has shown that collagen fibrils have tube-like mechanical properties where fibrils flatten under compression.[83] Sample preparation for AFM is much less invasive compared to that of TEM and SEM. The imaging can be done in air, vacuum or even in a liquid environment. Therefore, AFM has been used to study the self-assembly of the collagen fibrils in a buffer solution.[84]

One of the limitations of the use of AFM studies on collagen fibrils is that it cannot provide information on the interior of the fibrils. Further, only thin layers of fibrils can be studied. Additionally, fragments from the collagen fibrils can stick on the AFM tip and build up over time resulting in loss of resolution.

2.5.3. X-Ray Diffraction Studies on Collagen Fibrils

X-ray diffraction studies of the fibrillar collagens have been instrumental in elucidating the structure of the collagen triple helix. The triple helical structure was first proposed by Pauling and Corey[1] in 1951, further redefined by Ramachandran and Kartha[2] in 1954 and finalised by Rich and Crick[3,4] in 1961. Crystallographic studies in general are complicated for heterogeneous materials such as tissues and ECM samples due to lack of long-range order. Nevertheless, tendon materials provide a rare case of long-range order within the ECM that has been used to gain insight into the fibrillar collagens. These studies have shown that the fibril gap zones have greater molecular disorder in part because they contain the non-triple-helical telopeptides.[85] Furthermore, X-ray diffraction studies have confirmed that collagen molecules are packed in quasi-hexagonal array within the fibrils (see Figure 2.5.).[43] However, the tropocollagen units cannot be isolated from the fibrils as they are cross-linked within the collagen microfibrils and even then these units are not well ordered and are too large. Further structural studies on disease modified collagens can be expected to be even more complicated, if the sample becomes more heterogeneous and disordered then the X-ray diffraction becomes less useful.

X-ray diffraction studies have been insightful in studying synthetic triple-helical collagen peptides due to the small size and higher degree of inherent order. More details on the mimetic collagen sequences involved in these studies are presented in Section 2.6.5.

2.5.4. Synthetic Collagen-like Peptides

Synthetic collagen-like model peptides are small representations of the triple-helical native sequence in length of 30-60 amino acids per polypeptide chain.[86] These mimetic peptides contain high amounts of GPO triplets at the ends of the peptide sequences to help with self-assembly and give higher thermal stability. In the middle of the triple helix, a guest sequence of interest can be introduced, for example, the sequence of an integrin binding site.[87–90] These peptides have been extensively studied by X-ray diffraction, where variations of different native sequences and even impact on the pathological effects on structure stability have been investigated.[86,91] Solution and even solid-state NMR studies have been carried out for these mimetic peptides. In this thesis, a selection of synthetic model peptides which represent different variations of the native sequence was employed to study backbone dynamics via ^{15}N solid-state NMR (Chapter 6.). The following sub-sections provide an overview of the types of peptides studied in this thesis.

2.5.4.1. Imino Rich Peptides

Imino rich peptides used in this thesis were selected to gain a better understanding of tightly packed collagen triple helices. These triplets are highly conserved in the native collagen type-I sequence, e.g. in mammalian sequences in the α1 chain there are 41 GPO triplets conserved out of 43, and in the α2 chain, 24 of 33.[92] Studies so far using solid-state NMR and MD simulations have led our group to hypothesise that these imino-rich regions act as “expansion joints” to maintain molecular ordering within the fibril. This is due to the Pro residue in the GPO triplet being in a metastable structural state, where it dynamically samples a range of endo and exo conformations.[92,93] These results suggest that these triplets in the collagen molecule can provide regions where the molecule can rapidly expand, compress or bend due to the flexibility provided by Pro in any GPO triplet being able to convert between both conformations with no significant change in energy. The clustering of GPO triplets in collagen fibrils (see Figure 2.11.) suggests that this molecular flexibility property is conferred across the fibril and may provide site specific flexibility around cell binding sites.

X-ray diffraction structural studies on these imino rich model peptides have shown that prolyl rings in the Y position have a strong preference towards the exo conformation (up to 95%)[93] with hydroxyprolines further stabilised in the exo conformation by the gauche effect and $n \rightarrow \pi^*$ interaction.[50–54] This exo conformation of the ring is known to stabilise the molecular structure of collagen triple helix, and propagates the organisation of the backbone structure towards an ideal collagen 7/2 helical symmetry.[33,94,95] Prolines in X position in GPY triplets favour the endo

conformation, but in GPO triplets the ratio between endo and exo conformations is almost 50:50 (from X-ray diffraction studies of model peptides (PDB 3B0S)[96] and ssNMR studies).[92,93]

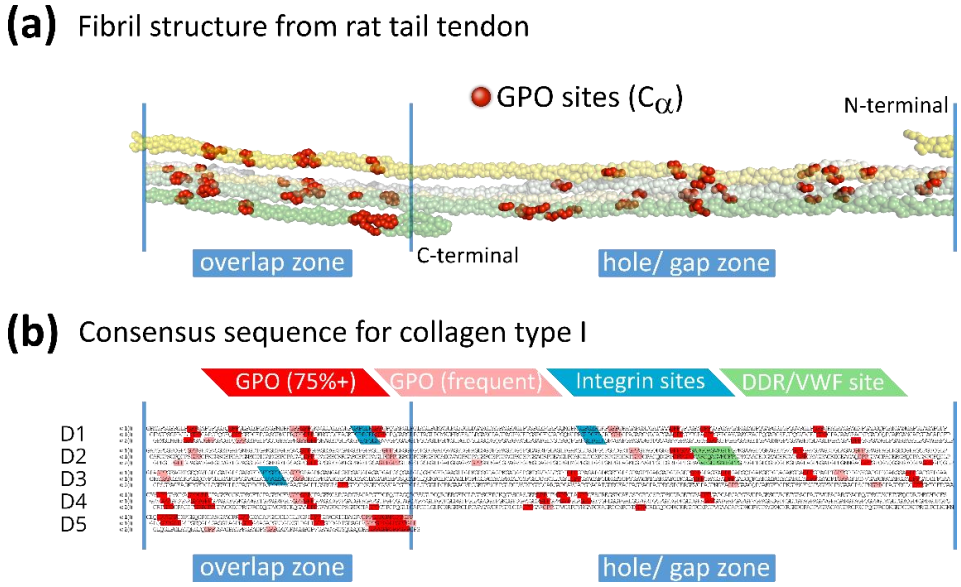

Figure 2.11. (a) Collagen type-I fibril structure modelled from X-ray diffraction data on rat tail tendon; derived from PDB code 3HR2.[43] Only Ca atoms are shown (as spheres); the red spheres represent Cα atoms of the residues in GPO triplets. (b) The collagen type-I sequence laid out as molecules are arranged in the fibrils with the most frequent amino acid at each position across all mammalian species, shown with stagger between the three polypeptide chains that make up the triple helix. Highly conserved GPO triplets (75%+) within collagen type-I mammalian sequences are shown in red, others in pink; integrin binding sites are highlighted in blue, and the DDR/VWF binding sites are in green. (a) and (b) are from reference [92].

2.5.4.2. Peptides with Gly → Ala Mutation

In this thesis, two pathological collagen sequences were investigated, one with a Gly-Ala mutation in each alpha chain and the other with deletion of one amino acid from the native GXY triplet structure of the collagen triple helix sequence. Mutations of Gly to either Ala, Asp, Glu, Cys, Ser, Arg and Trp in the native collagen sequences are pathological, causing diseases such as osteogenesis imperfecta (OI), Ehlers-Danlos syndrome IV and chondrodysplasia.[39–42] In these diseases, the mutation can cause kinks in the collagen triple helix that can affect its stability and cause formation of irregular collagen fibrils resulting in altered morphology and stability,[37,97] and further affecting major ligand binding site interactions with other proteins, such as integrins and MMPs.[38] X-ray diffraction studies of the Gly to Ala mutated model peptides have shown that this mutation destabilises the collagen triple helix. The methyl group of the Ala side-chain is pointing into the core of the triple helix, which results in all three strands being pushed apart (Figure 2.12.).[98,99] The stabilising hydrogen bonding between interchain Gly N-H and Pro C=O that is present in the $(GPO)_{11}$ peptide is replaced by water-mediated interchain hydrogen bonds between Ala N-H and Pro C=O (Figure 2.12.

(b) and (c)). Furthermore, untwisting of the triple helix in the Ala-mutated peptide leads to the three polypeptide chain ends being out-of-registry. Solution-state NMR studies have shown that the three Ala residues are non-equivalent, compared to the symmetric $(GPO)_{11}$ peptide and that only one of the three Ala amides can form an internal hydrogen bond.[100]

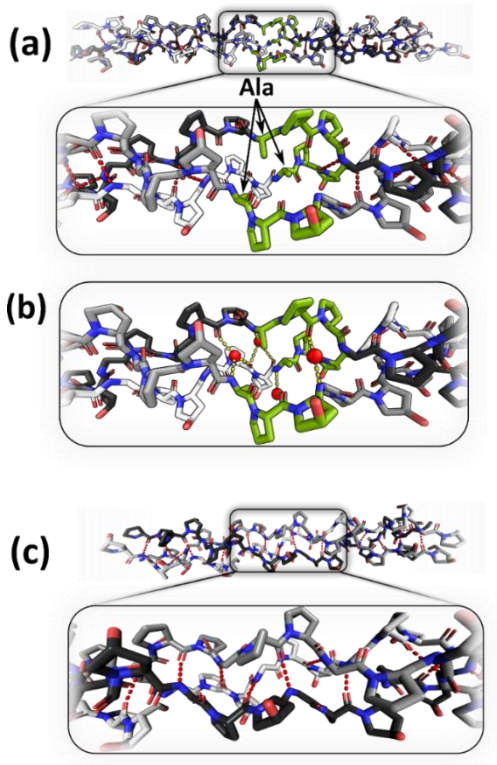

Figure 2.12. (a) Top: crystal structure of $(GPO)_5APO(GPO)_5$, with APO triplets coloured in green, PDB code 1CGD.[101] Bottom: close-up of the APO substitution site in the middle of the sequence; Ala is highlighted with black arrows. Interchain hydrogen bonding is represented with red dashed lines. Note that there is no interchain hydrogen bonding in the APO substitution site. (b) Water-mediated hydrogen bonding shown with water molecules represented as red spheres. (c) Top: sequence of the $(GPO)_{11}$ shown with interchain hydrogen bonding highlighted with red dashed lines; PDB code 1CAG.[25] Bottom: close-up of the middle of the chain representing intact hydrogen bonding in the $(GPO)_{11}$ peptide. Structures were rendered using PyMOL.[26]

2.5.4.3. Peptides with One Residue Deletion

Destabilisation of hydrogen bonding is also observed for the interrupted peptides, where one residue is removed from the polypeptide sequence. Deletions in the X or Y positions in the collagen native sequences are common and can occur naturally, especially in non-fibrillar collagens. In collagen type-VIII, there are eight such deletion sites, which are associated with specific higher order structural features in the form of highly extended hexagonal arrangements of the triple helices.[102] More than 20 breaks in the GXY sequence are observed in collagen type-IV.[103,104] These interruptions in the native sequence appear to be site-specific and are necessary to maintain higher-order structure (sheet-like molecular arrangements) and

proper function.[105] However, additional interruptions can be pathological.[106,107] Deletions in collagen type-I do not occur in the native sequence and are always pathological, for instance, they can lead to impaired triple helix formation, reduced extracellular secretion and alpha-chains being post-translationally over-modified resulting in several different types of OI.[108] The X-ray crystal structure of the triple helical deletion peptide with a sequence of $(GPO)_5GPGPO(GPO)_4$ shows two triple helices in the unit cell that have similar conformations, but they are not identical with all six polypeptide chains different from one another.[109,110] The interruption site shows an unusual out-of-phase hydrogen bonding motif (Figure 2.13.), where not all glycine residues are participating in the internal hydrogen bonding. X-ray diffraction results show that this deletion peptide exhibits greater variability of available conformation than the parent $(GPO)_{11}$ molecule.

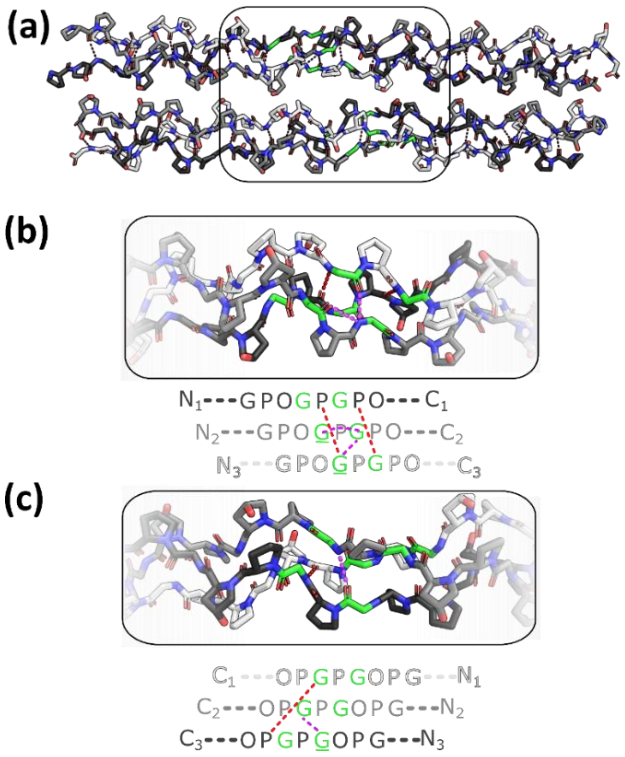

Figure 2.13. (a) The asymmetric unit of the $(GPO)_4GPGPO(GPO)_4$ peptide crystal structure; PDB code 1EI8.[109] Leading chains are coloured in white, middle in grey and lagging in dark grey. Gly in the deletion and post-deletion triplets are highlighted in green. Hydrogen bonding is shown with red dashes in the crystal structures. (b) Top: detail of the hydrogen bonding in the deletion site of the top peptide. Below: schematic representation of the sequence showing hydrogen bonding. Hydrogen bonds between Gly N-H and Pro C=O are shown in red, between Gly NH and Gly CO in purple. Gly highlighted with green underscores are hydrogen bond acceptors. (c) Top: Details of the hydrogen bonding of the deletion site of the bottom peptide in (a). Bottom: schematic representation of the sequence showing hydrogen bonding. Structures were visualised using PyMOL.[26]

2.5.4.4. Integrin Binding Site Containing Peptide

The last group of peptides studied in this thesis contain a specific amino acid sequence (GFOGER) that is recognised by the integrin $\alpha2\beta1$ I domain, found in the $\alpha1$ chain of type-I collagen.[87,111] Integrin binding is essential for cellular processes, e.g., cell adhesion, migration and cell differentiation. The X-ray crystal structure of this peptide has shown that there is no loss of the standard interchain hydrogen bonding between Gly N-H and C=O with residue in the X position.[112,113] The overall symmetry of the GPO triplets in the N- and C-terminal zones are close to 7/2, but the central zone of the sequence is slightly shifted away. However, it does not exhibit quite the 10/3 symmetry of natural collagen type-I triple helices either.[4,114] Furthermore, there are small kinks or bends at the edges of the central zone of the GFOGER that manifests in deviations of 9.08° and 8.48° respectively from the geometrically straight $(GPO)_{11}$ peptide. When bound to the integrin, this deviation becomes greater by 14.58° and 15.88° at the edges of the central zone. Figure 2.14. highlights the flexibility of this peptide backbone pre- and post-binding to the integrin.

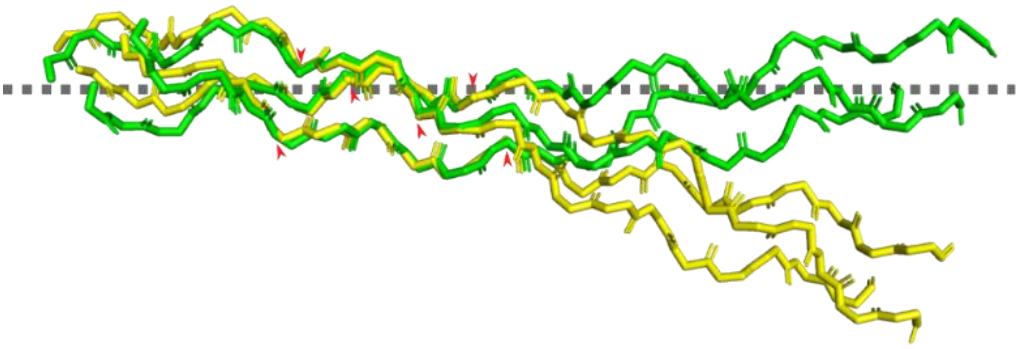

Figure 2.14. The figure is adapted from J. Emsley et al.,[112] modified using PyMol.[26] PDB code 1DZI (yellow) and 1Q7D (green). The figure shows main-chain atoms from the superposed unbound peptide (green), and the peptide bound to the integrin $\alpha2\beta1$ I domain (yellow); the $\alpha2\beta1$ I domain is not shown here for clarity. The peptides were superimposed based on the Cα atom coordinates of residues GFOGER from the central region, highlighted with red arrow heads. The grey dashed line represents the axis through the average coordinates of Cα atoms from equivalent residues on each strand in the central zone. A similar axis is calculated for the N- and C-terminal zones (GPO triplets before and after central zone) and these are used to calculate the angle of bends formed with the central zone. For the N and C-terminal zones, these are 8.48° and 9.08° for unbound and 15.88° and 14.58° for bound peptides. Structure was visualised using PyMOL.[26]

This synthetic approach was used in this thesis, together with [U-^{13}C,^{15}N] isotopic enrichment to study backbone dynamics of these peptides (see chapter 6. Table 6.1.). However, there are few things that one needs to keep in mind when results concerning the model peptides are processed. These mimetic helices contain high amounts of imino rich amino acids, therefore the NMR spectra obtained for these peptides might be different compared to the ECM

samples, where greater variation in the conformations of various residues can result in broader observed NMR signals.

2.5.5. Computational Studies on Collagen Structure

Computational studies has been used to carry out molecular dynamic (MD) simulations on the full length of native collagen sequence giving insights into the structure at different length-scales; however, it comes with a cost of computational time, that can be rather long, up to few months or even a year.[32,115,116] MD simulations have been used to assess the structural differences between collagen homo- and heterotrimers, where these studies have shown that homotrimers are more susceptible towards collagenase degradation, and more likely to unwind under the tensile forces that are applied.[72] Furthermore, more challenging sequences of the collagen triple helices can be investigated, such as osteogenesis imperfecta (OI) mutations of glycine residue to serine, where the trimerisation process of a synthetic model peptide was shown to increase two-fold compared to the normal collagen helix formation.[117] MD simulations have been applied to study glycated collagen-like peptides and have shown that the triple helix becomes more susceptible towards enzymatic digestion with collagenase.[118,119] The latest MD simulations performed on a collagen fibril surface model that is based on the X-ray diffraction data of the rat-tail tendon, showed that hidden binding sites become exposed after surface reconstruction as a result of molecular motion.[49]

The computational studies of the collagens help to better understand various structural studies derived from the experimental data that are limited in resolution. However, any new findings via MD simulations need to be supported experimentally.

2.5.6. NMR Studies of Collagen Structure and Dynamics

This section focuses on discussing studies that have been carried on ^{13}C, ^{15}N and ^{2}H NMR of the native collagens. The studies of the collagen proteins have relied on the developments in the NMR field, therefore the section will be presented more or less in chronological order.

The first structural and dynamical studies of the collagen proteins started in early 70s by Torchia et al. The first NMR spectrum of collagen was recorded on ^{13}C nuclei of calf Achilles tendon in 1973.[120] Further analysis on solubilised native collagen were performed on α1-CB2 triple-helical fragment extracted from rat skin that was 36 residues long.[121] These studies showed that the Hyp ring is less mobile compared to the Pro ring on the nanosecond timescale, as shown by effective correlation time (τ_{eff}). Initially these experiments were

performed at natural abundance using static solid-state ^{13}C NMR, concluding that the large number of signals obtained could not be characterised conclusively due to a large distribution in amino acid signals in the spectra. Therefore, selective labelling via either feeding or injection of isotopically labelled amino acids was performed on chick embryos. The first ^{13}C enriched NMR spectra of reconstituted chick calvaria revealed that the Cα linewidth of the glycine signal is motionally averaged at 20 °C showing half the linewidth at -95 °C. Furthermore, analysis of the relaxation of the ^{13}C signals revealed that the collagen backbone motions are anisotropic with an amplitude of about 30° and correlation times estimated to be in the range from 10^{-6} to 10^{-8} s.[122]

A subsequent more detailed analysis of backbone motions was performed using ^{13}C enriched reconstituted collagen fibrils. These studies showed that in the slow motional regime (10^{-4} s) the collagen backbone on the Gly ^{13}C carbonyl moves through an amplitude of 41°, where analysis assumed a reorientation about the long axis of the collagen molecule (the angle is the angular separation between two orientations that the nuclei studied is hopping between). Further studies of ^{13}C enriched Cα Gly revealed the presence of fast motions in the 1 – 5 ns range where the amplitude of motion was calculated to be 10°, which corresponds to segmental motion of the backbone in line with the molecule long axis. Additionally, in these studies the amplitude of motion was also investigated for mineralised and cross-linked collagens which were determined to be 9° and 5.5°, respectively at room temperature. These studies showed that the cross-linking does not affect the fast backbone dynamics of the collagen molecule.[122,123] The collagen side-chain molecular motions were studied via ^{13}C enrichment of Ala, Met, Lys and Glu amino acids.[124] The solid-state NMR studies of the line-shapes of the ^{13}C enriched amino acids showed chemical shift tensor (CSA) values (Δ_σ) between 29 and 34 ppm representing a fast reorientation of the molecular segments containing these residues. These latter studies were carried out for reconstituted collagen fibrils not in intact fibrillar collagens.

Further work in the Torchia group analysed cis and trans Gly-Pro and Hyp peptide bonds ratios in the ^{13}C-labelled chick calvaria collagen. It was found through selective labelling that in the solution-state, 16% of the Gly-Pro and 8% of the Hyp bonds in collagen are in the cis conformation. This result further confirmed that cis-trans isomerisation is the rate determining step in the propagation of folding in the collagen triple-helix.[125,126]

Besides ^{13}C labelling and NMR experiments, Torchia and co-workers' utilised ^{2}H-labelled amino acids and analysis of the ^{2}H line-shapes to yield information on the dynamics of side-chain motions via alanine-3,3,3-d_3 and leucine-d_{10} labelling. Only at lower temperatures

(-18 °C) were the characteristic ^{2}H Pake doublets detected. Application of a two-site jump model of Ala side-chains showed fast reorientations of the Cα-Cβ bond vector over an angle of 30° with a correlation time of about 10^{-7} s, measurements were carried out in temperature range from -18 to +18 °C on reconstituted collagen fibrils in buffer of 0.02 M $NaHPO_4$.[127,128] Additionally, ^{2}H-labelled proline was studied in collagen. The analysis of the line-shapes showed that the proline residues in collagen undergo ring puckering motions with root-mean-square amplitudes in range from 11° to 30°, in temperature range of -20 to +20 °C.[129]

All the previous studies mentioned were performed using static solid-state NMR, where most of the samples were in native state or in a salt buffer solution (for solution-state NMR). The first solid state magic angle (MAS) cross polarisation (CP) NMR spectra of natural abundance sample were recorded in 1978 by Schaefer et al.[130] The MAS frequency used was 3 kHz, the samples were comprised of lyophilised skin and tendon that further were compared with hydrated samples. The results showed that hydrated tendon gave rise to a CP NMR spectrum that was similar to that for lyophilised tendon, indicating that the tendon retains its solid-state like properties. However, skin collagen upon hydration became denatured. Besides tendon and skin samples, the authors also showed ^{13}C CP MAS spectra of native ivory that showed a typical signature NMR spectrum of native collagen. MAS enabled a dramatic improvement in resolution of the NMR spectra of collagens as well as opened a possibility of using more advanced NMR experiments and perform more detailed structural studies.[6,131,132]

The hydration studies of the bovine Achilles tendon showed that the ^{13}C CP MAS some of the resonance lines narrowed with increasing hydration level with no chemical shift changes for any of the residues indicating that there is no reorganisation during the rehydration of the collagen material.[133,134] Further studies by Huster et al investigated collagen molecular dynamics via ^{1}H-^{13}C dipolar couplings (using DIPSHIFT) and carbonyl CSA (using CODEX).[135] Their results showed that the hydroxyproline ring flipping is strongly dependent on water content, and hence it is slower than proline rind flipping due to the hydrogen bonding of the Hyp hydroxyl group. The collagen backbone and molecular helical geometry is strongly coupled to the prolyl ring conformation, ring-flipping is constantly twisting the collagen triple-helix from a less-tightly wound to a more tightly wound structure (see Section 2.5.4.1.). This work has primarily focused on monitoring the proline Cγ chemical shifts because the Cγ position is the one that moves through the greatest amplitude during ring flipping. The observed ^{13}C chemical shifts observed are the equilibrium population–weighted average of those for the respective endo and exo conformations.[92,93]

Solid-state NMR spectroscopy has also been performed on fibrinogenesis products of collagen type-I. In this study, ^{13}C chemical shift changes were used to imply how collagen molecular conformation changes as a result of fibril formation.[136,137] It was found that the there is a specific set of both backbone and side-chain conformational changes associated with fibril formation. By assessing the relative populations of the possible side-chain conformational states for different amino acid types, it was concluded that the drive for fibril formation may be the structural reorganisation of 10−15% of the collagen amino acids. Lysine, arginine, leucine, proline, and hydroxyproline residues were identified to be the most likely to undergo structural reorganisation as a result of fibril formation.

Recent collagen studies by NMR spectroscopy have focused on ^{15}N nuclei. A study on the backbone dynamics of the collagen protein in its native environment in bone was achieved by using paramagnetic doping of copper-ethylenediaminetetraacetic acid (Cu (II) EDTA) inside the bone matrix to enhance the sensitivity in the resulting NMR spectra. The paramagnetic relaxation enhancement effect is not homogenously applied to each of the glycine and proline/hydroxyproline sites due to lack of efficient mixing of sample and paramagnetic tag. Therefore, only the relative changes in relaxation as the function of the hydration level were studied. The results showed that proline/hydroxyproline mobility is highly dependent on its association with water.[138]

The latest advances in the NMR field are coming from DNP NMR studies, where paramagnetic radicals are used as an electron source to transfer magnetisation to surrounding proton nuclei, and further by CP to ^{13}C and ^{15}N nuclei to give a significant enhancement in the detected signal intensity. This methodology has been used for natural abundance bone material mixed with the AMUPOL radical and gave an approximately 30-fold enhancement in signal intensity.[139] Additionally, in the same work, direct evidence of the imino and aromatic amino acid interactions were observed. This has only been observed before in synthetic model peptides via X-ray diffraction studies.[140] Furthermore, DNP NMR in combination with [U-^{13}C]-Lys isotopic enrichment of mouse skin and in vivo ECM collagen has enabled observation of post-translational modifications of the collagen Lys (Hyl and its glycosylations) that has never been detected before using NMR spectroscopy.[59]

There is a variety of information available by using either solid or solution state NMR. However, a full complete understanding of the collagen molecule and fibrillar dynamics are still unknown. Structurally collagen is a simple protein, but dynamically it presents a highly complex system. Interest in recent years has increased, due to a greater realisation in the importance of the collagen protein in tissue function. Isotopic enrichment is required to gain better resolution and

insights via NMR spectroscopy. However, isotopic enrichment in in vivo studies can be rather expensive and time consuming. Therefore, our group has developed an in vitro cell based model where cells produce collagen rich ECM and varied isotopic enrichment can be introduced. This model has been validated by a one-off in vivo experiment.[6] In this thesis, an in vitro cell based model was used to produce ECM, where ^{13}C and ^{15}N nuclei were isotopically enriched in various amino acids followed by extraction of collagen fibrils. The effects of the glycation chemistry on the collagen structure and dynamics are also investigated.

3

Solid-state Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical technique used for determining the molecular structure and studying the dynamics of many chemical compounds. In our studies, the biological samples of interest are inherently solid, necessitating the use of solid-state NMR spectroscopy. This chapter introduces some essential theoretical foundations of solid-state NMR spectroscopy in order to help understand the experiments used in this work.[141–146]

3.1. Nuclear Spin Interactions in the Solid State

3.1.1. Basic Concepts

Atomic nuclei are comprised of protons and neutrons. In case when the number of protons and neutrons are both even, the nuclei are NMR inactive. On the other hand, when nuclei have an odd number of neutrons and/or protons, the nuclei are said to be NMR active. Each NMR active nuclei possesses a magnetic moment:

$$\hat{\mu} = \gamma \hat{\mathrm{I}} \tag{3.1}$$

here γ is the gyromagnetic (or magnetogyric) ratio which is a constant for a given nucleus ($T^{-1} s^{-1}$) and $\hat{\mathrm{I}}$ is a spin angular momentum. $\hat{\mathrm{I}}$ is a vector quantity; its direction and magnitude are quantised, and its units are $\hbar$ (Planks constant). The spin angular momentum is characterised by its spin quantum number I which is a constant for a given nucleus. For example, $I = \frac{1}{2}$ for ^{1}H, ^{13}C, ^{15}N, and $I = 1$ for ^{2}H and ^{14}N; nuclei with $I = 0$ are spin inactive, e.g. ^{12}C and ^{16}O, and cannot be used for NMR spectroscopy.

When a sample is placed in a static magnetic field, B_0, the spins align themselves in such a way that there is a small net magnetisation, M, generated. This net magnetisation precesses around the magnetic field at Larmor frequency:

$$\omega_0 = -\gamma B_0 \tag{3.2}$$

This net magnetisation can be manipulated by using radio-frequency (RF) pulses. The most important part of the NMR experiments is the nuclear spin interaction Hamiltonian, which describes the energy of interactions of the nuclear spins with their environment (external interactions) and with each other (internal interactions within the sample). The overall Hamiltonian describing the energy of a nuclear spin can be written as:

$$\widehat{\mathcal{H}} = \widehat{\mathcal{H}_Z} + \widehat{\mathcal{H}_{RF}} + \widehat{\mathcal{H}_{DD}} + \widehat{\mathcal{H}_{CS}} + \widehat{\mathcal{H}_Q} + \widehat{\mathcal{H}_J} \tag{3.3}$$

$\widehat{\mathcal{H}_Z}$ represents the Zeeman interaction, the interaction between the nuclear spin and the applied magnetic field of the NMR experiment; $\widehat{\mathcal{H}_{RF}}$ describes effects of RF pulses applied in the NMR experiment on the nuclear spin; $\widehat{\mathcal{H}_{DD}}$ describing dipolar coupling between the nucleus and any surrounding spins; $\widehat{\mathcal{H}_{CS}}$ describes the shielding interaction, the interaction of

the electron density surrounding the nucleus (which also interacts with the applied magnetic field) with the nucleus; $\widehat{\mathcal{H}_Q}$ is the electric quadrupole coupling interaction, which is only relevant for spins with I > 0, and $\widehat{\mathcal{H}_J}$ is the scalar or J-coupling, an electron-mediated through-bond coupling between nuclear spin interaction. The dipolar, shielding, quadrupole and scalar coupling are all anisotropic interactions in general, i.e. their magnitude depends on the orientation of the molecule containing the nuclear spin, with respect to the applied magnetic field. The work outlined in this thesis is mainly concerned with the shielding and dipolar interactions. These interactions are described in further detail in the following sections.

3.1.2. Chemical Shielding

In NMR spectroscopy, the main interaction is between the observed nuclei and the magnetic field. The distribution of electrons around the nucleus is not spherical, it is anisotropic, and hence different orientations of the electron distribution will interact with the applied magnetic field (B_0) differently. The electrons are said to be shielding a nucleus from the magnetic field or exerting a local magnetic field on the nucleus. The shielding effect from core electrons leads to an observable chemical shift in the NMR spectrum. This means that different nuclei which have different chemical environments will have different chemical shifts, for example, glycine molecule has two different carbon atoms, and as a result, they will be separated in the ^{13}C NMR spectra (Figure 3.1.).

Molecular tumbling in the solution-state averages out the anisotropy of the electron distributions. This allows the molecule to be treated as being spherically symmetric, giving rise to one chemical shift in the NMR spectrum (Figure 3.1. (a)). However, the lack of tumbling in the solid-state affords a line-shape that is obtained as a superimposition of all molecular orientations with respect to the magnetic field. This superposition gives a very broad peak in the NMR spectrum with a characteristic powder pattern (Figure 3.1. (b)). This orientation dependence is called chemical shift anisotropy (CSA). Each differently oriented nucleus will experience a slightly different degree of chemical shielding and will exhibit a difference in chemical shift. As a result, the NMR spectrum will display a broad line that represent all possible orientations of all the nuclei within the sample studied. The shape of the line will further depend on the symmetry of the nuclei.

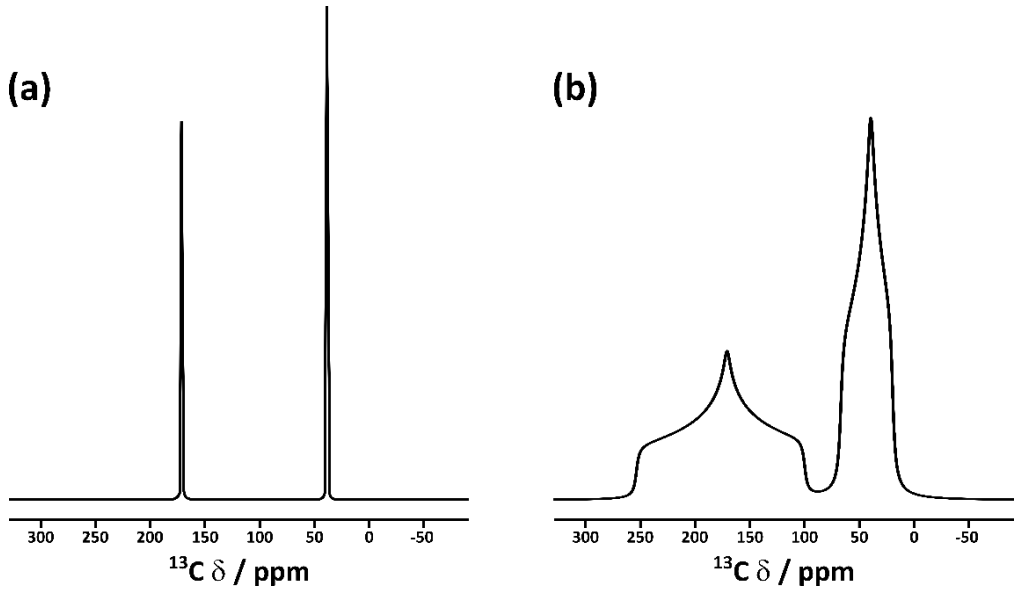

Figure 3.1. Simulated ^{13}C NMR spectra of α-glycine using SIMPSON simulation package.[147] (a) A solution-state NMR spectrum showing isotropic chemical shifts of carbonyl C=O and Cα signals at 178 and 42 ppm, respectively. (b) The powder line-shape of the carbonyl C=O and Cα nuclei.

3.1.3. Dipolar Interaction

As it was mentioned before, any nuclear spin possesses a magnetic moment (Equation 3.1), that generates a small, local magnetic field. If two spins are close to one another, they can interact directly. This through-space interaction is known as dipole-dipole or dipolar coupling. This interaction will depend on the distance between the two nuclei, the orientation of the intermolecular vector with respect to B_0 and the nuclei type. The dipolar coupling between two spins I and S can be written as:

$$d = \hbar \left(\frac{\mu_0}{4\pi}\right) \frac{1}{r^3} \gamma_I \gamma_S \frac{3cos^2\theta - 1}{2} \tag{3.4}$$

where μ_0 is the vacuum permeability, γ_I and γ_S are gyromagnetic ratios of spins I and S, respectively, r is a distance between spins I and S, and θ is an angle between the internuclear vector and the applied magnetic field, B_o. The dipolar coupling between spins can be homo- and heteronuclear, where spins I and S are the same species or different, respectively.

The dipolar coupling is a function of an internuclear separation, therefore it can be employed as a useful probe for measuring the internuclear distances in a system. This approach is widely adopted in the studies of different materials, including organic, inorganic and biological structures. Furthermore, this spatial separation and interaction of the spins is exploited by multiple NMR experiments used in this thesis.

3.1.4. Quadrupolar Interaction

Nuclear spins with $I > 1/2$ are referred to as quadrupolar nuclei. These spins possess a nuclear electric quadrupolar moment, that arises from an asymmetric distribution of a charge in the nucleus.[141,148] Quadrupolar nuclei will interact not only with external magnetic field of the NMR experiment and all local magnetic fields, but also with any electric field gradient (EFG) present at the nucleus. The electron density surrounding a quadrupolar nucleus typically provides an electric field gradient unless the nucleus is at a site of cubic symmetry.

The work carried out in this thesis is predominantly centred around interactions concerning dipolar and shielding interactions as discussed previously. Experiments involving quadrupolar nuclei were used in this work for performing relaxation experiments on a reference sample and will be discussed in more detail in the Section 4.2.2.

3.2. Experimental Techniques

The following sections outline the methodology used in the solid-state NMR experiments detailed in this thesis. Figures and explanations of the NMR pulse sequences that are used in this work are provided. In the figures, pulses are represented as vertical black rectangles; decoupling is indicated as horizontal grey rectangles and free induction decay as a decaying wave (vide infra).

3.2.1. Magic-Angle Spinning (MAS)

MAS is used to remove the effects of line-broadening arising from anisotropic nuclear spin interactions, e.g. CSA and dipolar coupling. In the MAS experiment, the sample is rotated at an angle of 54.74° (θ_R) with respect to an applied magnetic field, B_0 (Figure 3.2.). CSA and dipolar coupling contain an angular dependence of the form:

$$3cos^2\theta - 1 \tag{3.5}$$

This technique helps to increase resolution and hence simplifies analysis. To achieve optimal sensitivity and resolution, a spinning rate ω_r should be at least a factor of three greater than the largest anisotropy present.[149] However, the homonuclear coupling of the ^{1}H nuclei is strong and thus some degree of line broadening remains even at high spinning rates. High spinning rates are precluded for the biological samples studied here as it can leads to sample

dehydration and alterations to the structure of the material understudy. Therefore, the ^{1}H NMR spectra from the biological samples in this work typically exhibit broad signals. For the ^{13}C and ^{15}N spectroscopy used in this work, the MAS speeds of less than 11 kHz are suitable to achieve narrow signals.

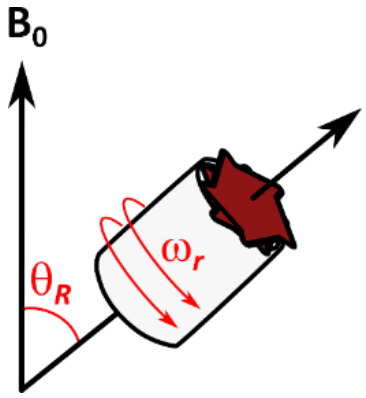

Figure 3.2. The orientation of the sample along an axis at the magic angle θ_R with respect to the magnetic field B_0 with a spinning rate of ω_r. The sample is placed in a small cylindrical rotor as shown above and rotated along the axis shown.

An artefact of MAS is the presence of spinning side-bands in the solid-state NMR spectra. These bands occur at frequencies that are multiples of the MAS rate on both sides of the observed signal (Figure 3.3.). The intensity of these spinning side-bands depend upon the spinning rate and the magnitude of the nuclear spin interaction being averaged by the MAS. At lower MAS rates, the spinning side-bands are observed with higher intensities. The anisotropy and asymmetry of the nuclear interaction being averaged can be extracted from the spinning sideband intensities.

The MAS rate is an important experimental parameter. If the MAS rate matches the frequency of the separation of the signals that correspond to a pair of dipolar coupled spins, then the spinning will not average out this dipolar interaction. As a result, the spectrum will show line broadening and splitting of these two signals. This is known as rotational resonance. The MAS rates used in this work were <11 kHz in all NMR experiments, where rotational resonance for the Cα and carbonyl are between 11 – 14 kHz for an alanine sample in a 9.7 T magnet.

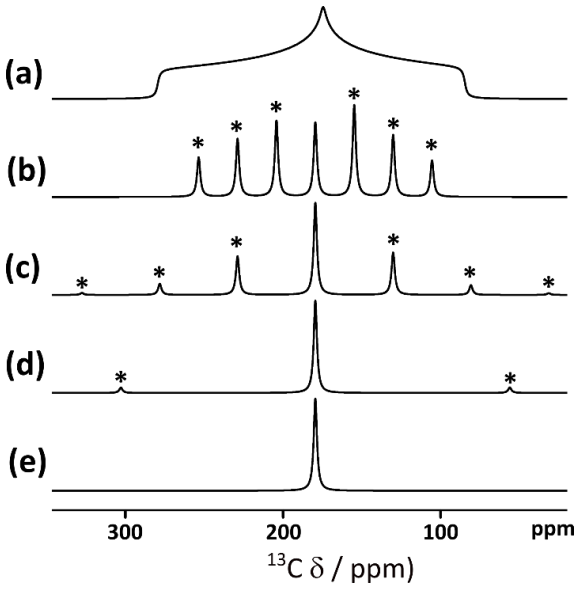

Figure 3.3. Solid-state ^{13}C 1D NMR spectra of powder of carbon atom simulated with δ_{iso} = 178 ppm, η_{cs} = 0.9 and Δ_{CS} = 100 ppm at (a) MAS frequency of 0 Hz, static, (b) MAS rate of 1 kHz, (c) MAS rate of 2 kHz, (d) MAS rate of 5 kHz and (e) MAS of 10 kHz. Spinning side-bands are highlighted with an asterisk. The spinning side-bands are separated in multiples of the MAS rate. Simulated ^{13}C NMR spectra were obtained using SIMPSON simulation package.[147]

3.2.2. Decoupling

^{1}H nuclei have a large gyromagnetic ratio and they are abundant in biological samples which results in broad lines in the ^{1}H NMR spectra due to multiple couplings. Protons also have an effect on other nuclei via heteronuclear dipolar coupling resulting in broadening of e.g. ^{13}C and ^{15}N solid-state NMR spectra that can make the assignment of the signals complicated. Furthermore, the resulting line-broadening affects the signal-to-noise ratio and decreases sensitivity of the NMR experiment.[150,151] The heteronuclear coupling between the proton and other nuclei can be removed by using decoupling. Decoupling involves irradiating the ^{1}H spins during the evolution and acquisition in order to interfere with dipolar coupling. There are many decoupling schemes used in solid-state NMR; the ones used in this work are two-phase modulation (TPPM),[152] small phase increment alternation (SPINAL)[153] and frequency-switched Lee-Goldburg (FSLG)[154] decoupling. These sequences use specific RF pulses with different powers and durations that are applied on the sample during the course of the experiment. These decoupling sequences have been reported to perform robustly on the similar type of samples employed in this work.[6]

3.2.3. Cross-polarization (CP)

The cross-polarization (CP) is one of the most routinely used NMR experiments. ^{15}N and ^{13}C nuclei exhibit long relaxation times and a low natural abundance, which together increases the experimental time required to achieve a desired signal-to-noise ratio. CP is designed to maximise the signal gain per unit time. CP experiments involve polarization transfer from abundant spins (i.e., those of ^{1}H) to ones with a lower gyromagnetic ratio, in this work ^{13}C and ^{15}N.

In the CP experiment, after performing the standard $\pi/2$ pulse on the ^{1}H spins that creates ^{1}H magnetisation in the x,y-plane, a cross-polarization pulse (contact time or spin lock pulse) is applied to both ^{1}H and ^{13}C/^{15}N spins for 1-5 ms. Finally, the signal is detected on the ^{13}C/^{15}N channel. Figure 3.4. shows a schematic representation of this pulse sequence.

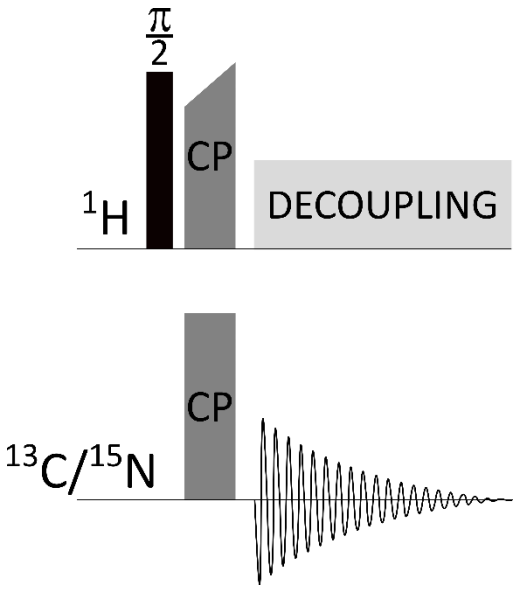

Figure 3.4. Simple CP pulse sequence representing a $\pi/2$ pulse that is applied on ^{1}H channel, further CP step is introduced where magnetisation is transferred from ^{1}H to ^{13}C or ^{15}N spins. Finally, acquisition is employed on ^{13}C/^{15}N channel while at the same time decoupling is performed on the proton channel.

To ensure an efficient magnetisation transfer, the Hartman-Hahn condition between ^{1}H and ^{13}C/^{15}N must be fulfilled: [155]

$$\omega_I = \omega_S \tag{3.6}$$

$$\gamma_I B_I = \gamma_S B_S \tag{3.7}$$

Where $\omega_{I/S}$ are the nutation frequencies of the spin-locking fields applied to the ^{1}H and $^{13}C/^{15}N$ nuclei. The signal sensitivity of the $^{13}C/^{15}N$ nuclei can be increased by a factor of $\left|\gamma_H/\gamma_{C/N}\right|$ via CP experiment. For ^{1}H-^{13}C CP the maximum enhancement is ca. factor of 4 and for ^{1}H-^{15}N it is ca. 10.

The dipolar interactions between ^{1}H and $^{13}C/^{15}N$ mediate the cross-polarization transfer between the spins, so anything that modifies the interaction will affect the CP efficiency. The initial rate of the cross-polarization is affected by the strength of the dipolar coupling, and thus magnetization transfer is slower for the spins that are further away from each other. The magnetization builds up to a maximum, after which relaxation of the accumulated $^{13}C/^{15}N$ transverse magnetization takes over and the CP signal intensity begins to decay. The optimal contact time for each sample will be slightly different as the transfer of the magnetization between different sites will vary, depending on the strength of the ^{1}H – heteronuclear dipolar coupling. This leads to the non-quantitative spectral intensities recorded by the cross-polarization experiment. The optimal duration of the spin locking time (CP step) during the magnetization transfer giving the maximum in the ^{13}C or ^{15}N signal also depends on the ^{1}H spin $T_{1\rho}$ time constant. $T_{1\rho}$ is the time constant for the relaxation of the ^{1}H spins during the spin-lock period and is influenced by the correlation time and the amplitude of molecular motion affecting the ^{1}H. The time scale of the motions determining $T_{1\rho}$ is governed by the nutation frequency of the locking pulse rather the Larmor frequency. For very short $T_{1\rho}$ the decay of the cross-polarization signal occurs very early and inhibiting a further build-up of the magnetization.

In the CP experiment, the only relaxation that is required between scans is that of the ^{1}H nuclei, hence the recycle delay of the experiment is dependent upon the spin-lattice relaxation of ^{1}H nuclei. ^{1}H relaxation is typically fast (T_1 < ~2 s) for the biological samples studied here and thus, more scans can be acquired within a fixed amount of time, leading to improved signal-to-noise ratios.

3.2.4. Two-Dimensional NMR Spectroscopy

The two-dimensional (2D) correlation experiments can be carried out to either indicate correlations between nuclear spins through dipolar or scalar interactions to resolve overlapping signals, or to relate two spin interactions to one another (Figure 3.5. (b)).

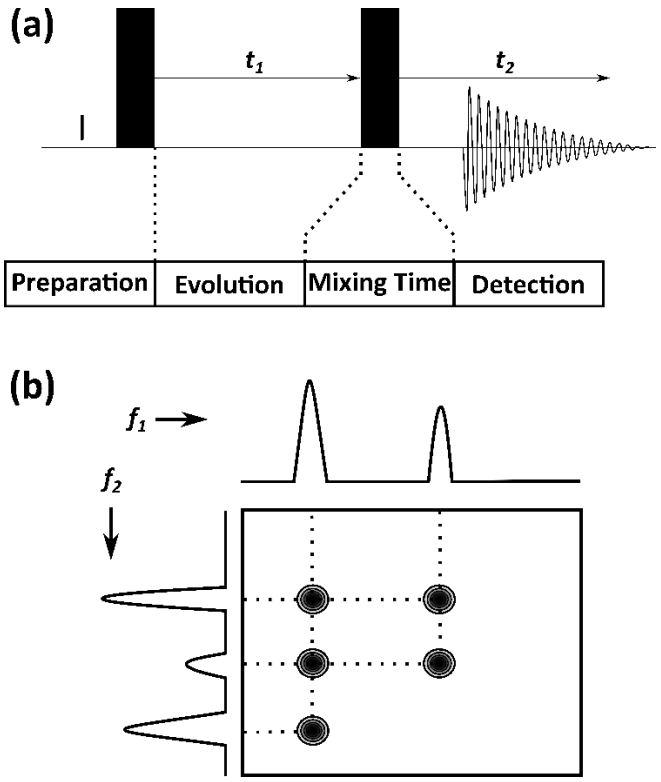

Figure 3.5. (a) General pulse sequence for a 2D NMR experiment showing four components: preparation pulses, increment delay (t_1), followed by mixing or reconversion pulse(s), finished with acquisition t_2. (b) Performing the Fourier transform on both dimensions yields a 2D spectra with f_1 and f_2.

In general, 2D experiment pulse sequence consists of four parts: preparation, evolution, mixing time and detection (Figure 3.5. (a)). After preparation of a pulse (or pulses), the system is allowed to evolve during t_1. Then mixing or reconversion of the pulse (or pulses) is allowed, and finally the free induction decay (FID) is recorded during acquisition. 2D experiments are often carried out by repeating 1D experiment with an increment delay (a short fixed-time interval repeatedly added to t_1). The collected 1D slices can be Fourier transformed (FT) to obtain both dimensions f_1 and f_2.

3.2.5. Proton Driven Spin Diffusion (PDSD)

The proton driven spin diffusion (PDSD) experiments used in this work correlates ^{13}C spins via their mutual dipolar coupling to intervening ^{1}H nuclei.[156] The resulting 2D ^{13}C-^{13}C correlation spectra thus show correlations between ^{13}C signals from ^{13}C up to 10 Å apart in distance in the sample, which is several times longer than a ^{13}C-^{13}C bond for instance. Thus, this experiment is employed in this work to gain structural information on the ^{13}C-Lys enriched samples (Section 8.4.) that were treated with paramagnetic complexes. Figure 3.6. (a) shows the PDSD experiment pulse sequence.

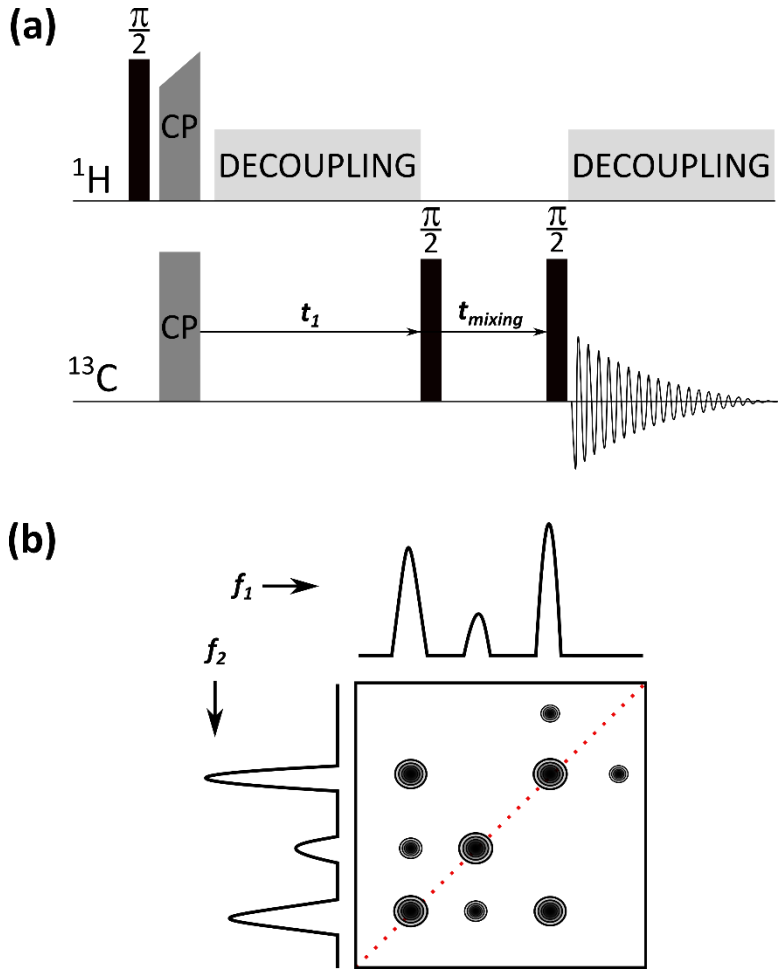

Figure 3.6. (a) The pulse sequence of the PDSD experiment. (b) ^{13}C-^{13}C 2D correlation spectra, where the main diagonal (red dots) representing all signals in the NMR spectra, the off-diagonal cross-peaks represent the dipolar coupling interaction. The 2D spectrum is symmetrical about the PDSD diagonal.

The pulse sequence starts with a CP step, followed by an increment delay to enable the evolution of the ^{13}C magnetisation. Following this, the ^{13}C transverse magnetisation is returned to the zero-quantum coherence by a $\pi/2$ pulse on the ^{13}C channel followed by mixing time. During this time, the decoupling is removed to enable proton diffusion between ^{1}H-^{13}C and ^{1}H-^{1}H nuclei network. Then another $\pi/2$ pulse is applied on the ^{13}C channel that enables ^{13}C detection. During this process, the decoupling on ^{1}H is re-introduced. During the mixing time, the magnetisation of ^{13}C 'diffuses' through the proton network via dipolar coupling to another ^{13}C nuclei. The length of the mixing time determines the range of the internuclear distances that can be explored. In general, PDSD NMR can be used for studying three-dimensional structures and probe long range spatial correlations. In Figure 3.6. (b) a simplified 2D NMR spectra is shown. The signals on the main diagonal represent self-correlated signals. Signals on the off-diagonal are ^{13}C-^{13}C pairs that are close in space. The mixing time employed for the experiments detailed in this thesis is 20 ms which enables to observe correlations between ^{13}C-^{13}C nuclei that are 5 – 6 Å apart.

3.2.6. Permutationally-offset Stabilised C7 (POST-C7) Experiment

POST-C7 is an experiment designed to correlate two nuclear sites that are directly dipolar coupled to one another. This experiment helps to map networks of ^{13}C-^{13}C bonds. In this work, the POST-C7 experiment was used to help resolve ^{13}C NMR signals in differently labelled collagen samples (Sections 7.2. and 7.3.). The experiment uses the symmetry-based recoupling sequences under MAS, which means that the rotor synchronisation is required for this sequence. The sequence is represented in the Figure 3.7. (a).[157]

The POST-C7 pulse sequence used in this thesis starts with a CP step by transferring magnetisation from surrounding protons to the ^{13}C nuclei. Following this, composite pulses are applied on the ^{13}C channel; the A block is designed to excite the DQ coherences. This pulse train is constructed of 7 C_{φ} elements, where each element contains a composition of pulses in different phases. Each element has a duration of $2T_r/7$, where T_r is a length of rotor period. After block A, the DQ is evolved for time t_1 for 2D experiment, followed by block B that contains reconversion pulses containing additional phase shift of $\varphi = \{\frac{\pi}{2}; \pi; \frac{3\pi}{2}; 0\}$ that returns the DQ coherence back to the single-quantum (SQ) which enables detection of the signal. During the evolution, the decoupling is removed to allow spin diffusion.

This experiment can be used as a simple 1D experiment, where the DQ filter will only be useful in observing species that are significantly dipolar coupled. Thus, this type of experiment is beneficial in ^{13}C enriched materials, where only signals from the enriched species will be observed, while the other carbon sites that are unlabelled will not be observed (see Figure 3.8.). In 2D experiments, the SQ dimension appears in f_1 and DQ dimension on f_2. The SQ dimension shows direct ^{13}C chemical shifts, where DQ contains a sum of signals that are correlated to one another due to close proximity to one another (or directly bonded). Figure 3.7. (b) shows a simple example of the assignment of the chemical shift and the correlations. This 2D experiment shows that it can be used to trace connectivity of the atoms, that can help with the structural assignment of the material.

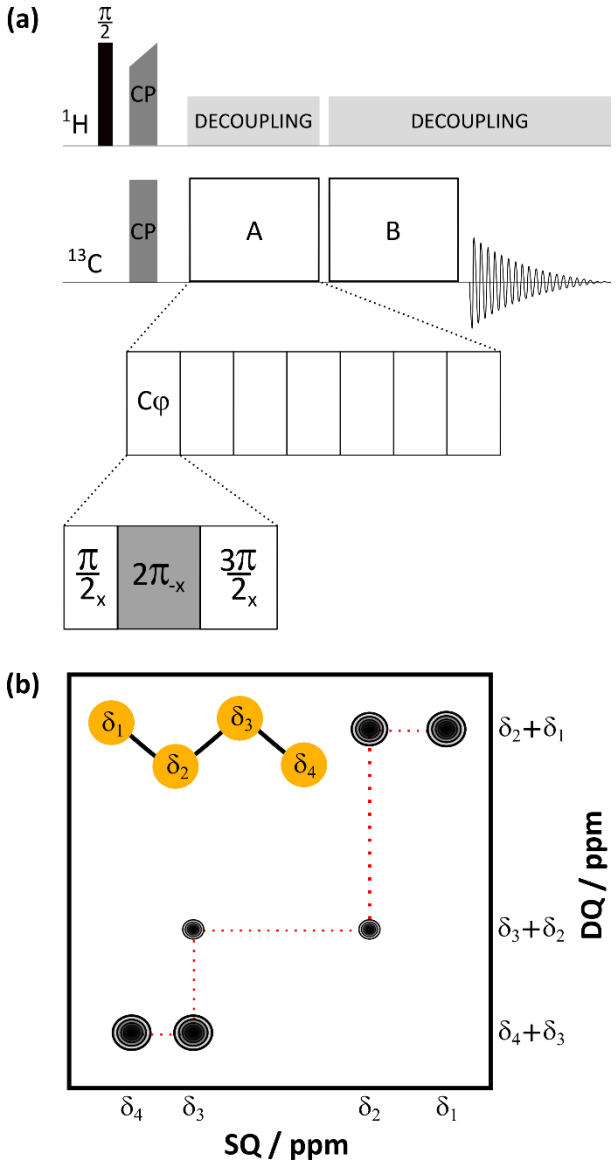

Figure 3.7. (a) The POST-C7 pulse sequence of 1D experiment is shown. The 2D correlation experiment uses the same sequence with the evolution time t_1 between blocks A and B. Blocks A and B represent the excitation and the reconversion sections, respectively. Each block contains 7 elements, where each element is a set of composite pulses that contain a phase modulation that is averaged to two 2π pulses. Rotor synchronisation is required to fit the one block into 2n rotor cycles. (b) A simplified example of a 2D SQ-DQ POST-C7 NMR spectrum of molecule as drawn. The isotropic chemical shifts of each atom are δ_1, δ_2, δ_3, and δ_4. The SQ dimension shows the isotropic chemical shifts, DQ dimension shows sum of the chemical shifts that are strongly coupled to one another.

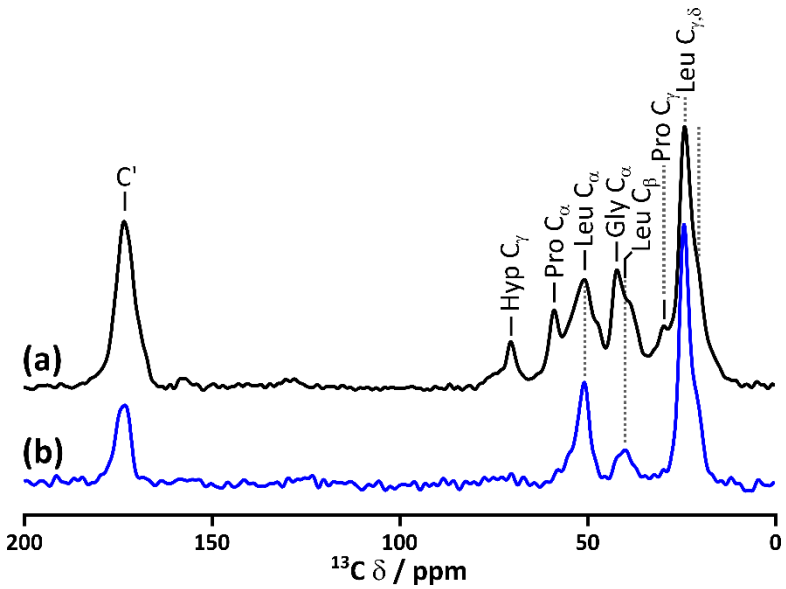

Figure 3.8. NMR experiments of (a) ^{1}H-^{13}C CP experiment and (b) ^{1}H-^{13}C POST-C7 experiment of U-^{13}C-Leucine enriched FSOB extracellular matrix that is cleaned with α-chymotrypsin digestion. CP spectra shows also unenriched ^{13}C amino acids that are at high concentrations in this material, including glycine, proline and hydroxyproline. POST-C7 spectra shows only the ^{13}C enriched species, that is leucine.

3.2.7. Heteronuclear Correlation (HETCOR) Experiment

The HETCOR experiment is a 2D heteronuclear correlation experiment, where cross-peaks represent correlation between spins ^{1}H and ^{13}C. This experiment was used in the samples where paramagnetic doping was introduced to help better observe a paramagnetic effect on the ^{1}H nuclei (Section 8.3.). This pulse sequence of this experiment is represented in Figure 3.9. The experiment starts with a $\pi/2$ pulse on the proton channel generating a transverse magnetisation. This is followed by a t_1 spin evolution during which the frequency switched Lee-Goldburg (FSLG) homonuclear decoupling is applied on the proton channel. During FSLG decoupling, the ^{1}H-^{1}H homonuclear dipolar coupling is supressed, which enables one to only observe ^{1}H and ^{13}C nuclei that are in close proximity. After the evolution period, the CP step is initiated followed by signal detection. This experiment yields a heteronuclear 2D NMR spectra that shows ^{13}C nuclei in the direct dimension and the protons in indirect; the proton dimension is better resolved compared to the direct detection of the protons. This experiment can provide structural information not only for different sites of the carbon nuclei but also on the protons, due to better resolution in the indirect dimension.

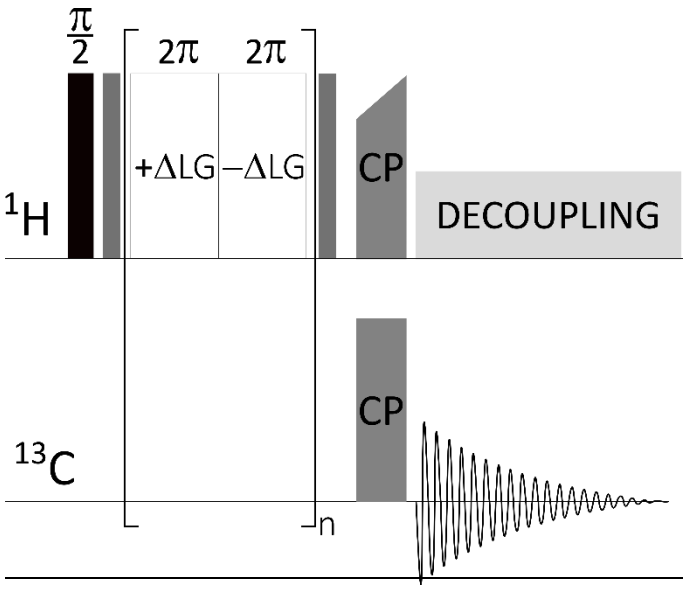

Figure 3.9. The representation of the FSLG HETCOR pulse sequence. The initial magnetisation is prepared by applying a $\pi/2$ pulse on the ^{1}H channel, followed by an additional pulse (dark grey rectangular) that places the magnetisation on the magic angle where it is maintained by a series of 2π pulses at alternating frequencies. The length of these alternating pulses determines the length of the t_1 increment. The transverse magnetisation is restored with another pulse followed by a CP step and signal detection.

3.2.8. Quadrupolar Experiments – Solid Echo

In all NMR experiments there exists a short delay between the final pulse and the acquisition of the signal which is known as dead-time. This is introduced to remove so-called 'ringing' in a coil that is a result of a residual voltage lasting at the end of the final pulse. Otherwise, the signal detected will contain artefacts that can obscure the signal interpretation. However, this can cause a problem with samples that have short T_2, which leads to a short length of the FID. In such cases, a substantial part of the beginning of the FID is lost in the dead-time. To overcome this issue, echo experiments are implemented, where dephasing of ^{2}H transverse magnetisation is recovered by introducing a $\pi/2$ refocusing pulse at a time τ after the initial $\pi/2$ pulse as shown in Figure 3.10. The second τ delay was set to be shorter than the first. This yielded an FID that did not begin at the intensity maximum of the echo but just before. This allows the data to be processed (by removing initial points) so that the processed FID begins at the exact echo maximum.

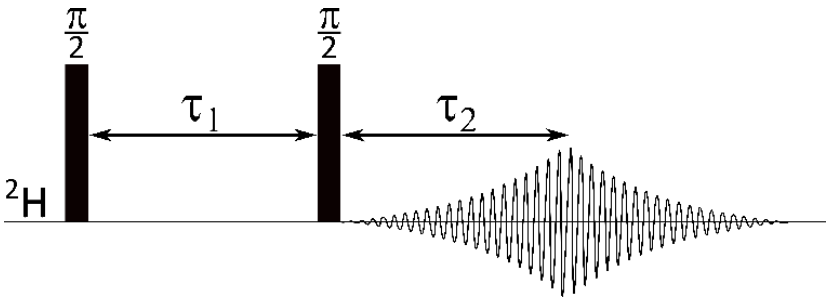

Figure 3.10. The representation of the echo experiment's pulse sequence. For ^{2}H nuclei, a solid echo was used, where both pulses are $\pi/2$ pulse, and where $\tau_1 > \tau_2$,with τ_2 being slightly shorter.

3.2.9. Dynamic Nuclear Polarization (DNP) Experiments

This type of experiment is used to gain enormous signal enhancement via the polarization transfer from unpaired electrons (which have a high gyromagnetic ratio) to the nuclei of interest. The theoretical signal enhancement limit is 660 for ^{1}H nuclei ($|\gamma_e/\gamma_H|$), and even higher for other nuclei. This technique offers fast signal detection and good signal-to-noise ratio. DNP was used in this thesis to help detect ^{15}N nuclei of the samples presented in Section 6.2. Due to the low abundance of Phe and Leu amino acids in the collagen native sequence, this technique enabled the detection of ^{15}N nuclei in these samples, where using the conventional NMR spectroscopy these signals were not resolved well enough. The pulse sequence example of the DNP experiment is shown in Figure 3.11.

Three components are important to obtain the DNP enhancement: (1) unpaired electron(s) that are near the spins of interest; (2) a microwave source to initiate the polarization transfer, and (3) an experimental temperature below 200 K to maximise the efficiency of the polarization transfer. The microwave-driven DNP mechanisms are categorised into four effects: the Overhauser effect (OE), the solid-effect (SE), the cross-effect (CE) and thermal mixing (TM). In this work, polarization transfer was performed via the cross-effect mechanism. The unpaired electrons are typically introduced to biological samples by adding small amounts of polarising agents that contain stable radicals. In this work AMUPol biradical was used, as it provides a good signal enhancement on similar materials that were used in this work, additionally, this radical has been optimised for CE-DNP.[158,159] Further details regarding the sample preparation used in this work are described in Section 5.6.2. The microwave source is a gyrotron.

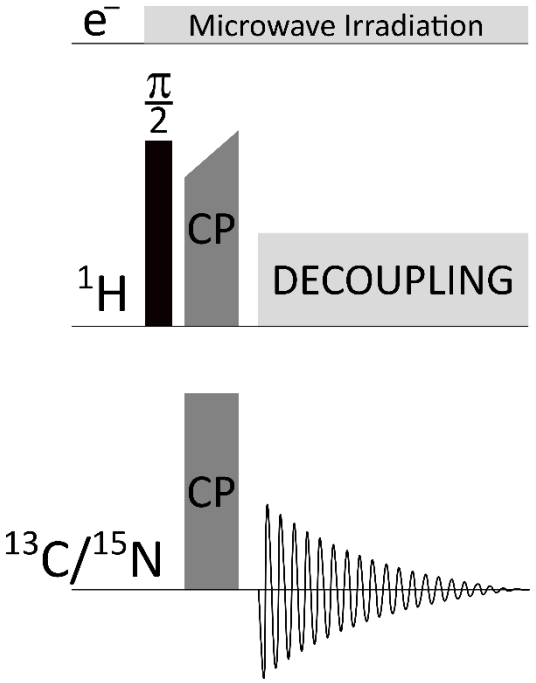

Figure 3.11. The cross-polarization pulse sequence employing the DNP enhancement. The cross-polarization sequence has not changed. The extra annotation is for the representation of the microwaves for excitation of the electrons.

4

NMR Relaxation and its Application to Complex Systems

In this thesis, we are interested in how chemical modifications to collagen affect its molecular mobility. To probe the effects of this chemistry on molecular mobility of the collagen fibrils , we use measurements of nuclear spin to assess molecular motion. As described previously, the relaxation times of nuclear spins inherently depend on the correlation time of the molecular motions within the sample that are near to the nuclei being studied. When the motion is anisotropic, i.e. functional group rotations, sidechain rocking, or protein backbone bends, the relaxation times additionally depend on the amplitude of the motion.

In organised structures like collagen fibrils, we expect there to be a distribution of relaxation times for any given residue type. For example, the relaxation time for ^{13}C spins in a given residue within the collagen sequence depends very much on the local environment of the residue. The local environment includes the local molecular packing (i.e. whether the residue is in the fibril hole or overlap zone), the nature of the neighbouring residues, and whether these residues have been affected by the applied chemical modification. We therefore required a method to assess not only the average or dominant relaxation time for a given residue ^{13}C or ^{15}N, but also to assess the range of relaxation times. Thus, we employed the inverse Laplace transform (ILT) to help separate differently relaxing components from the observed relaxation decay. Our hypothesis was that determining the range of relaxation times for a given residue type could indicate if there were several distinct environments for said residue and therefore provide useful information about the collagen fibril structure changes with collagen chemical modifications.

In the following section, the fundamentals of relaxation and NMR-based relaxation measurements will be discussed. The concepts presented here are taken from various texts available.[160–162] This will be followed by a discussion on the methodology of the ILT and its use on the relaxation data.

4.1. Introduction

Relaxation, as in all spectroscopies, describes a change in energy between two or more states, with the system of interest recovering back to its equilibrium state after perturbation. In NMR, the spin system is perturbed by a radio frequency (RF) pulse, where the net nuclear spin magnetisation, M, returns to its thermal equilibrium, aligned with the applied magnetic field of the NMR experiment, B_0. In the same way as the nuclear spins react to the external applied magnetic field, these nuclear spins also interact with the small local magnetic fields generated by each nucleus (dipole-dipole interactions) and by the surrounding electron density (shielding interaction). Therefore, the total field experienced by each nucleus will be slightly different. These small magnetic fields fluctuate with respect to time due to molecular motion moving the neighbouring spins and bonds containing the electron density. Thus, relaxation measurements in NMR can be a useful probe to study molecular motion.

Relaxation of nuclear spin magnetisation can be characterised by two processes: the longitudinal and transverse relaxation. The longitudinal relaxation describes how fast the nuclear magnetisation returns to its original value along the applied B_0 field (defined as the z-axis in the laboratory frame here). As the longitudinal relaxation changes the component of

the magnetisation parallel to B_0, the total energy of the system also changes. Thus, there must be thermal energy exchange with the surroundings, or "lattice". Longitudinal relaxation is therefore also known as spin-lattice relaxation. This relaxation process requires the return of the spin population distributed between nuclear spin Zeeman states to their Boltzmann equilibrium. Transverse relaxation describes how fast the transverse nuclear magnetisation recovers to its equilibrium value of zero. A RF pulse on the spin system creates nonzero transverse magnetisation, i.e. a non-equilibrium state. Transverse relaxation is a process of energy being transferred between spins to cause decoherence of the transverse magnetisation. Transverse relaxation therefore does not change the total energy of the system, and is also referred to as spin-spin relaxation. Both longitudinal and transverse relaxation are stimulated by local fluctuating magnetic fields; longitudinal relaxation is aided by the fields fluctuating parallel to the applied static field, B_0, and the transverse relaxation by fields perpendicular to the applied field axis.

Both relaxation processes decay exponentially with time and have characteristic time constants of T_1 and T_2 for longitudinal and transverse relaxation, respectively. These are known also as relaxation times and are a measure of the rate of the respective process. The rate of the relaxation of any magnetisation component is observed to be proportional to the deviation from the equilibrium, leading to the Bloch equations for the relaxation of both relaxation processes:

$$\frac{d}{dt}M_z = -\frac{(M_z - M_0)}{T_1}$$
$$\frac{d}{dt}M_x = -\frac{M_x}{T_2}$$
$$\frac{d}{dt}M_y = -\frac{M_y}{T_2} \qquad (4.1)$$

These equations can be solved to give the bulk magnetisation components as a function of time:

$$M_z(t) = M_0 + (M_z' - M_0)\exp(-\frac{t}{T_1})$$
$$M_x(t) = M_x(0)exp(-\frac{t}{T_2})$$
$$M_y(t) = M_y(0)exp(-\frac{t}{T_2}) \qquad (4.2)$$

where M_z' is at t=0, and $M_i(t)$is the time dependent magnetisation along the *i*-axes.

Another, significant relaxation parameter, can be measured for relaxation under the effect of the RF pulse. During the RF pulse, an effective field is generated in the rotating frame. As a result, the spins experience a much smaller field along the z-direction and the resulting relaxation behaviour is different from the normal longitudinal relaxation. This magnetisation relaxation is referred to as spin-lattice relaxation in the rotating frame or relaxation under the spin-lock field, abbreviated as $T_{1\rho}$. This relaxation describes the relaxation along the RF field of the applied pulse rather than the relaxation along the static field, B_0.

These three relaxation processes can be used to study molecular motions at different timescales covering motional regimes from femtoseconds to seconds. T_1 is sensitive towards fast molecular motions at picosecond to nanosecond timescales, where T_2 and $T_{1\rho}$ towards intermediate and slow molecular motions in the milli- to microsecond timescale (Figure 4.1.). Overlapping signals with the same chemical shift do not necessarily have the same relaxation rate, and we exploit this feature in this work to separate signals from protein residues in different environments.

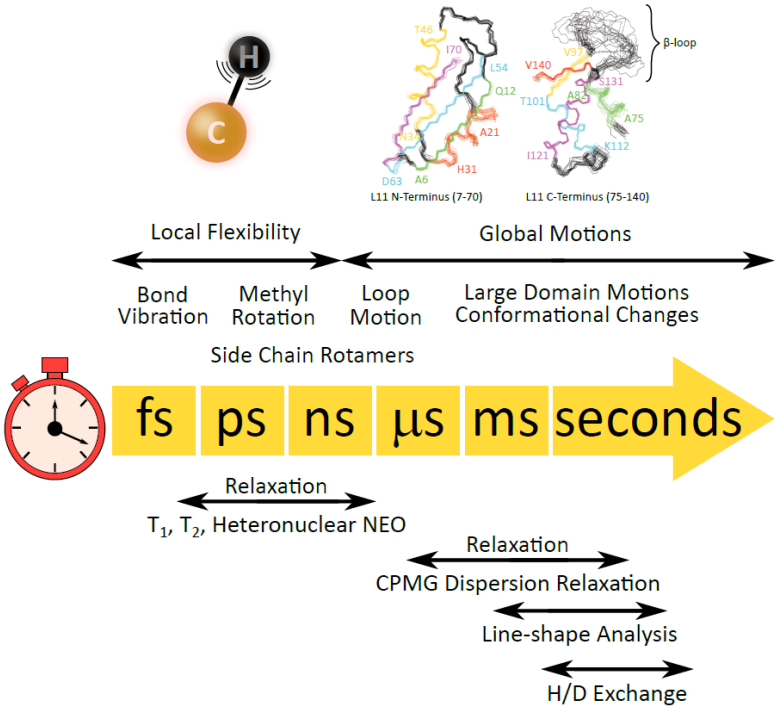

Figure 4.1. Different protein dynamics and time scales accessible to NMR are shown. Figure was adapted from reference [163].

As it was mentioned before, relaxation is caused by fluctuating magnetic fields. The time scale of these fluctuations can be characterised by the correlation time, τ_c. The correlation time corresponds to the amount of time it takes for a system to change from the initial state to

another. In other words, rapid molecular motions will give a small τ_c, while slow motions afford a large τ_c. In this work, T_1 relaxation on ^{13}C and ^{15}N is investigated. It can be shown that the T_1 is related to the correlation time by:

$$T_1 = \frac{1+\omega_0^2\tau_c^2}{\gamma^2 B_R^2 \tau_c} \tag{4.3.}$$

The observed T_1 will depend on the nucleus that is studied. Therefore the γ is the gyromagnetic ratio of the nuclei of interest, ω_0 is the Larmor frequency which will depend on the magnetic field nuclei is observed in, and B_R is the amplitude of the fluctuating random magnetic field that is present in the system. Figure 4.2. shows how T_1 varies with the correlation time according to equation (4.3).

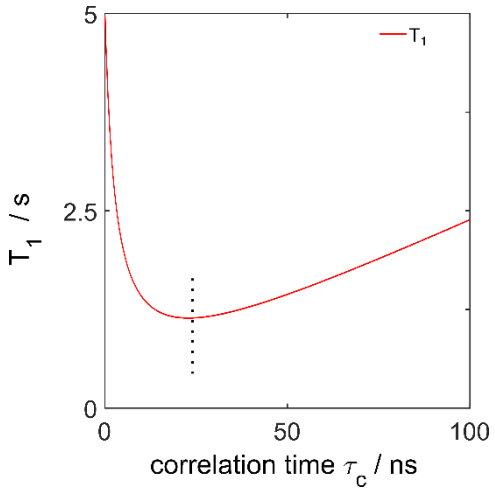

Figure 4.2. The variations of the relaxation constants T_1 with respect to the correlation time. The function was calculated for the ^1H nuclei at Larmor frequency of 400 MHz, the vertical dotted line indicates the T_1 minimum of approximately 25 ns (with $B_R^2 = 10^8$ T^2).

In NMR relaxation, the time scale is important. Relaxation of the observed spins is most effective when the time scale of the molecular motion matches the strength of the interaction frequency. In other words, protons will have the most effective relaxation (minimum of T_1) when the molecular motion is at the same frequency of the proton Larmor frequency. In the same material, multiple mobility regimes can be present at the same time, where these can be characterised by different correlation times.

4.2. T_1 Relaxation Measurements

In general, relaxation measurements are pseudo two-dimensional experiments, where the signals are recorded as a function of a delay time τ during which the relaxation occurs. The decay of the NMR signal intensity with respect to τ is analysed and the corresponding relaxation rate constant is extracted.

The inversion-recovery experiment is used for T_1 measurements in this work. The standard pulse sequence consists of $180^o_x - \tau - 90^o_{-x}$ –aquire, where the FIDs are recorded in the acquire segment with increasing τ. In this work, a variation of this sequence is used to increase a signal-to-noise ratio(Figure 4.3).[164] The first step of the pulse sequence is a CP step followed by a 90^o_i, with pulse phase of $i = +x$ or $i = -x$, generating magnetisation along the $-z$ or $+z$ direction in alternate scans, respectively, followed by the variable time delay τ. Figure 4.4. shows a representation of the net magnetisation behaviour in the inversion recovery experiment and the signal intensity change as a function of the τ delay time.

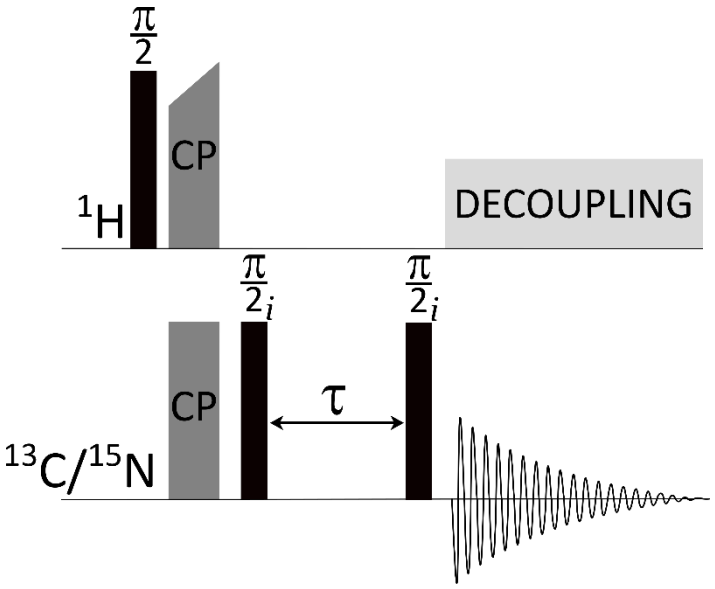

Figure 4.3. Representation of the pulse sequence for T_1 relaxation measurements. The experiment starts with a CP step, followed by a $\pi/2$ pulse on the $^{13}C/^{15}N$-channel. Then, a variable time delay τ is introduced, and another $\pi/2$ pulse is applied to put the magnetisation back in the x,y-plane for detection.

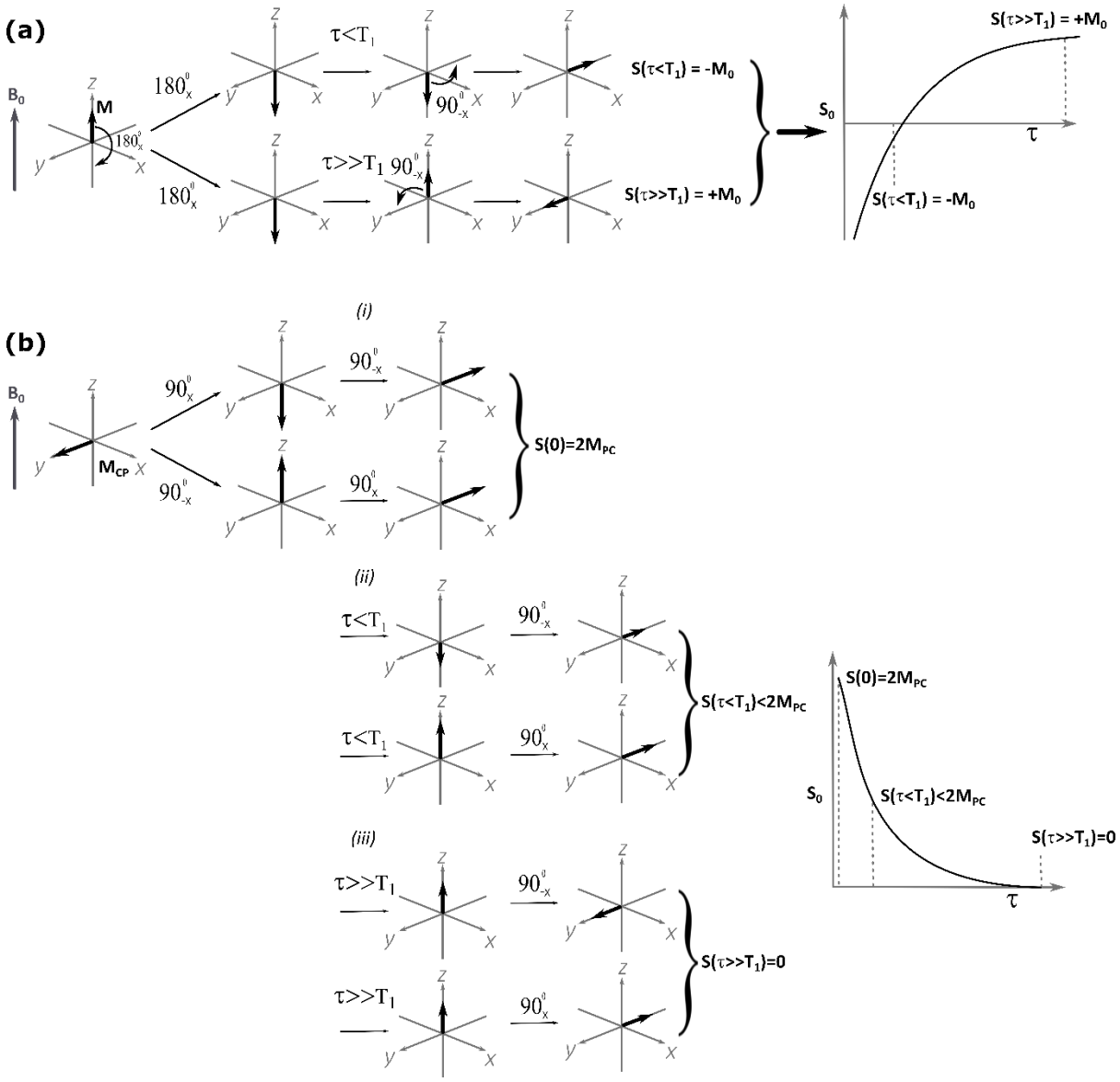

Figure 4.4. The net magnetisation behaviour during the inversion recovery experiment of the (a) $180_x^o - \tau - 90_{-x}^o$ –acquire pulse sequence and (b) of the modified pulse sequence with the incorporation of the CP step. In (a) M_0, the net nuclear spin magnetisation is along B_0 (z-axis), and after a 180_x^o pulse, the net magnetisation is along -z. The -z magnetisation decays during the τ delays, and the remaining signal at the end of τ, converted to observable transverse magnetisation with a 90° pulse. At very short τ delays, the observed signal intensity is large, as τ delay increases the signal intensity is decreasing, as shown by the change in size of the net magnetisation vector. In (b) after cross-polarization the net magnetisation is in the x,y-plane (transverse plane). Then a 90_i^o pulse is applied with pulse phase of $i = +x$ or $i = -x$, generating magnetisation along the $-z$ or $+z$ direction, allowing net z-magnetisation to decay as shown in *(i) – (iii)*. After waiting for τ, a 90_i^o pulse again generates observable magnetisation as shown in *(i) – (iii)*. The measured signal is the difference between the signal generated from decay of +z and -z initial magnetisations. In both, the graph represents how the signal intensity changes as a function of the time τ.

4.3. Relaxation Mechanisms

There are multiple mechanisms responsible of the nuclear relaxation in the sample. The observed relaxation will depend on the spin system present in the sample. Different sources of local fields can dominate the observed relaxation. These interactions that modify the local field at the nucleus are the shielding, dipole-dipole coupling and for spins with I>1/2, quadrupole coupling. The rate of the relaxation will depend on the strength of these interactions and the amplitudes of the motion that cause their time dependence. In quadrupolar systems, quadrupolar interaction will dominate the relaxation process; for ^{13}C and ^{15}N in the biological samples in this work, dipolar interaction is typically the main source of relaxation. For dipolar coupling, the transition rate is rate is $\propto 1/r^6$, where r is the internuclear distance between the relaxing spin and surrounding spins. The size of dipolar coupling between different nuclei that are present in biological materials are represented in Table 4.1. The size of the chemical shift anisotropy for the ^{13}C and ^{15}N are 12 – 18 and 4.0 – 4.5 kHz (at the spectrometer frequency of 400 MHz).[165,166]

Table 4.1. The size of the dipolar coupling involving ^{1}H, ^{13}C and ^{15}N are shown. The dipolar coupling was calculated using the Equation (3.4).

Nuclei Pairs	Distance / Å	Dipolar coupling / Hz	Environment
^{1}H-^{1}H	1.8	-20597	Non-bonded (in -CH_2)
^{1}H-^{1}H	4.0	-1877	Non-bonded
^{1}H-^{13}C	1.1	-22690	C-H bond
^{1}H-^{13}C	4.0	-472	Non-bonded
^{1}H-^{15}N	1.0	12179	N-H bond
^{1}H-^{15}N	4.0	190	Non-bonded
^{13}C-^{15}N	1.4	1116.	C-N bond
^{13}C-^{15}N	4.0	48	Non-bonded
^{13}C -^{13}C	1.5	2251	C-C bond
^{13}C -^{13}C	4.0	118	C=C bond
^{15}N-^{15}N	4.0	19	Non-bonded

Relaxation of the ^{13}C and ^{15}N nuclei in the materials used in this work is induced by fluctuations in dipolar interactions between NMR active nuclei as a result of a molecular motion. The main dipolar interaction in these unlabelled samples will be between $^{13}C/^{15}N$ and ^{1}H nuclei (Table 4.1.) which will dominate the $^{13}C/^{15}N$ T_1 relaxation. The T_1 value depends on the strength of ^{1}H – $^{13}C/^{15}N$ dipolar coupling and the timescale on which it is modulated by

molecular motions. In terms of the strength of ^{1}H-X dipolar interactions. For example, it is expected that the carbonyls C=O will have the highest T_1 value, followed by Cα, while the shortest T_1 values would be observed for the side-chains ^{13}C where each carbon atom has two or even three protons attached. In terms of the molecular motions present, the collagen triple helices can experience three different molecular motions, segmental, backbone and side-chain. Segmental motions are slow and involve many residues moving together. The backbone motions are local movements experienced by the nuclei that are located on the polypeptide chain – Cα, C' and amide N. Side-chain motions are experienced by all amino acid residues (except glycine) where the degree and the motion varies according to the residue considered. Side-chains can have group rotations and reorientations, in particular, methyl groups typically undergo rapid rotation, changing the proton positions in the process. Therefore, the side-chains carbons where there is a methyl group present will experience nanosecond timescale fluctuating fields from ^{1}H to a greater extent resulting in shorter T_1 values for sidechain ^{13}C, compared to the ^{13}C present in the collagen backbone.

Additionally, besides directly bonded protons, the surroundings of the nuclei observed can facilitate the relaxation of the^{13}C/^{15}N via proton spin diffusion (PSD). The ^{13}C/^{15}N nuclei are strongly coupled to their bound protons and further these protons can be affected by other neighbouring protons. For example, bound water in the biological samples can facilitate magnetisation transfer from one spin to another via spin flip-flop. A strong PSD effect can homogenise the observed relaxation of ^{13}C/^{15}N nuclei, where the observed relaxation constant for different residues can be the same across different protein regions.

4.4. Inverse Laplace Transform for Relaxation Data Analysis

In a standard NMR relaxation experiment, the intensity of a single peak is measured as a function of a relaxation delay τ. If the peak corresponds to a single site, the signal is expected to decay exponentially as was described in the previous section. However, where the observed signal intensity can contain a sum of multiple signals that are coming from different contributors that happen to have the same or very similar chemical shifts, the decay in intensities observed in relaxation experiments will be a combination of two or more relaxation decays with different relaxation rates. This is frequently the case for complex biological samples such as those studied in this thesis. Ideally, to interpret such relaxation data, one is required to be able to extract intensities and decays of the individual components. The sum of such exponential decays of the signal as a function of a decay time can be written as:

$$S(\tau)=\sum S_i(0)\exp(-\frac{\tau}{T_{1i}}) \quad (4.4)$$

where *i* is a number of components with their appropriate initial intensities $S_i(0)$ and decay time constants T_{1i} that are not known. If this sum is represented as an integral over τ, it can be written as the Laplace transform of the distribution of relaxation times $G(T_1)$:

$$S(\tau)=\int_0^\infty exp\left(-\frac{\tau}{T_{1i}}\right) G(T_1)dT_1 \quad (4.5)$$

By using the inverse Laplace transform (ILT) on the spectral intensities as a function of the τ delay in the relaxation experiment, the distribution of relaxation times $G(T_1)$ can be obtained:

$$G(T_1)=\frac{1}{2\pi}\int_0^\infty S(\tau)\exp\left(-\frac{\tau}{T_{1i}}\right) ds \quad (4.6)$$

The ILT is an ill-posed problem, it cannot give real, direct solution, but can give a statistically reasonable solution. It is plausible that there will be many possible functions $F(\tau)$ where:

$$F(\tau) =\sum G(T_{1i})\exp(-\frac{\tau}{T_{1i}}) \quad (4.7)$$

give a reasonable fit to the data as a function of τ. Knowing that there is no unique solution to the ILT, help to determine the most likely solution is required. If the data are noise free, then the function that fits best to the data will be the true solution. However, in case of NMR, the data is noisy, usually with 2–10 % noise level, and thus many possible solutions might give a similar fit to the data. To overcome this, a regularisation method is used. In this work, a least-squares fitting method is adopted, which imposes some independent constraint on the solution.

We used the program rilt.m (regularised inverse Laplace transform) which was written by Lari-Gabriel Marino (University of Parma, Italy),[167] further modified by Dr Veronica Wong (Duer group, University of Cambridge, UK). The script was adapted for MATLAB use from Stephen W. Provencher's program CONTIN, originally written in Fortran.[168,169]

The ILT has been used for processing relaxometry data, where the relaxation rate as a function of the magnetic field strength is measured.[170] The ILT approach was first introduced in solid-state NMR for separating overlapping second-order quadrupolar powder patterns of static samples by Lupelescu et al,[171] where results indicated that the ILT is sensitive towards noise and baseline distortions. However, results showed that the ILT can separate the

differently relaxing components; even up to three differently relaxing components can be separated when the signal-to-noise ratio is very good.

4.4.1. Testing the ILT – Ideal Data

In this section, application of the rilt.m code to the data is tested and discussed, in particular with reference to its accuracy when there is noise in the spectrum and thus in the intensities of the NMR signals being analysed. The reliability of the results from the code was tested in two ways: (1) by simulated relaxation data using Equation (4.2) to generate discreate signal intensities; (2) use of model sample with experimentally obtained relaxation values processed using rilt.m and compare to relaxation constant calculated via conventional methods and literature values.

In general, for good analysis of relaxation decay it is advised to spread the time points τ at which the NMR signal intensity is measured in an exponential manner. Furthermore, sampling close to $\tau = 0$ and in the vicinity of the expected actual relaxation value T_1 is advised. A problem arises for nuclei that have long T_1 relaxation constants. For example, the T_1 for ^{15}N can be up to minutes in very crystalline samples, which will result in long experimental times, especially if the sample has poor sensitivity that will require a large number of scans to get a reasonable signal-to-noise ratio. The usual number of sampling points in τ to obtain well-described exponential decay is around 30 to 40. Therefore, in this section the optimal number of sampling points will be evaluated. The effects of the noise will then be investigated as well to see to what effect it has overall on the precision of calculations.

4.4.1.1. Single Exponential Decay

The first example considers only single exponential decay, i.e. one relaxation constant. The decay data points (τ values) were spaced exponentially in range of 0.01 – 100 s, with the relaxation time constant set to 1 s. Calculations were carried out with added random noise at 0%, 1%, 5%, 10% and 20% to mimic the range of signal-to-noise ratio that can be seen in real NMR spectra. Additionally, the number of selected decay data points was varied to see what effect it has on the calculated $G(T_1)$ solution. In this case, τ values were selected to be 10, 20, 25 and 50 exponentially-distributed points within the previously given range (Figures 4.5. – 4.8.). The solutions were evaluated at the same number of τ points for each example, T_1 points were spaced out logarithmically between 0.01 – 100 s.

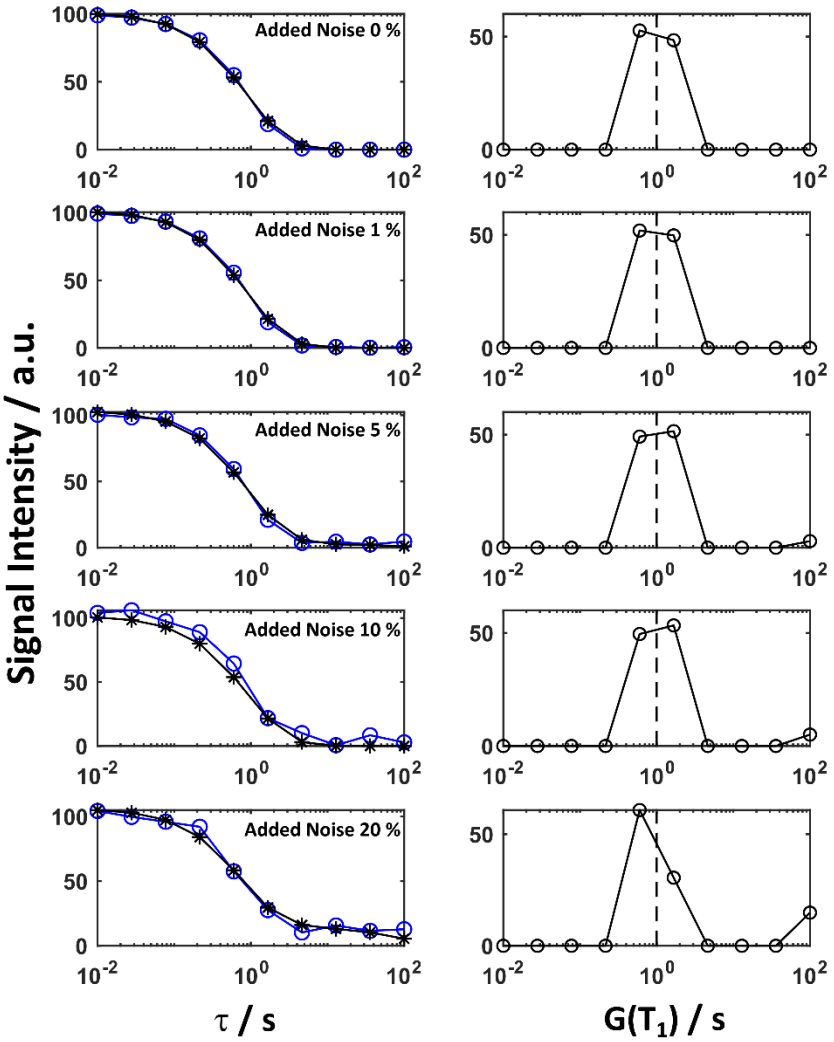

Figure 4.5. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with random noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT-fitted decay. A relaxation time constant of 1s was used to generate the exponential decay curve shown with 10 τ data points. Right: Black, circles joined by a solid line represent the calculated distribution of T_1 and the dotted vertical line shows actual T1 value of the simulated data. The solution is evaluated using 10 points. The rilt.m code estimates T_1 value between 0.59 – 1.67 s for data with 0%, 1%, 5%, 10% noise level. However, for data with 20% noise relaxation constant maxima is calculated to be at 0.59 s, with shoulder at 1.67 s. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise.

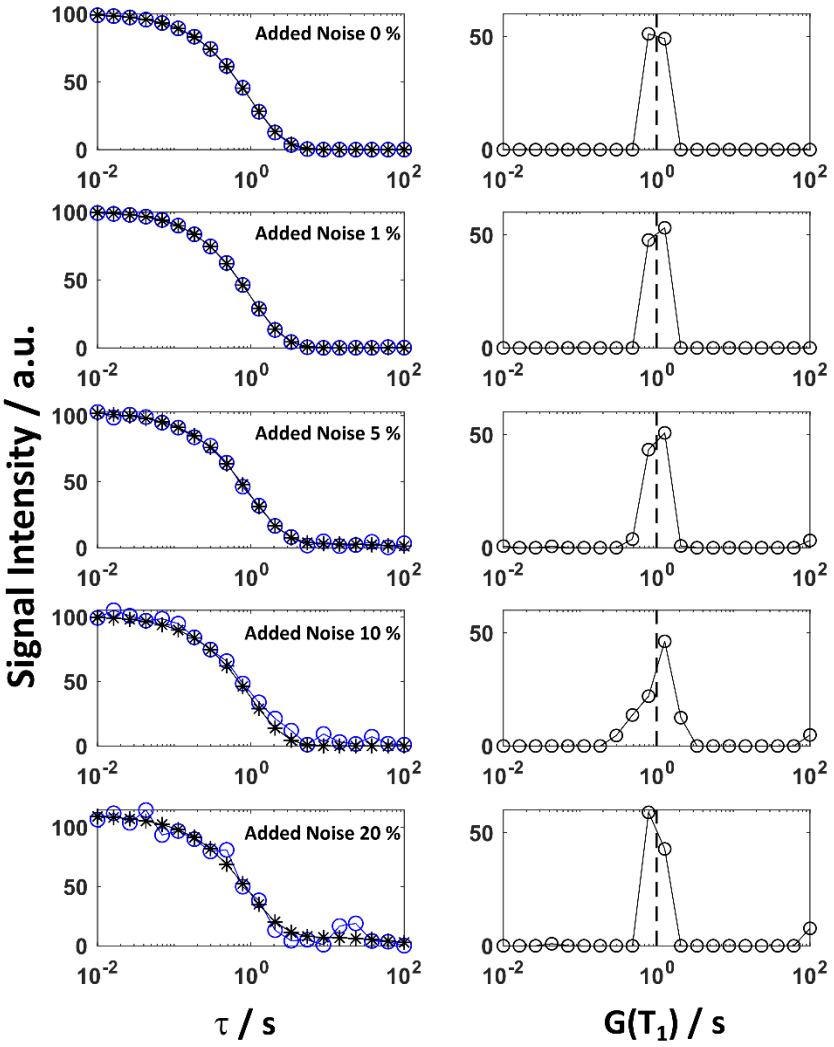

Figure 4.6. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. A relaxation time of 1 s was used to generate given exponential decay with 20 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical line shows actual distribution of the simulated data. The solution is evaluated using 20 points. The rilt.m code estimates T_1 value between 0.78 – 1.27 s for data with 0%, 1%, 5% and 20% noise level. However, for data with 10% noise relaxation constant maxima is calculated to be at 1.27 s with shoulder below 0.78 s. Additionally, small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise.

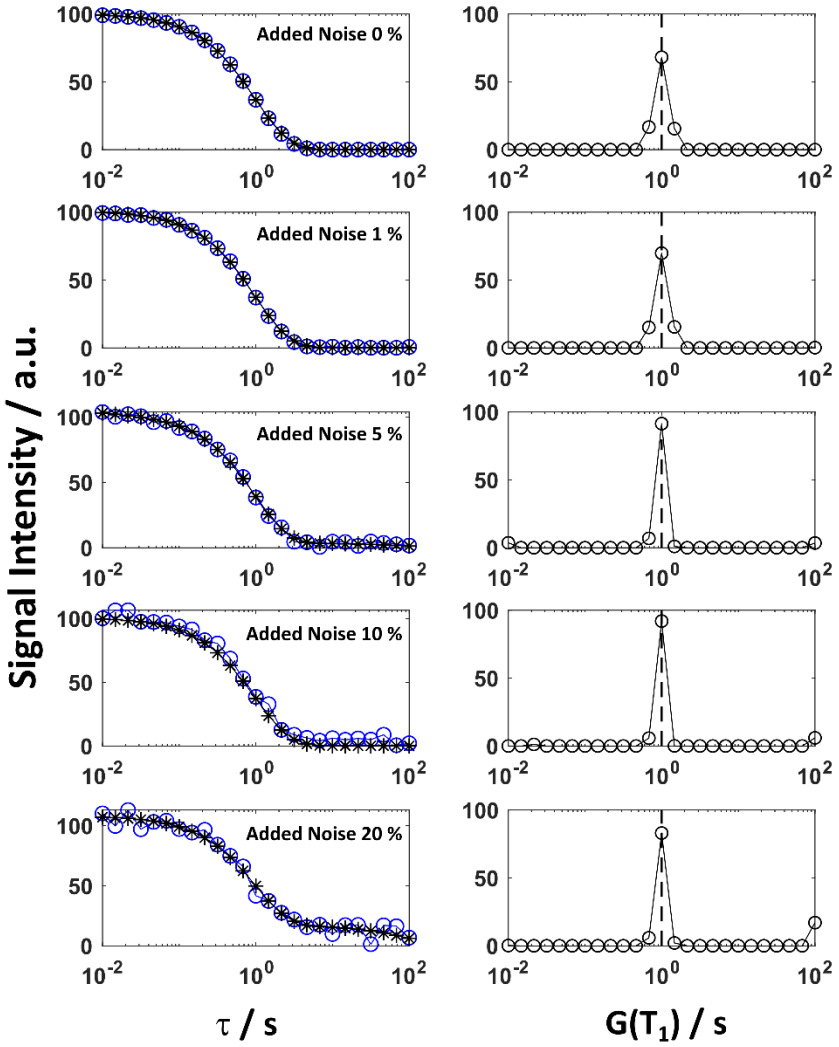

Figure 4.7. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. A relaxation time of 1 s was used to generate given exponential decay with 25 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical line shows actual distribution of the simulated data. The solution is evaluated using 25 points. The rilt.m code estimates T_1 value at 1 s for all data. Additionally, small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise.

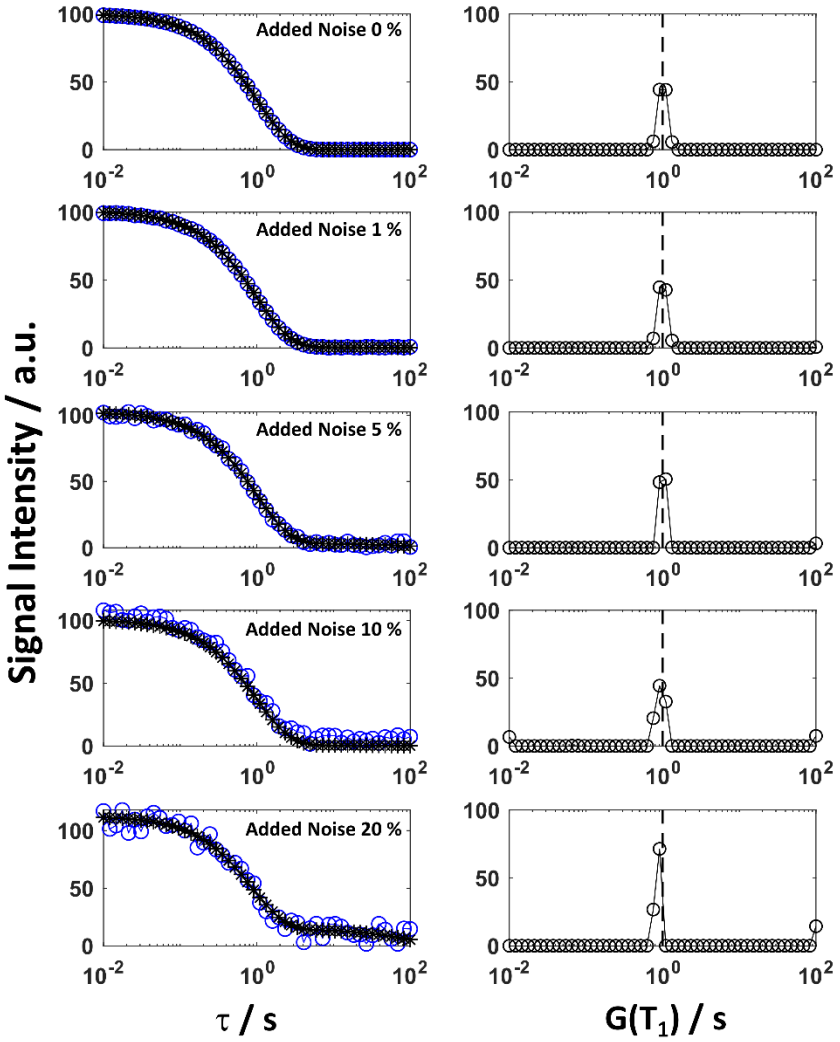

Figure 4.8. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. A relaxation time of 1 s was used to generate given exponential decay with 50 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical line shows actual distribution of the simulated data. The solution is evaluated using 50 points. The rilt.m code estimates T_1 value between 0.91 – 1.11 s for data with 0%, 1%, 5% and 10% noise level. However, for data with 20% noise relaxation constant maxima is calculated to be at 0.91 s. Additionally, small shoulder above 80 s is observed to increase in intensity for solutions using data with 1%, 5%, 10% and 20% noise.

The results show that the ILT gives a reliable estimation of the relaxation constant. It is clear that resolution in the T_1 dimension is a potential issue. With increasing number of τ and solution points for the $G(T_1)$, the width of the signal in the $G(T_1)$ decreases and becomes sharper. With increasing noise level, the estimation of the relaxation constant does not change within each data set (number of data points used), except, for the examples with 10 τ values with 20% noise level, 20 τ values with 10% noise level, and 50 τ values with 20% noise level where T_1 maxima is calculated to be at 0.59, 1.27 and 0.91s, respectively. However, the calculated value is not far from actual T_1 relaxation constant. This could be due to the noise level that the data contains, thus affecting fitting of the exponential decay.

4.4.1.2. Multiple Exponential Decays

A sum of two exponential functions with relaxation constants set to 1 s and 5 s with ratio of 2:1 in relative intensities were tested. The τ decay values and the noise level were selected to be the same as in previous examples (Figures 4.9. – 4.12.).

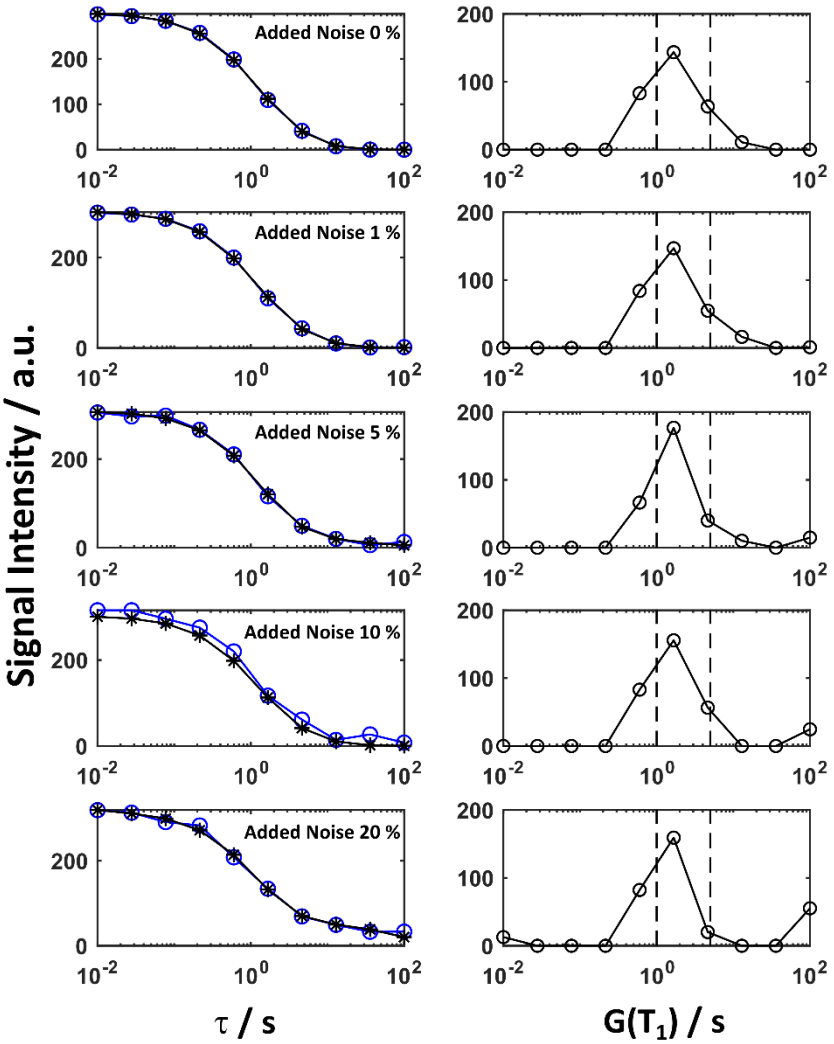

Figure 4.9. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. The sum of exponential decay was generated using relaxation constants set to 1 and 5 s with ratio of 2:1 with 10 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical lines show actual distribution of the simulated data. The solution is evaluated using 10 points. The rilt.m code estimates one T_1 maxima at 1.67 s in all examples. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise; for 20% noisy data additional signal below 0.03 s is observed.

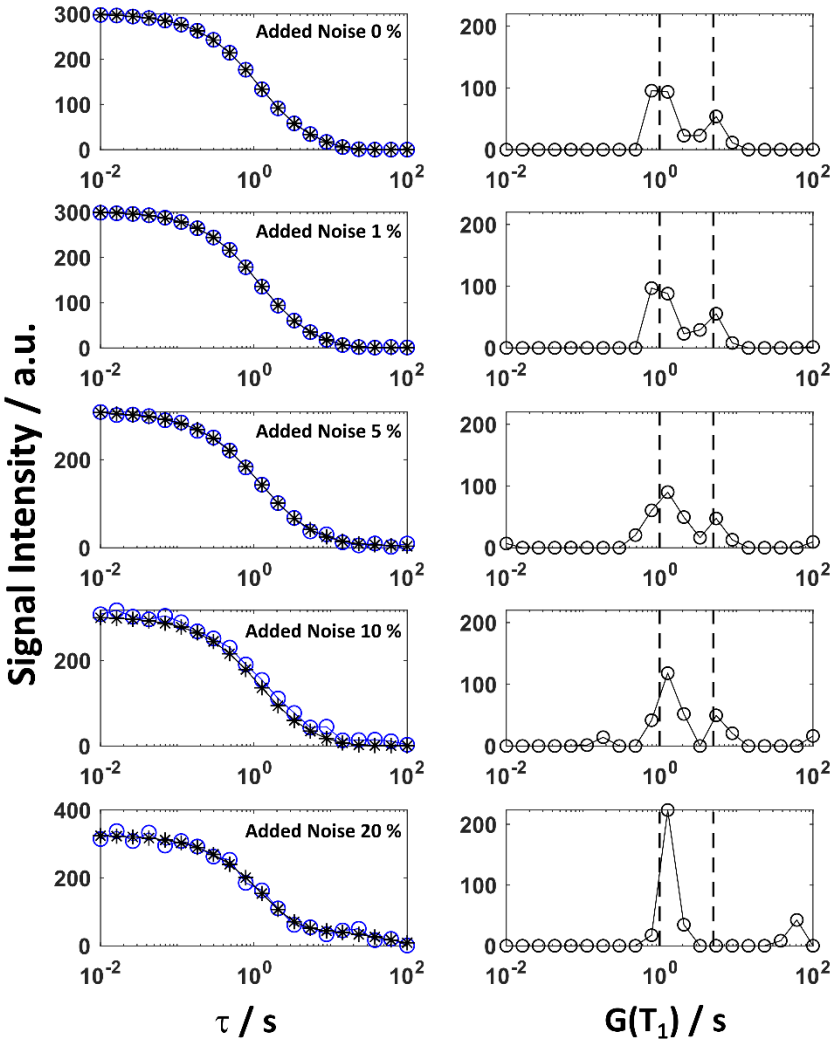

Figure 4.10. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. The sum of exponential decay was generated using relaxation constants set to 1 and 5 s with ratio of 2:1 with 20 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical lines show actual distribution of the simulated data. The solution is evaluated using 20 points. The rilt.m code predicts the first T_1 value between 0.79-1.27 s for data with 0% and 1% noise, and at 1.27 s for rest of the data, the second T_1 value is estimated at 5.46 s, except for the data with 20% noise which is calculated at 61.58 s. For the data with 10% noise, the third T_1 value is observed at 0.18 s. The ratios of calculated signals are (from top to bottom): 1.77, 1.75, 1.82, 2.38 and 5.26. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5% and 10% noise.

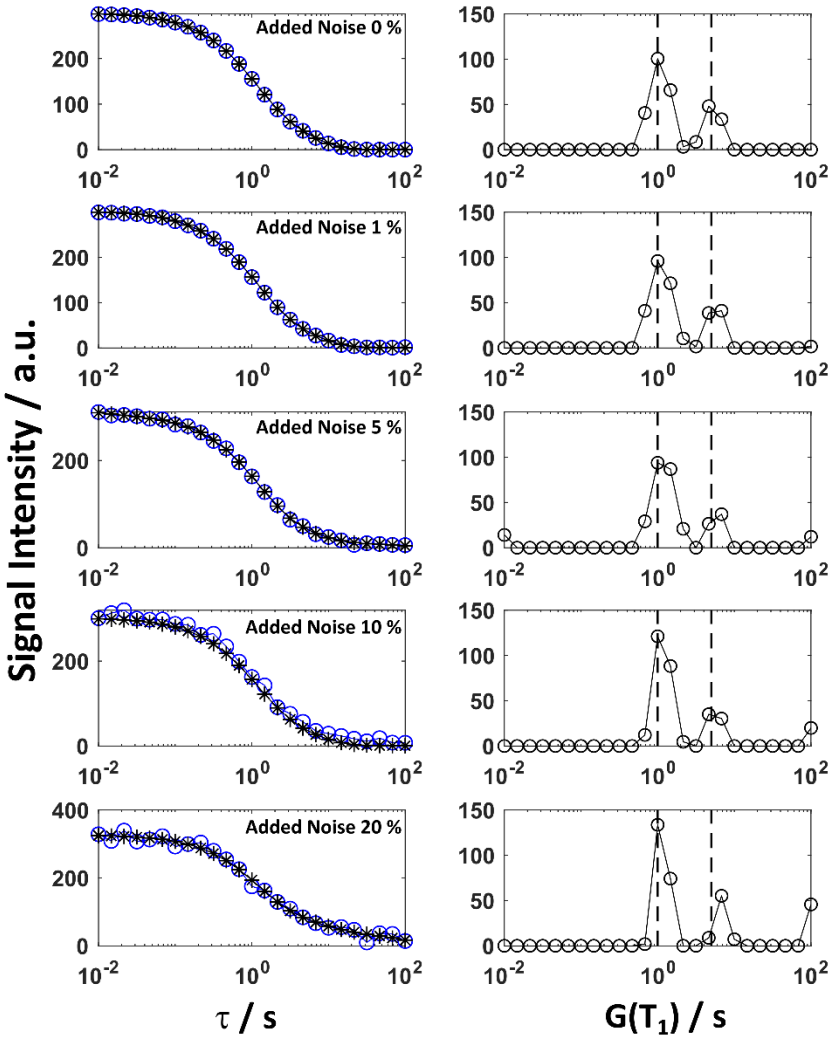

Figure 4.11. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. The sum of exponential decay was generated using relaxation constants set to 1 and 5 s with ratio of 2:1 with 25 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical lines show actual distribution of the simulated data. The solution is evaluated using 25 points. The rilt.m code estimates the first T_1 value at 1 s in all examples, the second T_1 between 4.64 – 6.81 s, except for the data with 20% noise second relaxation value is calculated at 6.81 s. The ratios of the two calculated T_1 components are (from top to bottom): 2.08, 2.43, 2.53, 3.46 and 2.42. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise.

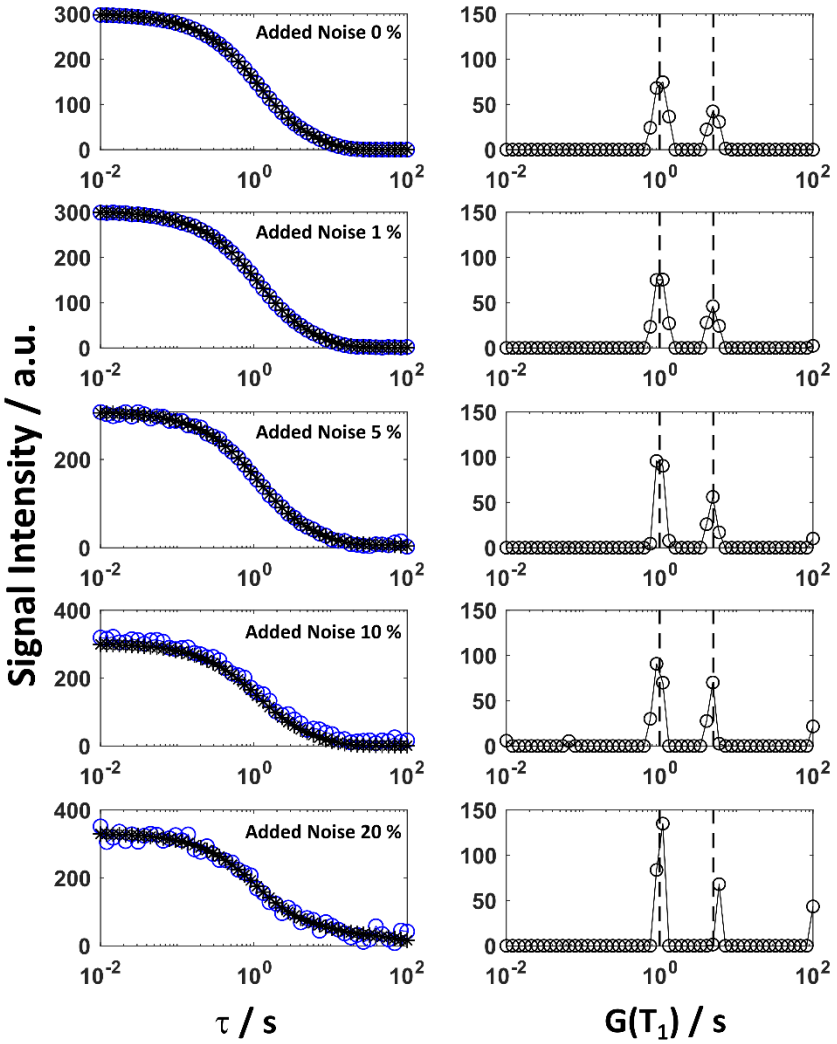

Figure 4.12. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. The sum of exponential decay was generated using relaxation constants set to 1 and 5 s with ratio of 2:1 with 50 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical lines show actual distribution of the simulated data. The solution is evaluated using 50 points. The rilt.m code estimates first T_1 value between 0.91 – 1.09 s for all data except for data containing 20% noise where T_1 value is at 1.09 s, the second T_1 relaxation value is calculated to be at 4.94 s in all examples except data with 20% noise where it is at 5.96 s. The ratios of the two calculated T_1 components are (from top to bottom): 1.76, 1.63, 1.71, 1.3 and 1.98. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise. Small extra T_1 value at 0.07 s is observed for data with 10% noise.

Similarly, as in previous examples where one relaxation constant was used, the higher number of decay and solution points give greater precision of the calculation. The estimation of the relative signal intensity ratios of two relaxation constants was not reliable as it is varying significantly between increasing number of τ values and the noise level added – there was no correlation detected.

With increasing noise, the T_1 distribution showed additional slow and fast relaxing components at extremes of $G(T_1)$. Solutions with 25 and 50 decay points are able to define two relaxation constants. However, when the noise is at 20%, the second T_1 component is calculated at 61.28 s for data with 20 decay points (Figure 4.10.); for data with 25 points the second T_1 constant is estimated only slightly higher than its actual value (5 s) at 6.81 s (Figure 4.11.). In all examples above 5% noise, extra signal was observed above 80 s, similarly to the examples with single exponential decay. This shows how sensitive the ILT can be towards noise, and how important it is to obtain good signal-to-noise ratio in relaxation experiments if we are to use the ILT.

In the rilt.m the number of the points used to calculate the $G(T_1)$ can be increased to gain better prediction for the relaxation constant. As it was observed previously that more points in exponential decay curve can help to gain better estimate of the relaxation constant. In the following examples, the optimal number of solution points in $G(T_1)$ will be investigated. In the first instance to test this, data containing 25 points in the exponential decay were used, and calculations were carried with one and two relaxation constants as in previous examples used. The relaxation data was then fitted with 50 τ values (Figures 4.13. and 4.14.).

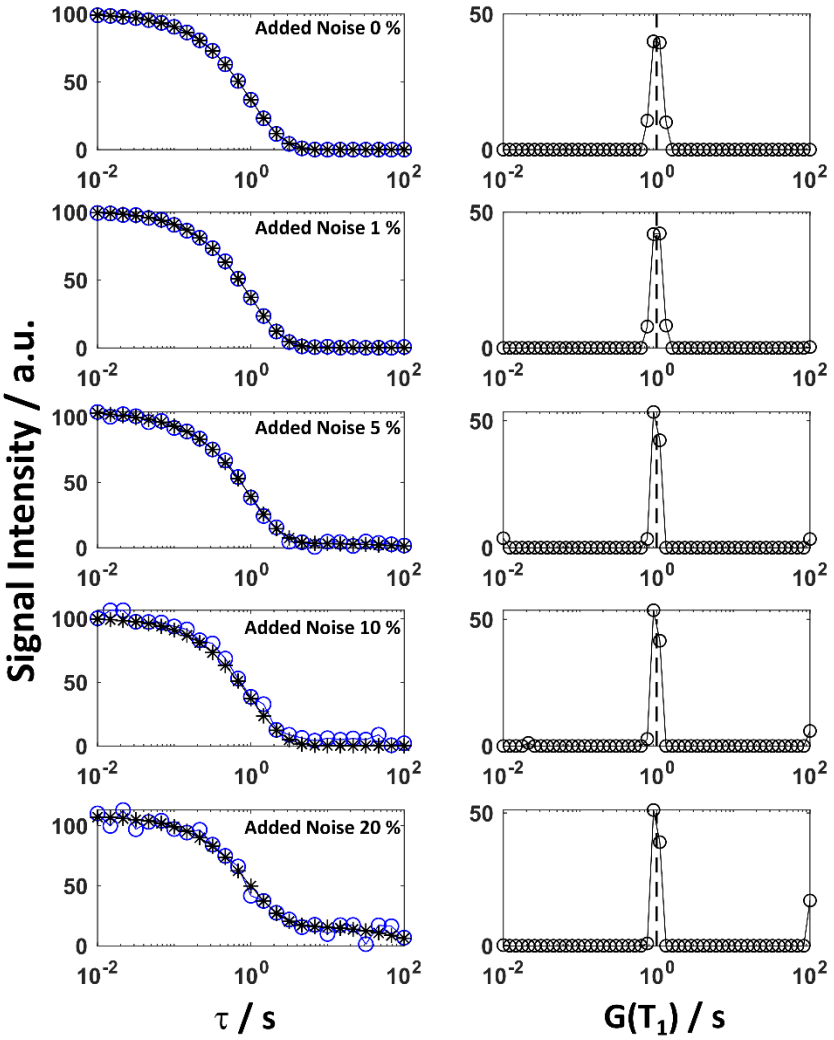

Figure 4.13. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. A relaxation time of 1 s was used to generate given exponential decay with 25 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical line shows actual distribution of the simulated data. The solution is evaluated using 50 points. The rilt.m code estimates T_1 value between 0.91 – 1.09 s for all data. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise. Small extra T_1 value at 0.02 s is observed for data with 10% noise.

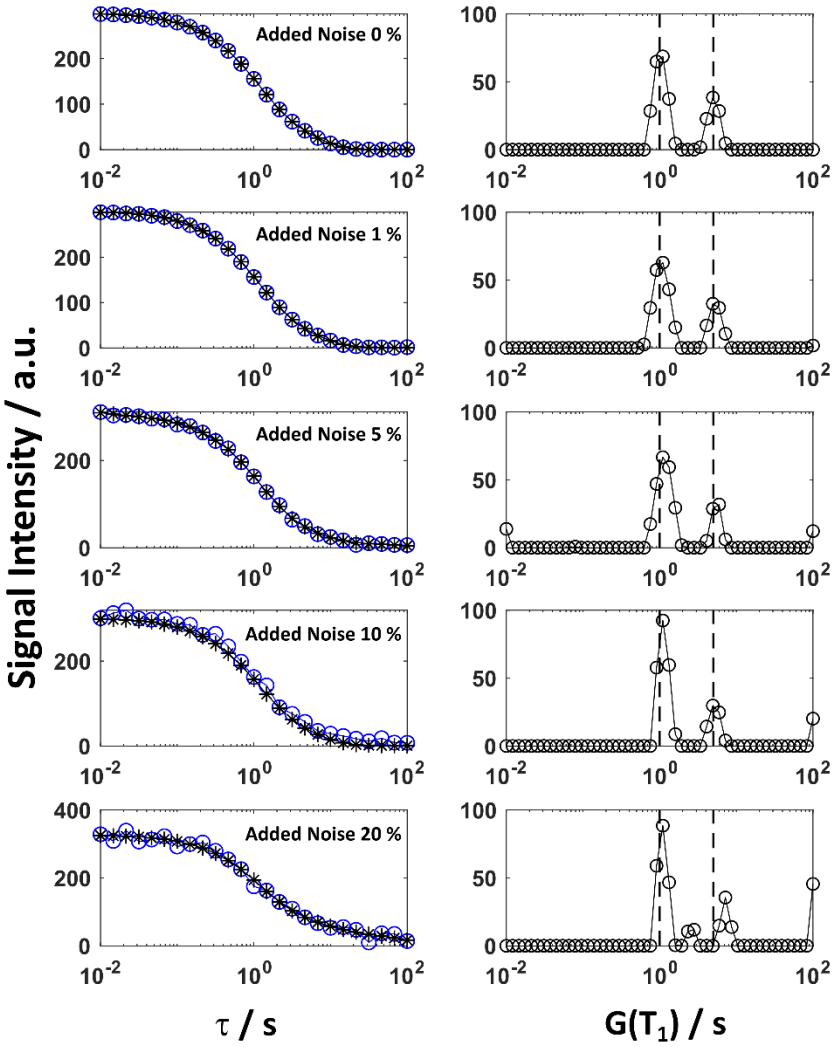

Figure 4.14. Left: Blue, round spheres joined by a solid blue line represent (from top to bottom) simulated relaxation data with noise of 0%, 1%, 5%, 10% and 20% added. Black, asterisks joined by a solid line represent the ILT fitted decay. The sum of exponential decay was generated using relaxation constants set to 1 and 5 s with ratio of 2:1 with 25 data points. Right: Black, circles joined by a solid line represent calculated distribution of T_1 and the dotted vertical lines show actual distribution of the simulated data. The solution is evaluated using 50 points. The rilt.m code estimates first T_1 value between 0.91 – 1.09 s for data with 0% and 1% noise, for rest of the data at 1.09 s. The second T_1 relaxation value is calculated to be at 4.94 s for data with 0% noise, and between 4.94 – 5.96 s for data with 1%, 5% and 10% noise. In example with 20% noise, second T_1 relaxation value is predicted at 7.19 s, additional third value is calculated between 2.33 – 2.81 s. The ratios of the two calculated T_1 components are (from top to bottom): 1.69, 1.92, 3.1, 2.09, 3.1 and 2.48. Additionally, a small shoulder above 80 s is observed to increase in intensity for solutions using data with 5%, 10% and 20% noise.

For both single and two exponential decays, the signal width in the $G(T_1)$ is decreased by almost a half as a result of increasing the number of τ points used to define G(T1). The results show for single relaxation constant the solutions are as good as for exponential data with noise at 0% and 1% and the relaxation decay defined by 50 points (Figure 4.13.).

It is worth noting that the solution of the relaxation constant when using relaxation data defined by 25 points is calculated to be exactly 1 s, whereas the solution for relaxation data defined by 50 points it is not. This is because each relaxation distribution contains values that are logarithmically distributed in a range that is set by user. In this case the range is set from 0.01 to 100 s, where for 25 values it happens to have one of the $G(T_1)$ values to be exactly at 1 s. For the 50 values, the logarithmic scale does not contain a value at exactly 1 s but it has values that are at 0.91 and 1.09 where the maximum of the $G(T_1)$ function is observed. This result shows that instead of reading the max value from the graph, an average between these two points could be given with an error as a width of the observed signal in the $G(T_1)$ frequency. This particular approach of taking an average of the relaxation constant is done for the analysis of the ^{15}N T_1 relaxation data of the model peptides (Chapter 6.).

In the example where the sum of two exponentials is processed, the precision of solution for the relaxation constants is calculated well for all data except for the example with 20% noise. The solution for the data with highest noise level gives three relaxation values with maxima at 1.09, a small signal between 2.33 – 2.81 and 7.19 s. Overall, the relaxation values are better estimated compared to the solution with 25 points (Figure 4.10.). However, the results show noticeable effects from noise that can give apparent relaxation constant values. Thus, one should exercise caution when using data that has 20% noise or higher.

As it was observed that the width of the signals in $G(T_1)$ decreased as more points are used in the solution, the program was further tested using only data with 10 decay values with 0%, 10% and 20% noise fitted with 10, 20, 30, 40, 50, 60, 80 and 100 points. This was carried out to determine if the precision of the solution will increase with increasing number of fitted points (Figures 4.15. – 4.16.).

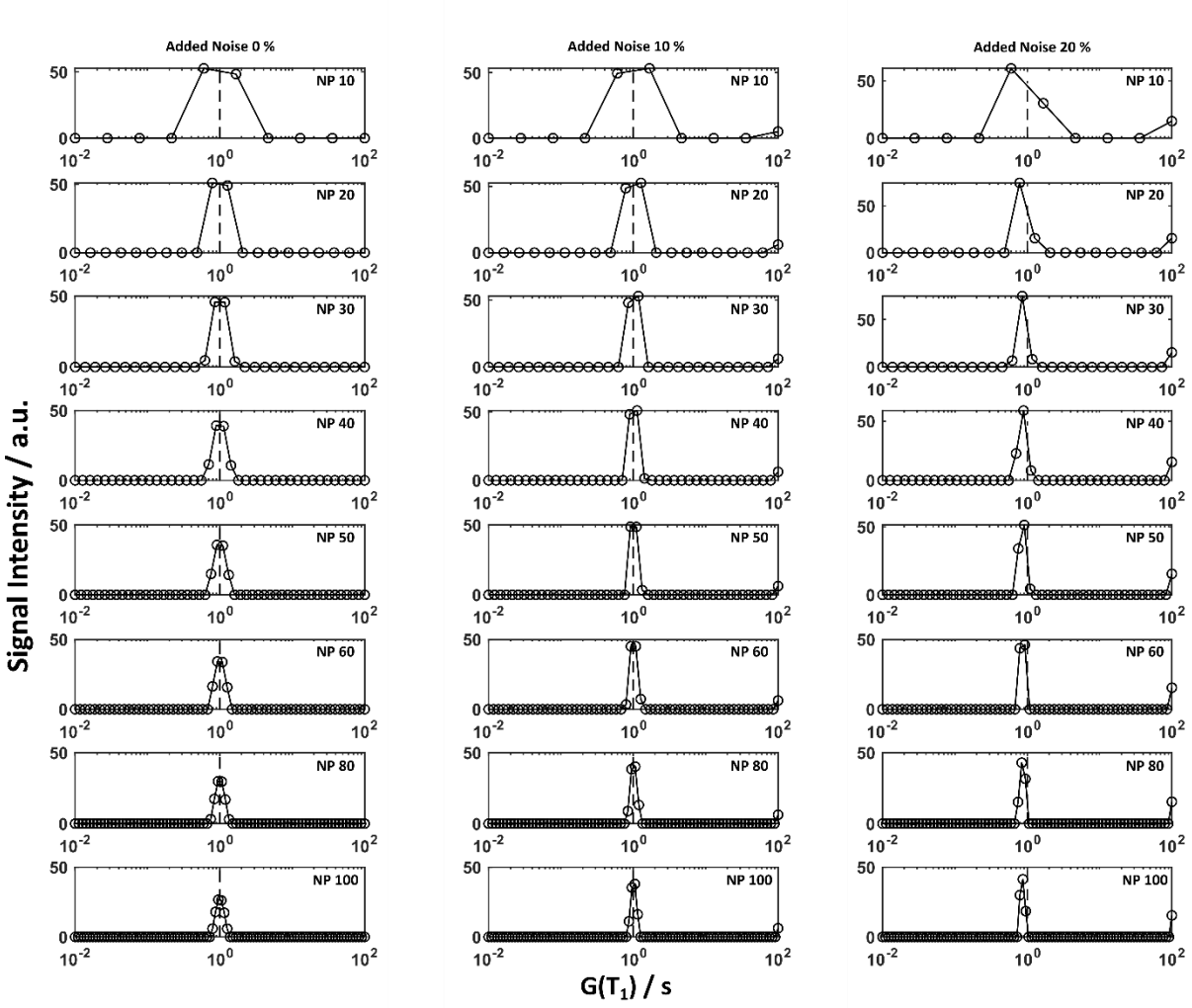

Figure 4.15. Black circles joined by a solid line shows calculated distribution of T_1 and the dotted vertical line at 1 s show actual distribution of simulated data. From left to right data used were with 0%, 10% and 20% noise level, fitted with (from top to bottom) 10, 20, 30, 40, 50, 60, 80 and 100 number of points (NP) for solution. Precision of calculated value is increasing with increased number of points used per solution.

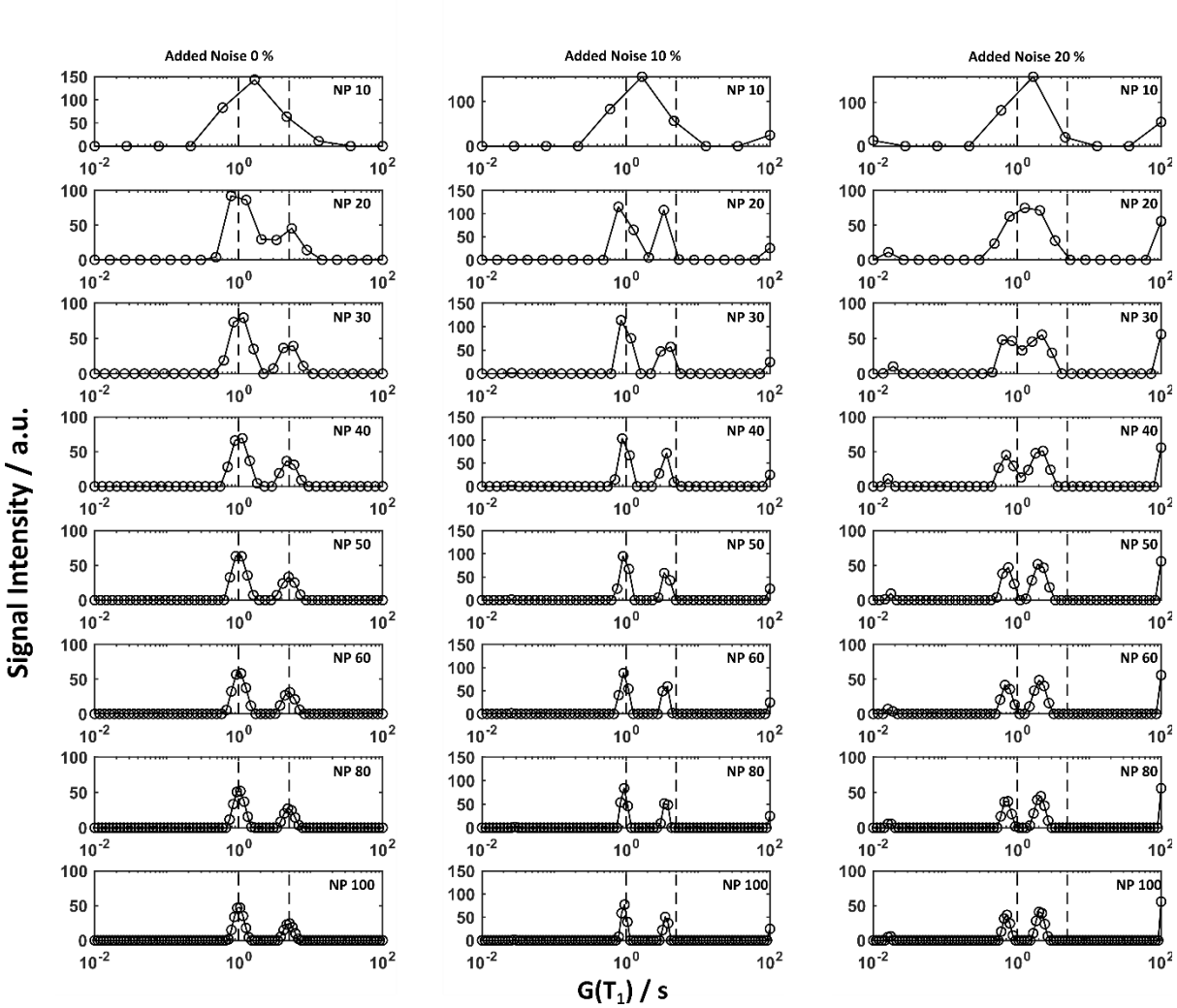

Figure 4.16. Black circles joined by a solid line shows calculated distribution of T_1 and the dotted vertical lines show actual distribution of simulated data. The simulated data consisted of sum of exponentials with relaxation constants with 1 and 5 s. From left to right, the data used were with 0%, 10% and 20% noise level, fitted with (from top to bottom) 10, 20, 30, 40, 50, 60, 80 and 100 number of points (NP) for solution. The precision of calculated values is increasing with increased number of points used per solution for data with 0% noise. For data with 10% added noise, the first T_1 relaxation constant is calculated correctly but the other is predicted to be at 3.5 s. The data with 20% added noise shows a shift to lower T_1 values in both signals.

The widths of the signals in $G(T_1)$ decrease with increasing number of points in the $G(T_1)$ solution in all examples. Thus, the relaxation constants are estimated with high precision, especially for data with one relaxation constant. However, for data with higher added noise, extra components are observed at both extremes of the $G(T_1)$ range. Noise has a greater effect on the precision of estimating relaxation constants for data with a sum of exponentials.

These results show that multiple considerations must be taken into account when the ILT is applied to experimental NMR data: (1) the number of the sampled τ points need to be as high as possible; (2) the precision of the prediction and resolution of the $G(T_1)$ solution can be increased if the solution is evaluated at more T_1 points than the exponential decay is sampled with; (3) the signal-to-noise ratio needs to be below 20%, otherwise the estimated T_1 values could potentially contain significant error. This method cannot be used to predict the intensity ratios of the relaxation components as this parameter is significantly affected by noise; however, a reasonable estimate of the signal intensity ratio can be obtained if the noise level is close below 5%. To gain better T_1 resolution, the range of T_1 solutions considered can be decreased, as the ILT algorithm uses T_1 values that are logarithmically distributed, meaning that there will be fewer points at the upper end of the $G(T_1)$ range.

4.4.2. Testing the ILT on Experimental Data of ^{2}H-PMMA-d$_8$

We next tested the ILT with experimental NMR data and compared the results with T_1 results using conventional relaxation analysis and literature values for T_1. The conventional analysis was done using Bruker software, where the integrated signal area was taken.

The sample selected was deuterated polymethyl acrylate (PMMA-d$_8$). This sample is a good test material as the alkyl and ester methyl groups have very similar quadrupole splitting of around 39.8 and 36.0 kHz, respectively[172] and in the static 1D spectrum they are indistinguishable (Figure 4.17. (a)). Furthermore, the relaxation constants in the literature are very different for these two methyl groups; the alkyl methyl has T_1 of 0.008 s and the ester methyl of 0.37 s. The methylene deuterons have a quadrupolar constant of around 120 kHz, and the relaxation constant in the literature for these deuterons is 2.4 s, although the work acknowledged that this T_1 constant might not be estimated correctly.

T_1 relaxation data were collected using inversion recovery solid echo pulse sequence for T_1 measurement (see Figure 3.10. of the pulse sequence used in section 3.2.8.). τ delays were selected to be 0.0001, 0.0002, 0.0003, 0.004, 0.009, 0.025, 0.05, 0.12, 0.36, 0.8, 2, 3, 5, 10, 15 s.

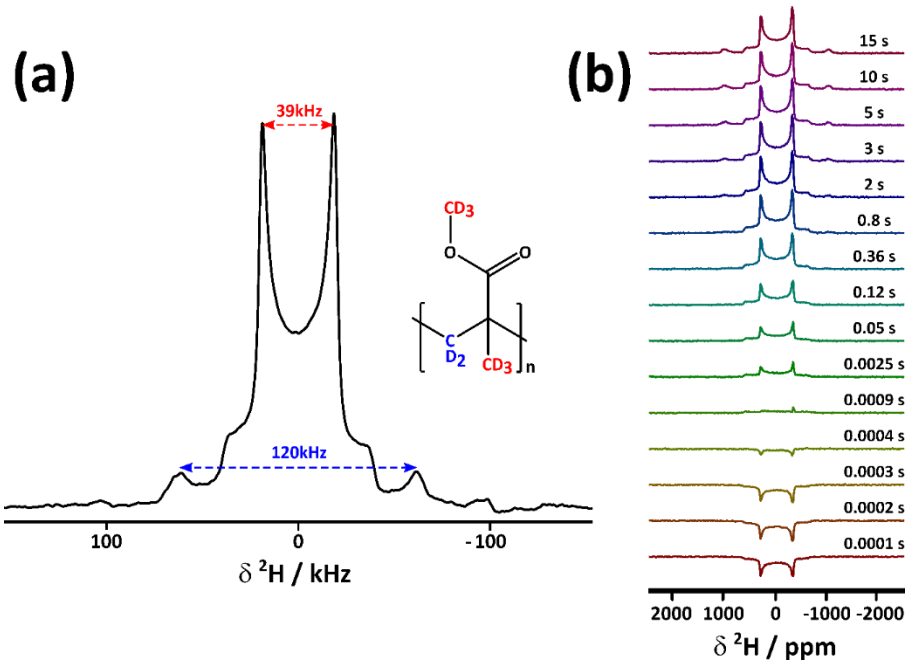

Figure 4.17. (a) 1D ^{2}H quadrupolar echo NMR spectra of PMMA-d_8. The line-shape was obtained with 512 scans with 5 s recycle delay; exponential line broadening of 1000 Hz was used in processing the FID. On the right, the structure of PMMA-d_8 is shown. In blue is highlighted the PMMA methylene group, in red, the methyl groups. (b) Stack of 1D spectra acquired using the inversion recovery experiment. τ delays are shown next to each spectrum. Each data set was acquired using 320 number of scans with 10 s recycle delay.

The NMR spectra prior to the ILT can be processed in two ways. If peaks in the NMR spectrum are well-resolved, then integration over the peak region can be performed, and a relaxation decay curve can thus be generated for each peak. Integrating the signal intensity helps to improve signal-to-noise ratio for each relaxation decay curve. Alternatively, the whole spectrum can be used, where relaxation decay curve for each point in the NMR frequency range can be independently processed by the ILT. This is necessary when frequency signals overlap or in the case of the ^{2}H powder pattern here, where the lineshape is expected to have different T_1 values at different frequencies. This latter approach was used in this example here of PMMA-d_8.

In this example, 4096 points were used across the whole frequency range for each 1D NMR spectra giving approximately one point per ppm. In the ILT analysis, a stack of such 1D NMR spectra are processed where for each stack of points with respect to the τ delay is fitted with a sum of exponential functions. This results in a 2D spectrum that demonstrates how the T_1 time relaxation constant varies through the whole spectral frequency range. The final 2D spectrum of the ^{2}H relaxation data of PMMA-d_8 after the ILT processing is shown in Figure 4.18. (a). In this example, the solution was estimated using 30 points for the G(T_1).

The contour plot in Figure 4.18. (a) shows three strong signal regions that are nearly symmetric around the spectral frequency 0 ppm. The strongest signals are observed at 298, 278, -318 and -298 ppm that correspond to the methyl groups in PMMA-d_8. Additionally, lower intensity signals at 1023 and -1023 ppm are observed and correspond to the methylene group. The separation between the signal pairs is 37.8 kHz, 35.4 kHz and 125.7 kHz, which are close to the ones reported by Early et al. (38.4, 38.0 and 124 kHz) corresponding to the alkyl and ester methyl groups, and methylene groups, respectively.[172] The differences in determined values with those in the literature are due to the resolution of the frequency range. Furthermore, experiments we performed here are carried out at 290 K temperature which was maintained with temperature control through the whole experiment. In the report from Early et al., it is simply stated that the measurements were performed at the room temperature, not mentioning if the temperature was controlled or not. It is common that the temperature can change during the experiment time by radio frequency coils warming up during the rapid pulsing.

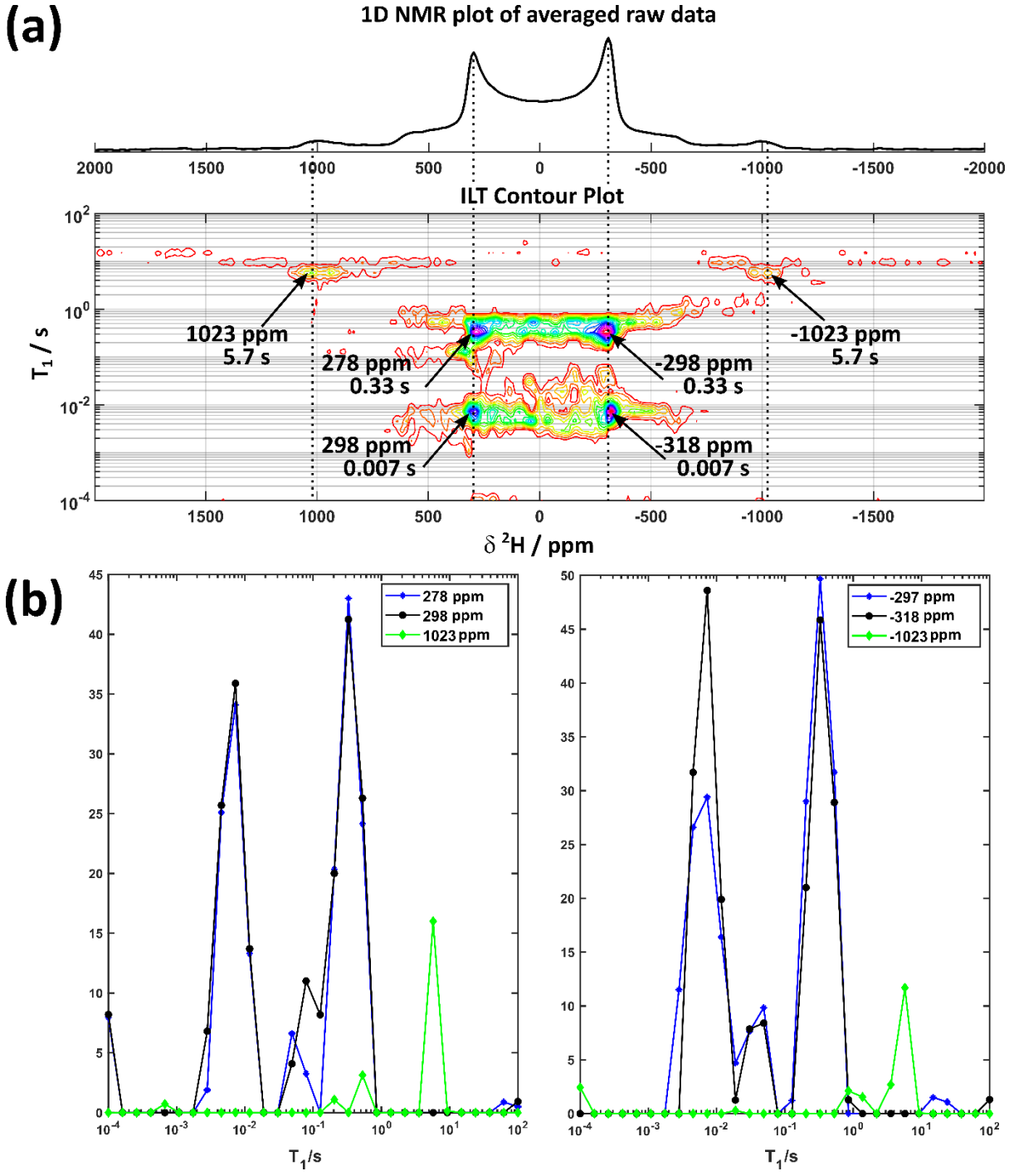

Figure 4.18. (a) The contour plot of the relaxation time constant T_1 versus a ^2H chemical shift of the PMMA-d_8. The T_1 region was estimated with 30 points exponentially spaced between 0.0001 and 100 s. On top of the contour plot is shown 1D ^2H NMR spectra after combining every four data points prior to the ILT with the peaks of the interest correlated to the contour plot highlighted with dashed lines. Chemical shifts and relaxation constants are highlighted with arrows. (b) ^2H T_1 relaxation time constants distributions associated with relaxation of PMMA-d_8 extracted slices from the contour plot along the dashed lines (on the left) at 1023, 298 and 278 ppm; and (on the right) at -298, -318 and -1023 ppm.

The extracted relaxation constants from the contour plot with their chemical shifts are summarised in Table 4.2. The distribution of $G(T_1)$ for each part of the ^2H lineshape of interest is extracted from the contour plot and shown in Figure 4.18. (b). The error is taken as a half-width of the $G(T_1)$ signal intensity.

Table 4.2. The relaxation constants calculated for the PMMA-d_8 using the ILT with errors. The T_1 region was estimated with 30 points exponentially spaced between 0.0001 and 100 s. Errors were taken as a half-width of the signal intensity (Figure 4.18. (b)).

^2H δ_{iso} / ppm	T_1 / s	Literature[172]	Assignment
102	5.7 ± 1.6	2.4	methylene
298	0.007 ± 0.003	0.008	alkyl methyl
278	0.3 ± 0.2	0.38	ester methyl
-298	0.3 ± 0.2	0.38	ester methyl
-318	0.007 ± 0.003	0.008	alkyl methyl
-1023	5.7 ± 1.6	2.4	methylene

The relaxation constants estimated for the alkyl and ester methyl groups are in good agreement with the literature values within their error estimates. The methylene group T_1 is more than two times longer than expected. However, it was noted in the previous work that this value might not be calculated correctly as the model used for estimating the T_1 constants were not suitable for these ^2H. The number of τ points used to calculate the T_1 value for the methylene residue here might not be enough, in Figure 4.17. (b) the number spectra that had sufficient signal intensity for this component to be used for the ILT is just five (from 2 to 15 s).

The error ranges of T_1 values shown in Table 4.2. are increasing with the increasing T_1 value. This is because the points for the solution were distributed exponentially across the $G(T_1)$ range, and thus the width of the $G(T_1)$ signal is greater for longer relaxation constants. This issue can be addressed in two ways: by using more points to estimate the $G(T_1)$ or performing the ILT on separate frequency ranges with adjusted $G(T_1)$. Second, the ILT could be performed on the specific regions of the NMR spectrum, where the $G(T_1)$ is specifically adjusted to gain the best resolution

For this example, the number of solution points was increased (Figure 4.19.); as it was shown in the previous section, this helps with obtaining better resolution and quality of the T_1 data even for very noisy data. In Figure 4.19. contour plots are shown after using a 60, 90 and 120 $G(T_1)$ points for relaxation constant estimation. Table 4.3. summarises the resulting relaxation constants with their errors. It was found that the relaxation constants do not change for the alkyl methyl group: for ester methyl group values are slightly increased and are in the range

of 0.32 to 0.38 s. T_1 for methylene groups covers an even broader range between 5.5 to 6.02 s. With increasing T_1 solution points for $G(T_1)$, more than three regions of relaxing components are observed (Figure 4.19). However, these lower intensity T_1 components are not real as the frequency regions they arise from are not symmetric around 0 ppm. These extra signals are a result due to noise and experimental artefacts.

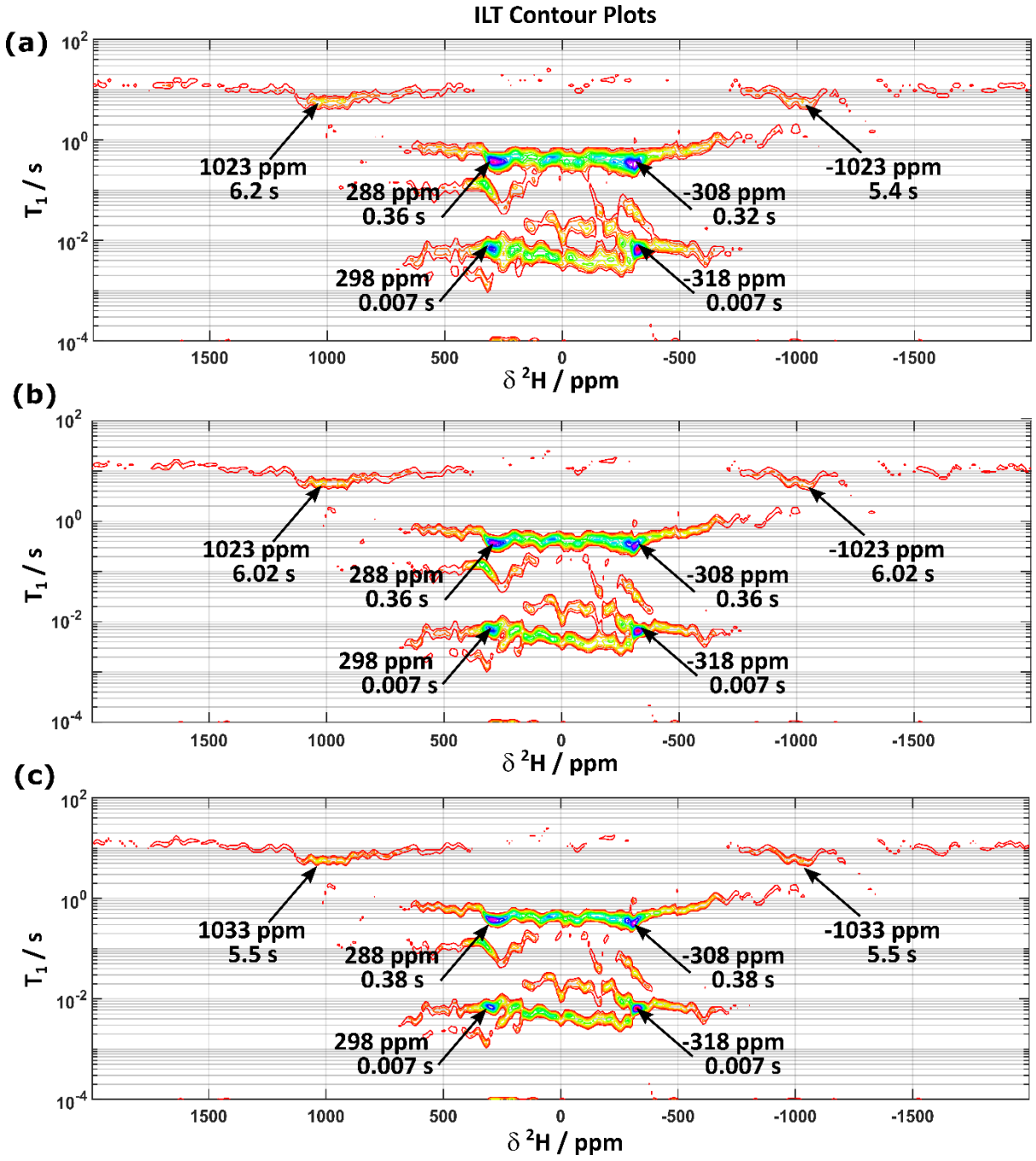

Figure 4.19. The contour plot of the relaxation time constant T_1 versus a ^2H chemical shift of PMMA-d_8. The T_1 region was estimated with (a) 60, (b) 90 and (c) 120 points exponentially spaced between 0.0001 and 100 s.

Table 4.3. The relaxation constants calculated for the PMMA-d_8 using the ILT with errors shown below. The T_1 region was estimated with 60, 90 and 120 points exponentially spaced between 0.0001 and 100 s. Errors were taken as the half-width of the signal intensity.

60 G(T_1) points		90 G(T_1) points		120 G(T_1) points	
^{2}H δ_{iso} / ppm	T_1 / s	^{22}H δ_{iso} / ppm	T_1 / s	^{2}H δ_{iso} / ppm	T_1 / s
1023	6.02 ± 1	1023	6.02 ± 1	1033	5.5 ± 0.7
298	0.007 ± 0.002	298	0.007± 0.002	298	0.007 ± 0.001
288	0.36 ± 0.11	288	0.36 ± 0.12	288	0.38 ± 0.07
-308	0.32 ± 0.12	308	0.36 ± 0.1	-308	0.34 ± 0.07
-318	0.007 ± 0.002	-328	0.007± 0.002	-318	0.007 ± 0.001
-1023	5.4 ± 1.3	-1033	6.02 ± 1.1	-1033	5.5 ± 0.8

To find out if the T_1 estimated for the methylene residue is the actual relaxation constant, the range of G(T_1) was changed to 1 – 10 seconds and the ILT was performed only on the methylene regions in the NMR spectra. The resulting contour plots are shown in Figure 4.20. For the given signal, multiple relaxation constants are observed in a range from 5.3 – 6.7 s for the positive spectral frequency range and 5.4 – 6.4 s for the negative spectral frequency. The relaxation constants with their chemical shifts are summarised in Table 4.4. The relaxation constant for the methylene group can be given as 6.3 ± 2 s where this value was taken as an average across both contour plots.

Overall, this experimental example has shown that the ILT can be successfully used to separate differently relaxing components from an NMR spectrum, as shown by the contour plot in Figure 4.18. (a) where three differently relaxing regions are separated. In case of using the conventional relaxation processing (Bruker software), the relaxation constant for the alkyl and ester methyl groups was calculated to be 0.07 s. The calculated T_1 value is not matching the ones found in literature, nor estimated by using the ILT. No T_1 was found for the signals corresponding to the methylene regions due to low signal-to-noise ratio.

The relaxation constants estimated with the ILT for the PMMA-d_8 of alkyl and ester methyl groups are in good agreement with the literature values within error estimates. However, the T_1 constant calculated for the methylene group did not match the constant reported in the literature, although, this may be because the previous work estimated this T_1 with a large error. Increasing the number of points used for the solution of the G(T_1) from 30 to 120 points increased the resolution of the contour plot in the G(T_1) dimension and reduced the error estimates. However, curving of the 2D signals in the T_1 dimension was observed, which is due to poor signal-to-noise and experimental artefacts, such as spectrometer's deadtime or DC offset,[171] and inherent anisotropy of T_1 across the ^{2}H powder patterns. Interestingly, the

estimated relaxation constants did not change significantly for alkyl and ester methyl group ^{2}H between analysis methods.

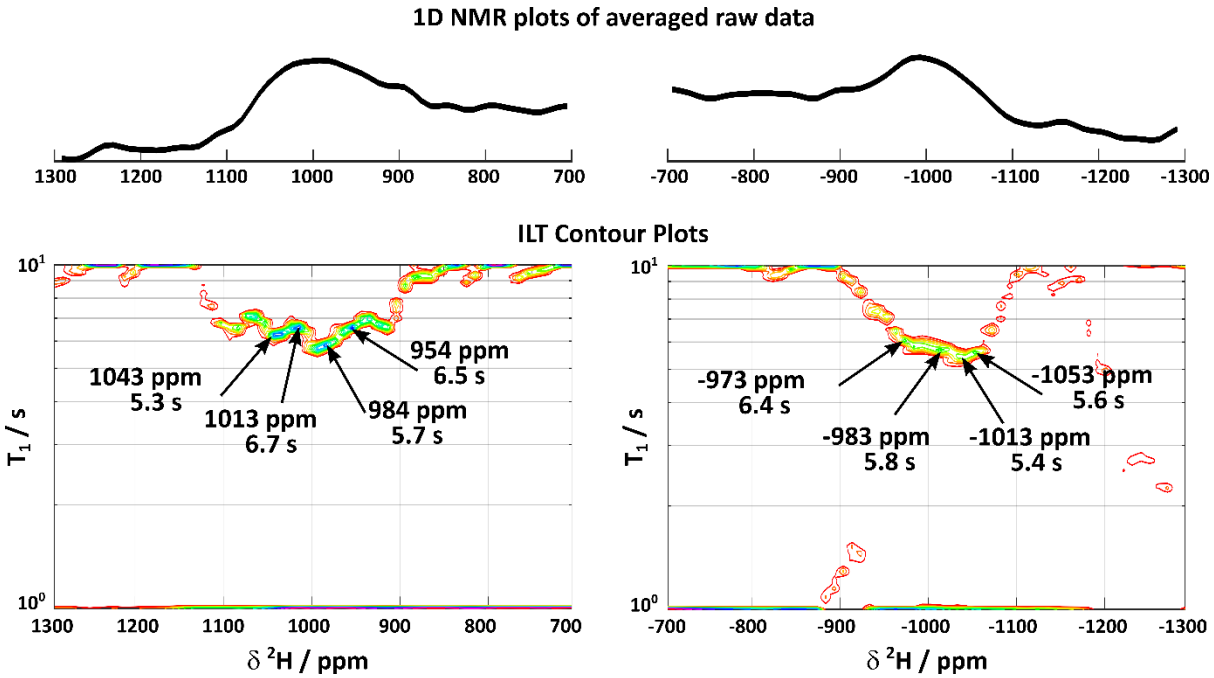

Figure 4.20. The contour plot of the relaxation time constant T_1 versus a ^{2}H chemical shift of the methylene groups in PMMA-d_8. The T_1 region was estimated with 60 points exponentially spaced between 1 and 10 s. On top of the contour plot is shown 1D ^{2}H NMR spectra after combining every four data points prior to performing the ILT.

Table 4.4. The T_1 relaxation constants extracted for the methylene regions in the PMMA-d_8 from the ILT contour plot (Figure 4.20.). The values obtained were averaged to give a final T_1 value.

^{2}H δ_{iso} / ppm	T_1 / s
1043	5.3 ± 1
1013	6.7 ± 1.5
984	5.7 ± 1.1
954	6.5 ± 1.5
-973	6.4 ± 1.5
-983	5.8 ± 1.1
-1013	5.4 ± 1
-1053	5.6 ± 1.1

5

Materials and Methods

5.1. Synthetic Model Peptides

Synthetic model peptides were prepared following procedure reported by Professor Richard Farndale.[93,173] Briefly, peptides (0.1 mmol scale) were prepared in a microwave-assisted automated peptide synthesiser (Liberty™, CEM) as C-terminal amides on a TentaGel R RAM resin (loading of 0.19 mmol/g, Rapp Polymere) following a standard Fmoc-based solid phase peptide synthesis (SPPS) strategy.[174] Crude peptides were purified using reverse phase high-performance liquid chromatography (HPLC) (PerkinElmer Life Sciences LC200) on an ACE Phenyl-300 column (250×21.2 mm, Hichrom Ltd) using a linear gradient of acetonitrile in water containing 0.1% trifluoroacetic acid. The purity of the peptides was assessed by matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometry before ionisation. Final purity was estimated to be >95%. Mass of the peptides were measured and packed into Kevlar™ inserts, which fit into a 4 mm MAS rotor. Table 5.1 below summarises

the peptides used in this work showing the peptide sequence with isotopic enrichment highlighted. Isotopically enriched Fmoc-[U-$^{13}C^{15}N$]-amino acids were purchased from Cambridge Isotope Laboratories. [U-$^{13}C^{15}N$]-isotopically enriched peptides were prepared by Dr Dominique Bihan.

Table 5.1. List of model peptides used. The first column denotes the sequence of the model peptide with the isotopically enriched amino acid highlighted in bold and marked with an asterisk. The second column represents peptide code used in the thesis text. The third column contains the molecular weight of peptide as obtained by MALDI-TOF. The final column details the mass of the peptide sample packed in the solid-state NMR rotor insert. In the given sequences G represents glycine, P – proline, O – hydroxyproline, A – alanine, R – arginine, F – phenylalanine, E – glutamic acid.

Peptide sequence	Peptide code	Molecular weight / (mol/g)	Mass of sample / mg
[U-$^{13}C^{15}N$] – enriched peptides			
$(GPO)_5$**G*P***$O(GPO)_5$-NH_2	P1	2967.2	11.0
$(GPO)^5GP$**P*G***$PO(GPO)_4$-NH_2	P2	2951.2	13.0
$(GPO)_5G$**P*G***$PO(GPO)_4$-NH_2	D1	2838.1	10.0
$(GPO)_5$**G*P***$P(GPO)_4$-NH_2	P3	2951.2	14.1
$(GPO)_5AP$**P*G***$PO(GPO)_4$-NH_2	A3	2965.2	21.7
$(GPO)_5GPP$**A*P***$O(GPO)_4$-NH_2	A4	2966.2	20.5
$(GPO)_5$**A***$PO(GPO)_5$-NH_2	A1	2976.2	8.1
$(GPO_5GFOGE$**R*G***$PO(GPO)_4$-NH_2	I1	3363.6	10.0
$(GPO)_5$**G*F***$OGER\ (GPO)_5$-NH_2	I2	3363.6	9.0

5.2. Preparation of Extracellular Matrix from in-vitro cell culture

5.2.1. Isolation of Foetal Sheep Osteoblasts (FSOB)

The parent cells of the foetal sheep osteoblasts were isolated by Mr Rakesh Rajan (Chemistry Department, University of Cambridge). The full experimental details are reported in reference.[5] Briefly, foetal sheep osteoblasts (FSOBs, 18-weeks old) were isolated from a humanely sacrificed foetus. The femurs were removed from the foetus and washed several times with 1% Trigine (Medichem International). Muscle and non-osseous tissues were stripped off to expose bone which was then sectioned into small pieces and washed with 70% ethanol followed by multiple washes with Minimum Essential Medium (MEM; Invitrogen) until all traces of ethanol were removed. Bone pieces were then transferred into flasks containing Dulbecco's Eagle Medium (DMEM, 1 g/L glucose; Invitrogen) containing bacterial collagenase A (0.5 mg/mL) and dispase II (3 mg/mL) both purchased from Roche Diagnostics. Bone strips were incubated at 37 °C in a shaking water bath for 3 h to release osteoblasts. Then, bone

sections were rinsed in DMEM containing 20% foetal calf serum (FCS, Invitrogen) to stop enzymatic digestion. Cell suspension and rinse medium were passed through a 40 µm cell strainer (Appleton Woods), then centrifuged at 1000 g for 5 min at room temperature. The cell pellet was resuspended in 25 mL DMEM Complete medium (containing 10% FCS, L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate (30 µg/mL; Sigma) and L-glutamine-penicillin-streptomycin (10 mL/L; Sigma)) and was then transferred into two T-175 cm^3 cell culture flasks (Nunc) and stored in an incubator at 37 °C in a 5% CO_2 atmosphere. All supplements were filter sterilised (0.22 µm filter, Appleton Woods) before addition. When the cell cultures were 70% confluent, the cells were washed with 10 mL of phosphor-buffered saline (PBS; Invitrogen). Then the content of the flask was treated with 10 mL of 0.25% trypsin containing 1mM EDTA (Sigma) and incubated for 5 min at room temperature until the cells became detached from the flask. Trypsin was neutralised by 10 mL DMEM complete media, 10 µL of cell medium were collected and cells counted using counting chamber (Thermo Fisher). The remaining cell suspension was transferred to a 50 mL centrifuge tube and centrifuged at 1200 rpm for 5 min. The supernatant was removed, and cells were resuspended in DMEM containing 10% dimethyl sulfoxide (DMSO, Sigma) (freezing medium). The concentration used was 1 million cells per 1 mL of freezing medium. 1 mL of cell mixture then was transferred to 2 mL cryotubes (Thermo Fisher) and placed in 'Mr Frosty™' freezing container and stored in -80 °C for 24 h. Cryovials were then stored in liquid nitrogen until further use. The average amount of cells from one T-175 flask were around 6-7 million, producing 6-7 cryotubes.

When the cells were employed for experimentation, the cryotube was removed from liquid nitrogen, thawed to room temperature, centrifuged at 1000 g for 2 min to remove DMSO and resuspended in 25 mL Complete DMEM medium (the media was changed every 48 h). When the cells reached 70% confluency, they were trypsinised as described above. Then they were either resuspended in freezing medium for further storage or resuspended in 10 mL Complete media and split into four T-175 flasks and cultured further as described previously. For cells that were used for growing matrix, they were cultured until matrix peel-off was observed.

5.2.2. Isotopic Enrichment of Extracellular Matrix

When all cell lines reached 100% confluency, isotopically enriched amino acids were added to the medium (1:1 concentration ratio). The table 5.2 below summarises all the isotopically enriched amino acids and the final concentration of the enriched amino acid in the cell culture used in this work. All isotopically enriched amino acids were purchased from Cambridge Isotope Laboratories. The stock solutions of each amino acid were prepared in

phosphor-buffered saline (PBS; Invitrogen) and sterilised passing through 0.22 µm filter. Media with appropriate isotopically-enriched amino acid were changed every 48 h for FSOB. Cell culture was maintained until the peel off the matrix was observed from the flask around 10 days for FSOB cell line.

Table 5.2. List of isotopically enriched amino acids and their concentrations used.

Isotopically enriched amino acid	Concentration (mg/L)
[U-$^{13}C^{15}N$] -Alanine	*,a
[U-$^{13}C^{15}N$] -Arginine	84
[U-$^{13}C^{15}N$] -Glycine	30
[U-$^{13}C^{15}N$] -Leucine	105
[U-$^{13}C^{15}N$] -Lysine	146
[U-$^{13}C^{15}N$] -Phenylalanine	66
[U-$^{13}C^{15}N$] -Proline	69*,2

* No Alanine and Proline were supplemented in the media used in this work; therefore, concentrations were obtained from different supplier: (a) DMEM (Life Technologies) and (b) McCoy's media (Invitrogen).

5.2.4. Extracellular Matrix Clean-up

In this work, collagen extraction methodology was developed in collaboration with Miss Rui Li and Dr Uliana Bashtanova (Department of Chemistry, University of Cambridge). The following method was developed to gain a higher quality of isotopically enriched collagen material that contains fewer contaminants (these include proteins that are not required in this study which become isotopically labelled during the isotopic enrichment procedures). The following techniques were employed to compare and determine the quality of the "cleaned-material" after each cleaning step:

(1) NMR spectroscopy – comparison of unlabelled in-vitro grown ECM material against commercially available "pure" collagen type-I.
(2) Fluorescence measurements – monitoring changes in the tryptophan ($\lambda_{ex/em}$ 280/348 nm) fluorescence emission band.
(3) Transmission electron microscopy (TEM) – providing information on the content of the overall matrix after homogenisation.
(4) Amino acid analysis by high-performance liquid chromatography mass-spectrometry (LC-MS) – amino acid composition was assessed in each step of the process and compared to the amino acid content of the commercially available "pure" collagen type-I.

During this work, a three-step clean-up method was developed. This method consisted of sequentially performing (1) a freeze-thaw procedure to lyse the cells and remove cell related proteins; (2) a detergent clean-up to dissolve and further remove membrane related proteins; and (3) enzymatic digestion of other ECM proteins and polypeptides by α-chymotrypsin leaving predominantly collagen behind. These procedures were employed for the FSOB matrix. The average mass of the lyophilised material following each procedure is given for all types of matrices obtained from one T-175 cm^3 flask.

Matrix treatment with the freeze-thaw method

This process was used to lyse the cells by disrupting them via formation of the water ice-crystals. During this process, cells swell and break releasing proteins located in the cytoplasm, that are further washed away. When matrix peel-off was observed, the matrix was washed with PBS buffer (2×10 mL, 37 °C) to remove all residues of the media, then 5 mL of PBS was added to the flask and left at -80 °C. After 24 h the flask was retrieved and allowed to thaw to the room temperature (298 K), followed by a PBS wash (2×10 mL, 37 °C) to remove all released proteins and residues from the cells, followed by water wash (2×10 mL, 37 °C) to remove phosphate containing species. The matrix was then lyophilised for 24 h, affording a dry material. This sample was then stored at -20 °C until further use. Overall, the average dry weight of the FSOB matrix from the one T-175 cm^3 flask is around 10-11 mg. Variations in the overall mass of the sample may be attributed to the passage number of the cells. It has been reported that the proliferation rate, protein expression and matrix production can be affected as the passage number increases.[175] The first type of ECM material that was used in this study was the freeze-thaw matrix.

Matrix treatment with detergent wash and DNase

In this process, Triton-X 100 was used to solubilise and remove membrane proteins, followed by DNase treatment to remove any residual DNA in the matrix material. This procedure can be used after freeze-thaw method, which further increases efficiency of the detergent treatment of the matrix. But in this work, it was used as described further. When matrix "peel-off" was observed, the medium was discarded, and the matrix was washed with PBS (2×10 mL, 37 °C). Pre-prepared extraction buffer (10 mL, comprising of 20 mM NH_4OH (Sigma), 0.5% Triton X-100 (Invitrogen) in PBS, pH 8.0) was added, and the matrix was

incubated at 37 °C for 10 min in a 5% CO_2 incubator. The matrix and the extraction buffer were transferred to a 50 mL tube and diluted by the addition of 10 mL of PBS. The mixture was left at 4 °C. After 24 h, the extraction buffer was decanted and the remaining matrix was washed PBS (2×10 mL, 37 °C) and subsequently incubated with DNase 10 µg/mL in DNase working buffer (100 mM Tris-HCl (Invitrogen), 25 mM $MgCl_2$ and 1mM $CaCl_2$) for 30 min at 37 °C in a 5% CO_2 incubator. The matrix was then washed with DNase working buffer 10 mL once, followed by washes with PBS and water (each 2×10 mL, 37 °C). The material was lyophilised for 24 h and stored at -20 °C until further use. The average dry weight of the material was around 5-6 mg from one T-175 cm^3 flask.

Enzymatic digestion of matrix by α-chymotrypsin

The enzymatic α-chymotrypsin digestion was employed to remove other extracellular matrix proteins leaving predominantly collagen behind. α-chymotrypsin preliminary hydrolyses the peptide bonds on the C-terminal side of the hydrophobic residues: tryptophan, phenylalanine, tyrosine and leucine.[176] Additionally, lower activity is observed for the C-terminal side of methionine, serine, threonine, valine, histidine, isoleucine, alanine and glycine. The rate of the cleavage is affected by the neighbouring amino acids in the peptide bond; there is no cleavage observed for neighbouring amino acids with proline.[177] Collagen triple helix is rich in proline, thus we hypothesised that most of the extracellular proteins would be removed and mainly collagen will be left in the matrix. It is plausible that small quantities of the collagen get digested, e.g. propeptides (not rich in proline), however, the cleavage will be slower compared to the other extracellular proteins.

The matrix was detergent cleaned as described in the previous section. After the DNase treatment and the subsequent washes, the matrix was then further washed with α-chymotrypsin working buffer (2×10 mL) consisting of 100 mM Tris-HCl (Invitrogen), 10 mM $CaCl_2$ (Sigma) with pH 7.8. Next, the matrix was incubated with 4 mL of 0.05 mg/mL α-chymotrypsin (Sigma) in an α-chymotrypsin working buffer for 24 h at 37 °C in 5% CO_2 incubator. Following this, the matrix was washed with PBS (2×10 mL) and with water (2×10 mL). The material was then lyophilised for 24 h and stored at -20 °C. The average dry weight of α-chymotrypsin treated matrix was between 3-4 mg.

5.2.4.1. Solid state NMR Analysis

All samples were grown and treated as described in previous sections. Samples were packed in Kevlar™ inserts, 1D ^{13}C NMR spectra were recorded at 10 kHz MAS. Other NMR parameters used as defined in Section 5.6. NMR results are summarised in Figure 5.1. A sample of lyophilised bovine Achilles tendon (Sigma) was used as received from the supplier, and the ^{13}C NMR spectrum was recorded, serving as a "reference" spectrum of the clean collagen type-I material.

In the freeze-thaw treated sample, there are strong unsaturated fatty acid signals observed at 129.7, 128.2, 30.1, 27.5, 22.9 and 14.1 ppm, corresponding to lipid bilayers from the cell membrane.[178] This result indicates that during the freeze-thaw procedure, when the cells are ruptured, PBS and water washes do not remove all cell debris. However, following detergent treatment with Triton-X-100, there is a significant decrease in unsaturated fatty acid signals. The freeze-thaw and detergent treated samples show higher signal intensities in highlighted regions (orange arrows in Figure 5.1). Where these signals correspond to the other extracellular matrix proteins.

The ^{13}C NMR spectrum of α-chymotrypsin treated matrix is in agreement with the reference spectrum (Figure 5.1.), implying the matrix sample contains a high content of collagen type-I. One notable difference between the spectra is the presence of a sharp signal at 32.5 ppm in the reference spectrum. We initially predicted that this signal could correspond to the presence of unsaturated fatty acids as the commercially available collagen type-I collagen contains unknown amounts of lipids. However, the other, stronger signals that are characteristic of lipids (see FSOB freeze-thaw sample Figure 5.1. (a)) are absent from cross-polarization spectra. It is likely that these lipids more mobile compared to the lipids in the freeze-thaw treated FSOB matrix, and thus the cross-polarization is not working efficiently on lyophilised bovine Achilles tendon material. The lipid signals are only observed under ^{13}C INEPT solid-state NMR experiment, giving strong signals corresponding to the lipid carbon atoms (Appendix A), that further suggests that these lipids are more mobile and less likely to be observed under cross-polarization experiment. We, therefore, hypothesise that this signal corresponds to Glu carboxylate sidechain Cγ carbon. The intensity of this signal is known to be reduced significantly when the sample is soaked in D_2O and freeze-dried, which suggest that this signal is not corresponding to the lipid signal. The hydration level in the tendon sample changed, further suggesting that the intensity of this signal is correlated to the water content in the sample. The hydration level in the reference sample is unknown, and it is highly plausible that the material has absorbed atmospheric water thus becoming more hydrated compared to

freshly lyophilised matrix samples. Further, if matrix treatments are considered, we infer that signal intensity change at the 32.5 ppm can be further attributed to the subtle pH changes between the compared materials. The difference in chemical shift between the acidic and conjugated base forms for the Glu (Cγ) is 3.5 ppm.[179] The matrix is subjected to detergent and α-chymotrypsin treatments which include adjustment of pH, for detergent treatment pH is at 8.0 and for α-chymotrypsin digestion at pH 7.8. The water used in this work has a pH of 5.5; therefore, the pH of the matrix is somewhere between 7.8 and 5.5 pH. Thus, the matrix could be partially in conjugated base form, where the chemical shift of Cγ should be at 35.7 ppm (not seen in the spectra due to overlap with other amino acids), and intensity of acid form has decreased. The tendon material is subjected to acid treatments using a modified method by Einbinder and Schubert.[180] From the sharp peak in the NMR spectrum, it is clear that the reference sample contains an acidic form of the Glu rather the conjugated base. Further, when the matrix material is washed with Tween 20 (0.5% in PBS, pH 8.0, Sigma) after α-chymotrypsin digestion, the signal at 32.5 ppm is lost. Therefore, we conclude that the signal at 32.5 ppm is Glu Cγ, whose intensity and chemical shift depends upon matrix material hydration level and pH.

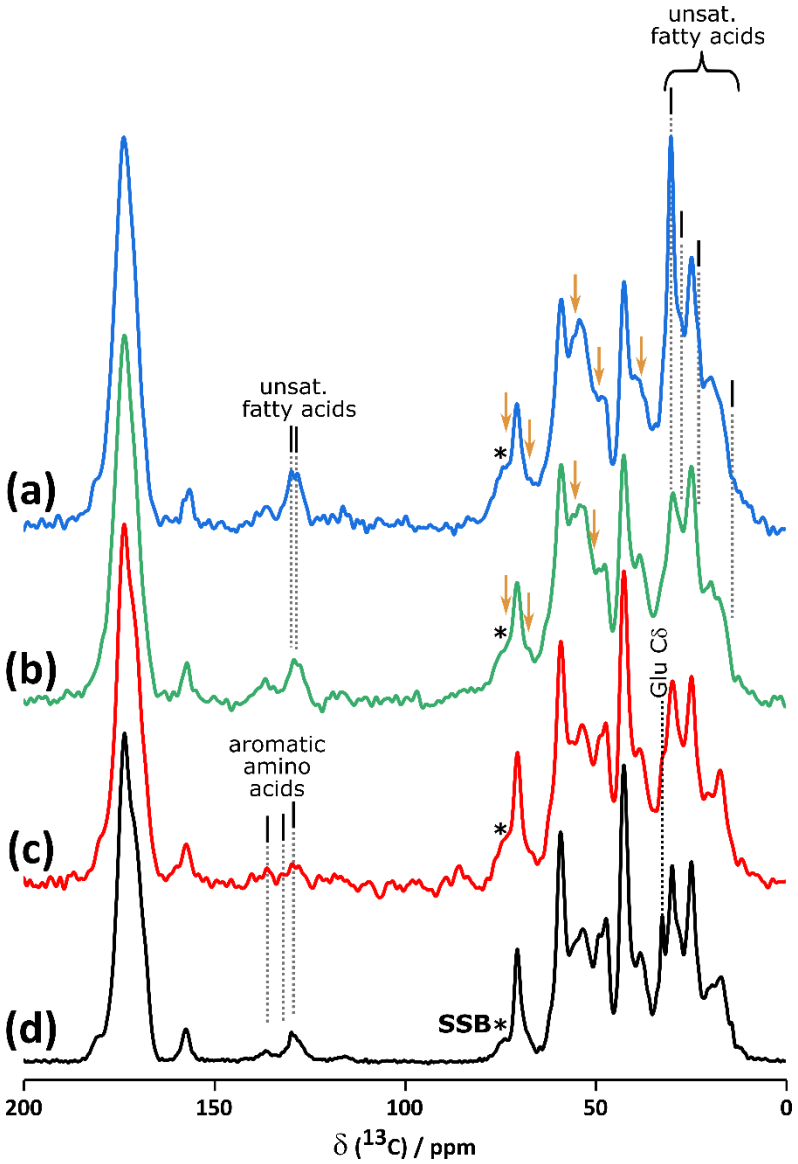

Figure 5.1. The ^{13}C 1D CP NMR spectrum of FSOB matrix after (a) freeze-thaw, (b) detergent and (c) α-chymotrypsin treatments; and (d) commercially available bovine Achilles tendon. Unsaturated fatty acids are represented with dashed lines at 129.7, 128.2, 30.1, 27.5, 22.9 and 14.1 ppm in freeze-thaw and detergent treated samples. Aromatic amino acids are highlighted with dashed lines in the α-chymotrypsin treated matrix and Achilles tendon samples. Orange arrows highlighting other extracellular proteins; where signals at around 75-80 ppm (corresponding to glycosylations) are overlapping with spinning sidebands (SSB) highlighted with an asterisk. Spectra were obtained using a 400 MHz magnet, operating at 290 K temperature. MAS frequency was 10 kHz; FSOB samples were acquired using 10 240 and tendon with 40 960 number of scans.

5.2.4.2. TEM Analysis

TEM analysis was performed to assess the content and appearance of the collagen fibres within the FSOB matrix and other cell lines after different treatments (Figure 5.2 (a)-(f)). The commercially available bovine Achilles tendon was used as a 'reference' of clean collagen type-I (Figure 5.3 (g)-(i)). Samples were processed for TEM analysis following the procedure described in section 5.5., negative staining with uranyl acetate was used.

The TEM images of the freeze-thaw and detergent treated matrix (Figure 5.2 (a)-(d)) show collagen fibrils that are surrounded by a high amount of aggregates of other proteins and cell debris. However, the characteristic banding of the collagen fibrils (Figure 5.2 (g)-(i)) is vague due to these aggregates absorbing most of the uranyl acetate stain. In the α-chymotrypsin treated sample (Figure 5.2 (e)-(i)) the amounts of these aggregates are considerably less. Furthermore, the banding of the fibrils is more pronounced; gap and overlap zones can be clearly distinguished, even sub-banding of the fibrils is observed (Figure 5.2 (f)). In Figure 5.2 (e) collagen fibrils seem to be frayed on the fibril edges. This could be either the unravelling of the collagen triple helix (1) due to the α-chymotrypsin digestion indicating that the enzyme concentration used is too high or (2) this is a result of handling the matrix sample with the ultra-sonication for too long. It is also plausible that these are proteoglycans with their GAG chains, these proteins do look like small 'brushes' as shown in this image coming off the collagen fibrils.

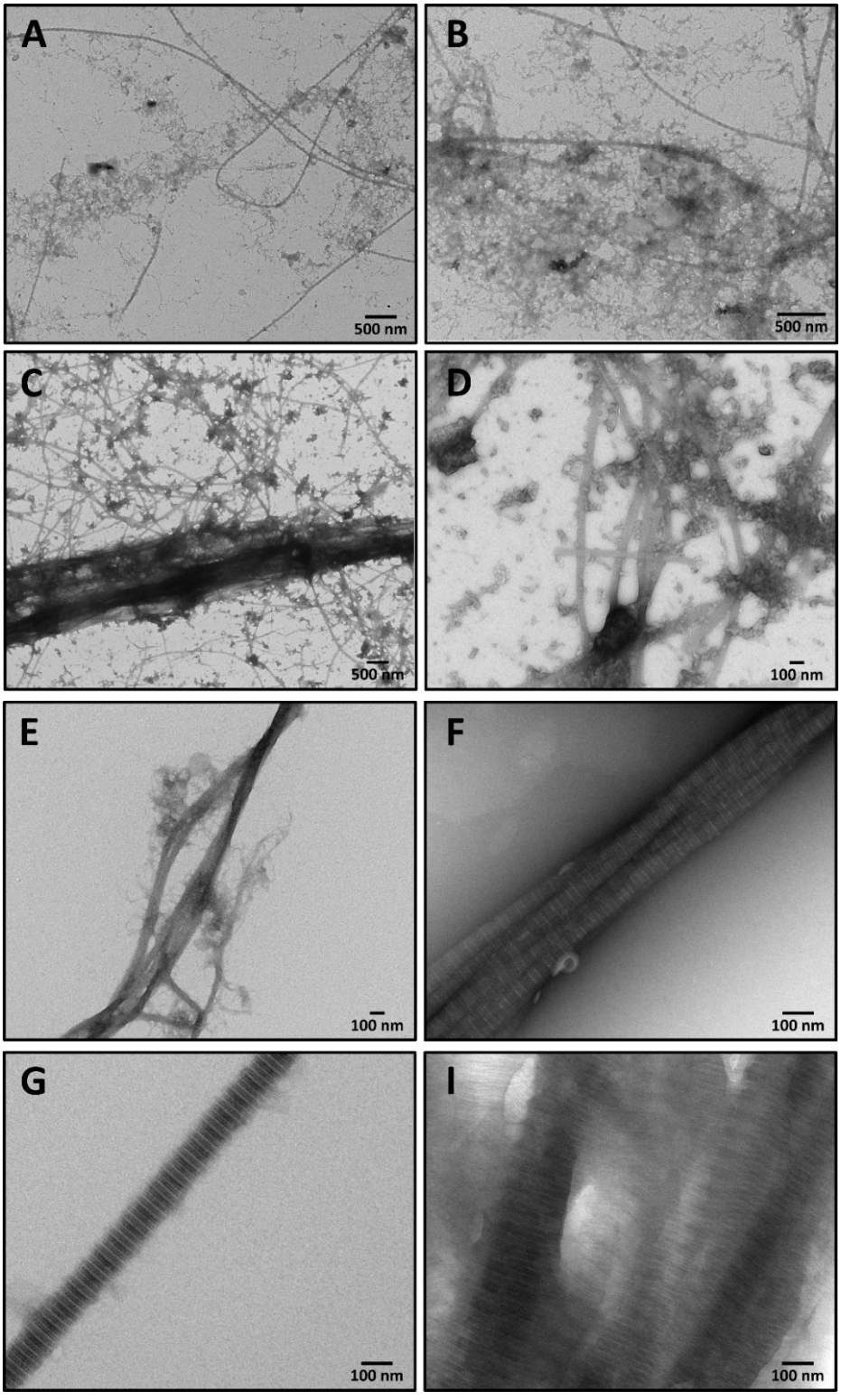

Figure 5.2. Bright-field TEM images of negatively stained FSOB samples after (a) and (b) freeze-thaw treatment, (c) and (d) detergent treatment, (e) and (f) α-chymotrypsin digestion. (g) and (i) are bovine Achilles tendon sample. The TEM images of (a) – (d) contain high amounts of dark grey, irregular chunks that are aggregates of the other smaller extracellular proteins and cell debris. The α-chymotrypsin digested matrix (e) and (f) shows collagen fibrils with their characteristic banding pattern that represents gap and overlap zones as in the 'reference' sample of bovine Achilles tendon (g) and (i).

5.2.4.3. Fluorescence Analysis

In proteins, fluorescence comes from aromatic amino acids: tryptophan, tyrosine and phenylalanine. These residues have distinct excitation and emission wavelengths and quantum yields (Table 5.3.). In the collagen type-I, only tyrosine residues will contribute to the fluorescence due to the lack of tryptophan in the native collagen sequence and low quantum yield of the phenylalanine.

Table 5.3. Fluorescence properties of tryptophan, tyrosine and phenylalanine in measured in water at neutral pH. Quantum yields are taken from reference 183.

Amino acid	λ_{ex} / nm	λ_{em} / nm	Quantum yield
Tryptophan	280	355	0.13
Tyrosine	275	304	0.14
Phenylalanine	258	283	0.02

In general, fluorescence spectra of collagen type-I is a sum of overlapping fluorescence bands. In addition to the expected tyrosine fluorescence with excitation/emission maxima at 275/305 nm, there are other possible fluorophore pairs contributing towards the fluorescence of the collagen type-I, these species are summarised in the Table 5.4. below. The intensities and ratios of these fluorophores will depend upon the collagen source (skin, tendon, etc.), age and the conformation of the collagen backbone and surroundings.

Table 5.4. Reported fluorophores in the collagen type-I with their reported excitation and emission fluorescence.

Fluorophore	λ_{ex} / nm	λ_{em} / nm	Comments	References
Tyrosine aggregate (excimer-like species)	270	360	Result of mutually interacting tyrosine residues.	[181]
3,4-dihydroxy-phenylalanine (DOPA)	285	330	Post-translational modification; consequence of collagen aging and can be formed under UV irradiation.	[182]
Dityrosine	325	400	Can be formed photochemically or enzymatically.	[183-185]
Advanced glycation end products (AGEs)	330-360	440-450	Non-enzymatic reaction products between collagen and surrounding sugars.	[186]

Fluorescence measurements were performed on a Cary Eclipse Fluorescence spectrophotometer (Agilent, Santa Clara, CA, USA) with a 10 mm path length quartz cuvette (SUPRASIL®, Hellma Analytics, Mulheim, Germany) at 295 K. The water solvent background was recorded and subtracted for all fluorescence spectra due to the Raman scattering of the water overlapping with the emission of the tyrosine. For all samples, the fluorescence emission was measured at 275 and 285 nm, no emission was observed when the sample was excited in the range of 325-360 nm. In this study, the fluorescence emission spectrum of matrix samples was measured after each treatment, as described in Section 5.2.4. Additionally, a reference sample of 'pure' collagen type-I was selected to be commercially available bovine Achilles tendon. The excitation at 275 nm gives a characteristic fluorescence of the collagen type-I, with tyrosine emission maxima at 309 nm with broadening at higher range (320 - 380 nm) where this broadening can be assigned to the excimer like species and DOPA. Further, excitation at 285 nm (targeting DOPA) shows an increase in the range of 320 - 380 nm emission. The tyrosine emission is still observed; however, the overall signal has shifted due to the decrease in tyrosine and increase of the emission of the excimer like species and DOPA. Furthermore, extra, sharp peaks are observed at the 359, 375 and 423 nm that are artefacts from scattering. Fluorescence measurements of the freeze-thaw and detergent treated matrices show strong tryptophan emission at both 275 and 285 nm excitation with maxima highlighted in Figure 5.3. This fluorescence emission is originating from the other extracellular proteins present in the matrix. A significant difference is observed upon the α-chymotrypsin digestion, where a decrease in tryptophan and increase in the tyrosine fluoresce is observed, showing that most of the extracellular proteins are removed. The emission in the range of 320 – 380 nm is weaker compared to the pure collagen type-I, suggesting either: (1) the composition of the in-vitro collagen is different, where the excimer species and DOPA might be present at lower concentrations; (2) the collagen in the matrix might be denatured slightly after α-chymotrypsin digestion, therefore the fluorescence of the free tyrosine has increased.[187] Overall, fluorescence spectroscopy of the matrix shows that it is a relatively easy and quick method to check if the α-chymotrypsin digestion has removed other extracellular proteins.

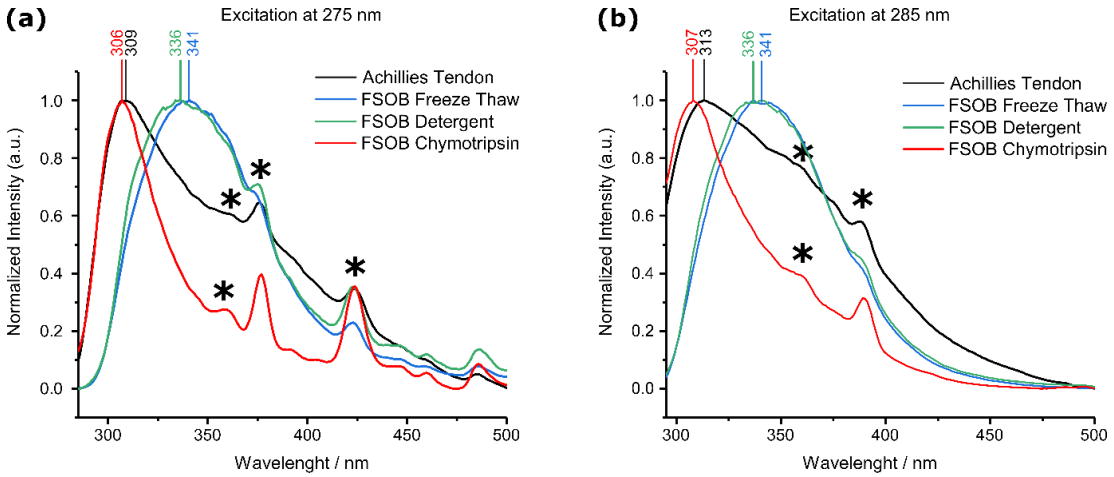

Figure 5.3. Fluorescence emission spectra of bovine Achilles tendon and FSOB matrix samples treated as described in Section 5.2.4. excited at (a) 275 nm and (b) at 285 nm. Sharp shifts at (a) and (b) 359, 375 and 423 nm are due to the light scatter and direct reflectance of light from material in the cuvette. All emission spectra are normalised to their maximum intensity to aid comparison.

5.2.4.4. Amino Acid Analysis

The amino acid analysis was performed by Dr Peter Sharratt (Department of Biochemistry, University of Cambridge). The procedure is described in reference[111] with slight modifications for handling the extracellular matrix samples. Briefly, to carefully weighted, lipolyzed sample a solution containing L-norleucine (100 µL, 250 nM in 0.1 HCl, Sigma) was added as an internal standard. The mixture was incubated for 1 h at 4 °C and centrifuged at 3000 g for 15 min. The sample was transferred to a pyrolysed tube, where the residual liquid is removed via centrifugal evaporator (SpeedVacc). Then the sample was placed in a hydrolysis vial containing a mixture of concentrated HCl and phenol (0.5 mL) together with dodecanthiol (0.68 µL, Sigma) at 4 °C. The hydrolysis vial was evacuated and flushed with argon four times. Finally, the vial was placed in an oven at 115 °C for 22 h for amino acid hydrolysis. The sample was placed in a desiccator over solid NaOH for 40 min until the acid was removed. The sample was then dissolved in sodium citrate loading buffer (pH 2.2, Sigma), centrifuged at 3000 g for 15 min and then filtered through a 0.2-µm filter. The filtrate was injected into a loading capsule placed in a Pharmacia Alpha Plus series amino acid analyser (Biochrom Ltd, Cambridge, UK). Chromatography was performed on a sodium system ion exchange resin eluting with buffers over the pH range 3.2 to 5.45. Peak detection was achieved by mixing eluate with ninhydrin at 135 °C and measuring the absorbance at 570 and 440 nm. The area of the signals was taken of each amino acid and compared with respect to the internal standard and scaled against it to find mol%. In Table 5.5. the amino acid analysis is shown for all materials used in this work; furthermore, FSOB matrix samples are compared with different cleaning methods as described in Section 5.2.4. All results are compared with respect to the bovine Achilles

tendon collagen (Sigma) that is assumed to be “pure” collagen type-I. These results show that with each step of the cleaning method, the content of the matrix becomes more collagen type-I like.

Table 5.5. The table shows the mol% of each amino, data is normalised against L- norleucine, that was used as an internal standard (250 nM). Due to hydrolysis, amino acids Asp and Asn are combined as Asn; and Glu and Gln as Glu. The mass of each material used is indicated in brackets. HO-Pro and HO-Lys are hydroxylated amino acids, Hyp and Hyl, respectively. Gly, Pro, HO-Pro and HO-Lys are highlighted and compared across all materials, these amino acids are primary components of the collagen type-I.

Residue	Achilles Tendon (0.62 mg)	FSOB		
		Freeze-thaw (0.82 mg)	Detergent (0.74 mg)	α-Chymotrypsin (0.71 mg)
Asp	5.88	9.45	8.04	6.37
Thr	1.91	4.63	3.45	2.28
Ser	3.11	4.68	3.76	2.98
Glu	10.66	12.49	12.11	10.12
Gly	**22.00**	**7.81**	**13.39**	**19.47**
Ala	8.27	5.41	6.56	7.47
Val	2.61	5.61	3.98	2.74
Met	0.09	1.91	1.61	0.99
Ile	1.58	4.20	2.98	1.73
Leu	3.42	7.88	5.71	3.70
Tyr	0.79	3.84	2.36	1.05
Phe	2.35	4.36	3.30	2.53
His	0.77	2.26	1.58	0.96
Lys	3.43	7.30	5.12	3.30
Arg	8.74	8.84	9.17	8.98
Pro	**12.22**	**6.43**	**8.94**	**11.72**
HO-Pro	**11.16**	**2.54**	**7.17**	**11.86**
HO-Lys	**1.00**	**0.37**	**0.77**	**1.76**

The chromatography of amino acid analysis shows that with each cleaning step, the FSOB matrix becomes more collagen-rich. The mol% of amino acids in the α-chymotrypsin digested sample are the closest to the tendon material, compared to the other materials. Higher content of the hydroxylated amino acids (Lys and Hyp) in the α-chymotrypsin treated sample is observed; this can be attributed to the other collagen types present in the matrix. The FSOB detergent and freeze-thaw treated sample analysis shows lower mol% of the Gly and Pro amino acids, indicating high presence of other extracellular proteins in the matrix. The amino

acid mol% differences between bovine and FSOB α-chymotrypsin digested matrix could be due the origin of the reference sample (tendon collagen) from a different animal and has originated from a different body part. Thus, the total hydroxylation content and the overall ratios of the amino acids in the collagen type-I of these materials could be due to this reason.

5.2.4.4. Conclusion on the Procedure

The material analysis using solid-state NMR, TEM, fluorescence and LC-MS show that the α-chymotrypsin digestion can give matrix that is high in the collagen type-I. Furthermore, it was found that the fluorescence can be used to assess the processed matrix in a relatively fast manner and without further sample processing that is required for the other procedures. The TEM analysis showed that after α-chymotrypsin digestion small proteoglycans are still present, furthermore, the digestion might cause slight denaturation of the collagen fibrils within the matrix. Denaturation could further be demonstrated by the fluorescence measurements where the tyrosine emission is stronger compared to the reference collagen sample, indicating the increase of the tyrosine exposure to the solvent. On the contrary, the solid-state NMR results show the α-chymotrypsin digested matrix gives strong collagen type-I signals, and there are no significant differences when compared to the reference sample. The amino acid analysis shows the Gly, Pro and Hyp amino acid ratios of the α-chymotrypsin digested matrix are similar to the reference sample; however, the amounts of the hydroxylation are different, higher in the FSOB sample. This result indicates that the fibril composition and the structure might be slightly different in the in-vivo grown extracellular matrix, where one would expect the fibrils to contain other collagen types (type V and III). This is not a surprising difference, as these materials have different origins. In conclusion, a α-chymotrypsin digestion can be employed to study isotopically enriched collagen fibrils. The solid-state NMR spectrum assignment of such treated samples can be done with confidence.

The detergent treated samples show a significant decrease of the unsaturated fatty acids NMR signals, indicating that this method is efficient for removing cell debris and its content compared to the freeze-thaw method. This method can be employed if, for example, interactions between other extracellular matrix proteins and collagen are of interest. In solid-state NMR spectra, these different proteins should be distinguished by their chemical shifts, as different secondary structures (α-helix, β-sheet and collagen triple helix) give distinctive chemical shifts.

5.3. Glycation Experiments

5.3.1. Preparation of Buffer Solution

150 mM phosphate buffer saline (PBS) $K_2HPO_4 \cdot 3H_2O$ (2.28 g, 10 mmol, Sigma) and KH_2PO_4 (1.36 g, 10 mmol, Sigma) were separately dissolved in 10 mL of deionised water (DIW). KH_2PO_4 solution was added slowly, dropwise to the $K_2HPO_4 \cdot 3H_2O$ solution until a pH of 7.4 was reached, this gave a final solution with a concentration of 1M. Finally, 15 mL of this solution was diluted to 100 mL with DIW to give 150 mM PBS.

5.3.2. Glycation with Ribose 5-phosphate of Decellularised ECM

Ribose 5-phosphate (R5P, 0.069 mg, 0.3 mmol, Sigma) was dissolved in 6 mL of 150 mM PBS buffer solution to yield 50 mM R5P solution. Decellularised ECM of FSOB (three flask product) was placed in 6 mL of 50 mM ribose solution with a pH of 7.4 and placed in shaking incubated at 37 °C (170 rpm) for 6 weeks. At the end of the glycation, ECM was washed with DIW to remove excess sugar and PBS solution. Samples were freeze-dried for 24 h, stored at -20 °C for further use. The final product was a light-yellow solid.

5.4. Paramagnetic Probe Sample Preparation

5.4.1. Preparation of Ln(III)DTPA Complexes (Ln = Pr, Eu and Dy)

Gadopentic acid ($H_2Gd(DTPA) \cdot H_2O$) as a white solid was purchased from Sigma. The remaining Ln(III)DTPA samples were prepared as described here. Appropriate $Ln(III)Cl_3 \cdot xH_2O$ (0.0013 mol, Sigma) was dissolved in DIW to yield 0.25 M solution. 5M NaOH (Sigma, ca. 5 mL) was added to the salt solution to yield a cloudy suspension with a pH of 11. Resulting $Ln(III)(OH)_3$ salt was collected and washed with DIW to reach a pH of 7. Further, solid was transferred to diethylenetriamine pentaacetic acid (DTPA, Sigma) dissolved in water (0.0013 mol, 0.05M), the mixture was neutralised using 5M NaOH, then refluxed under normal atmosphere until the solution became clear (ca. 30 min). A clear solution was collected, filtered through a 0.2 µM filter, further water was removed by freeze drying. Solid was collected and stored at 4 °C for further use. Yields were: 77%, 69% and 75% for $H_2Pr(DTPA) \cdot H_2O$,

$H_2Eu(DTPA)\cdot H_2O$ and $H_2Dy(DTPA)\cdot H_2O$, respectively. Praseodymium complex was in a light green colour, rest of Ln(III) complexes were white.

5.4.2. Treatment of ECM with Paramagnetic Complexes

Ln(III)DTPA complex and $CuCl_2$ salt (Sigma) separately were dissolved in D_2O (Sigma) to give 10 mM stock solutions. These solutions were then further diluted to provide 2 mM solutions in D_2O.

Decellularized ECM sample (two T-175 cm^3 flask product) was placed in 2 mL of prereader 2 mM paramagnetic solution. The sample was then homogenised using microtip probe sonicator (Q700 Sonicator, Qsonica) to gain a cloudy suspension (see Section 5.5. for further details of the sample preparation). The sample was centrifuged at 10 000 rpm for 5 min, then washed with D_2O (5 mL) and centrifuged again in total 2 times to remove any residual paramagnetic complexes.

5.5. Sample Preparation for Bright-field Transmission Microscopy

The commercially available collagen was reconstituted prior to the TEM analysis by incubating it in the 1.7 M acetic acid overnight, followed by homogenisation. Resulting gel was washed twice with DIW via centrifugation, then diluted to obtain 10 mg/mL collagen solution. Further, collagen solution was neutralised with 1N NaOH until pH 7 was reached allowing a gel to form. This reconstituted collagen was further used for TEM analysis (and other experiments where text refers to reconstituted collagen).

Decellularized ECM samples were homogenised using microtip probe sonicator (Q700 Sonicator, Qsonica) on ice to avoid denaturation until a cloudy suspension was obtained. Decellularized matrix from one flask (T-175) was placed in 15 mL tube with 2 mL of DIW. Pulse mode programme was used with pulse-on time of 5 seconds, and pulse-off mode with 30 seconds in total of 30 to 60 seconds of process time with amplitude of 20 – 30%. Processing time and the amplitude varied from sample to sample, with thicker samples the processing time was longer than for the much thinner or smaller amounts of the ECM samples.

10 µL of either of solutions were adsorbed onto glow-discharge 400 mesh copper/carbon film grids (EM resolutions) for 2 min. Grids were further rinsed onto two droplets of DIW; then the

grid was placed on 2% aqueous uranyl acetate solution (UA, Sigma) droplet for 30 seconds. Grid was removed from UA droplet; excess UA was removed by tapping the grid on the filter paper and left to dry. This staining method is known as negative staining. Positive staining would require sample incubation with UA, followed by DIW wash afterwards. Grids were viewed using a FEI Tecnai G^2 electron microscope run at 200 keV using 10 µm objective aperture. Images were acquired using AMT Camera software. Further image processing was done using ImageJ software.

Other TEM sample preparation to visualise paramagnetic attachment to the collagen fibrils were done similarly, instead of using UA as a staining reagent, the appropriate paramagnetic complex solution (2 mM) was used.

5.6. Solid State NMR Methods

5.6.1. Sample Preparation

Samples for solid-state NMR were freeze-dried to obtain a uniform baseline for all materials used in this study. In this work, 3.2 mm outer diameter zirconia rotors with Vespel caps were used for DNP experiments and 4 mm outer diameter zirconia rotors with Kel-F caps for the rest of the solid-state NMR experiments. Samples intended for the use with 4 mm rotors were packed in Kevlar™ inserts that fit into 4 mm rotors. These inserts have 30 µL volume; if the dry weight of the sample was more than 30 µL, then the sample was placed in rotor using Teflon seal tape to ensure stable spinning.

For 3.2 mm rotors samples were carefully sandwiched between small amounts of Teflon seal tape to ensure that the sample is positioned in the centre of the rotor, this was done to help maintain stable spinning during acquisition. Teflon seal tape consists of polytetrafluoroethylene (PTFE) and will not give ^{13}C NMR signal under cross polarization experiments as there are no ^{1}H nuclei present in the PTEF structure. ^{13}C NMR signal will be observed only in direct (pulse acquire) ^{13}C NMR experiments.

5.6.2. Sample preparation for DNP NMR experiments

For DNP experiments, AMUPol (0.958 g, 0.0015 mol) was prepared as a 38 mM solution in D_2O/H_2O (70/30 ratio) with no glycerol.[159] To the sample was added 3 - 6 µL of the radical solution, spun down using centrifuge (500 rpm, 1 min) and left incubating in the fridge at 4 °C

overnight before DNP NMR experiments. In Table 5.6. is summarised the amounts of samples and volumes of radical solution used.

Table 5.6. Mass of FSOB material and volume of AMUPol (38 mM) radical used. Isotopic enrichment is highlighted in the material with an asterisk.

Sample	Mass of sample / mg	Volume of AMUPol (38 mM) / µl
FSOB Leucine* α-Chymotrypsin	8	3
FSOB Leucine* α-Chymotrypsin + R5P	12.5	6.5
FSOB Phenylalanine* α-Chymotrypsin	8	3
FSOB Phenylalanine* α-Chymotrypsin + R5P	10	4

5.6.3. Spectrometers Used and Chemical Shift Referencing

Solution-state NMR experiments were recorded on Bruker DCH cryoprobe spectrometer at 11.25 T superconducting magnet, operating at 500 and 125 MHz for ^{1}H and ^{13}C frequencies, respectively. Experiments were recorded by the NMR service at the Chemistry Department, University of Cambridge. Samples were prepared in D_2O/H_2O solution, and spectra were referenced to TSP at 0 ppm.

Solid-state NMR experiments performed at the Department of Chemistry, University of Cambridge, were recorded using Bruker Avance I spectrometer with 9.4 T superconducting magnet, with frequencies of 400, 100, 60 and 40 MHz for ^{1}H, ^{13}C, ^{2}H and ^{15}N nuclei, respectively. Experiments were recorded using 4 mm double and triple resonance probes. Variable temperature measurements were obtained using a Bruker SmartCooler™ BCU II cooling unit. All ^{13}C NMR experiments were carried out at 10 kHz MAS spinning speed, ^{15}N at 6 kHz, ^{2}H at static or 10 kHz MAS, unless stated otherwise.

DNP NMR experiments were performed at FMP with the aid of Dr Wing Ying Chow. The experiments were performed on Bruker Avance III wide-bore spectrometer operating at 9.4 T magnetic field, with ^{1}H Larmor frequency of 400 MHz. The spectrometer is equipped with a gyrotron operating at 9.7 T and 3.2 mm ($^{1}H/^{13}C/^{15}N$) triple resonance low-temperature MAS (LTMAS) probe, with MAS rate set to 8889 Hz unless stated otherwise. DNP-enhanced experiments were carried out under 30 mA microwave radiation. The sample temperature was set to 100 K, with the actual experiment temperature of 110 K. Cooling was enabled with a variable temperature nitrogen gas flow of 2000 L/h.

Chemical shifts reported in this work were referenced against:
Glycine Cα = 43.1 ppm for ^{13}C (TMS = 0 ppm)
Glycine N = 32.3 ppm for ^{15}N (ammonia = 0 ppm)
Ethanol methyl carbon = 1.11 ppm for ^{1}H
D_2O ^{2}H = 0 ppm

5.6.4. Solid State NMR Experimental Parameters

Detailed background and pulse sequences used for the solid-state NMR experiments were described in Chapter 3. This subsection of materials and methods will cover general NMR parameters used for one and two-dimensional NMR experiments.

5.6.4.1. One-dimensional NMR experiments

^{1}H direct polarization experiment The ^{1}H 90° pulse length was set to 2.5 µs; the power level was optimised using U-^{13}C-Glycine. Delay time between acquisitions was set to 2 s. All spectra were recorded using 16 transients.

^{1}H-^{15}N/^{13}C Cross-polarization The ^{1}H 90° pulse length was set to 2.5 µs with a contact time of 2.5/3.5 ms using ramp pulse on ^{1}H and square pulse on ^{13}C/^{15}N nuclei. The delay time between scans was set to 2 s. For experiments using paramagnetic nuclei contact time was varied depending on paramagnetic species used; appropriate contact times used are given in figure captions. TPPM decoupling at 100 kHz applied on ^{1}H was used for all experiments. The number of scans was varied between samples, to yield comparable signal intensities. The appropriate number of scans given for each spectrum in the figure captions.

^{13}C-^{13}C SQ-DQ PC7 ^{1}H-^{13}C CP parameters were used as described previously, with ^{13}C 90° pulse length of 3.57 µs. At 10 kHz MAS, 70 kHz POST-C7 pulse sequence was applied on the ^{13}C channel to excite double quantum coherence in 0.4 ms, followed by another 0.4 ms of PC7 pulses to return magnetisation to zero quantum. During both PC7 pulse blocks and acquisition delay, SPINAL64 decoupling was applied at 100 kHz.

5.6.4.2. Relaxation Experiments

^{13}C/^{15}N T_1 relaxation Initial CP conditions were used as described previously. Variable delay list was introduced between 90° pulses on ^{13}C/^{15}N channel with 3.57 µs for ^{13}C, and for ^{15}N nuclei, it was calculated prior to each experiment and varied.

Variable delay list for ^{2}H NMR inversion recovery experiment with solid echo was: 0.0001, 0.0002, 0.0003, 0.004, 0.009, 0.025, 0.05, 0.12, 0.36, 0.8, 2, 3, 5, 10, 15 s.

Variable delay list for ^{13}C NMR inversion recovery experiment was: 0.01, 0.1, 0.2, 0.3, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.5, 2.75, 3, 5, 10, 15, 30, 60 s.

Variable delay list for ^{15}N NMR inversion recovery experiment was: 0.0001, 0.5, 1, 2, 5, 10, 15, 20, 30, 60, 120 s.

5.6.4.3. Two-dimensional NMR

^{13}C-^{13}C PDSD Cross-polarization condition was used as described previously. ^{13}C pulse was set to 3.57 µs. After CP PDSD pulse sequence was applied using mixing times of 20 ms. During acquisition decoupling was set to 100 kHz on ^{1}H channel.

^{1}H-^{13}C/^{15}N HETCOR with FSLG ^{1}H 90° pulse length was 2.5 µs, followed by refocusing pulses with 1.52 µs length, with further 4 composite pulses at 8.17 µs and one with 1.52 µs. Afterwards, ^{1}H-^{13}C/^{15}N CP step with 500 µs contact time was performed with ramp pulse on ^{1}H and square pulse on ^{13}C/^{15}N nuclei. Followed by auction with 100 kHz decoupling on ^{1}H nuclei.

5.6.4.4. DNP Experiments

^{1}H-^{15}N/^{13}C Cross-polarization experiment ^{1}H 90° pulse length was set to 2.75 µs. Contact time was 0.75 and 1 ms for ^{1}H-^{13}C and ^{1}H-^{15}N transfers, respectively, with a shaped pulse on ^{13}C/^{15}N and square on ^{1}H nuclei. 89 kHz TPPM decoupling was applied on the ^{1}H channel in all experiments. Recycle delay was set to 2 s; the number of acquisitions was different for each sample.

5.6.4.5. Quadrupolar Nuclei

^{2}H solid echo ^{2}H 90° pulse length was optimised using pulse acquire sequence on D_2O. 90° pulse length was 3.2 µs, with eco delays of 60 and 55 µs. The second echo was always set to be shorter to ensure that the echo maxima are recorded.

5.7. Circular Dichroism Experiments

Circular Dichroism (CD) experiments were performed on an Applied-Photophysics Chirascan qCD spectrometer using a 10 mm path-length quartz cuvette. Samples were dissolved in

deionised water to obtain concentration of 0.04 mM. The sample concentrations in all measurements were adjusted if required to maintain a high-voltage below 700 volts. The signal was collected from 300 to 210 nm at 1 nm intervals with accumulation time of 4 s per point was adopted. For equilibrium melting experiments, the CD spectra was collected over a temperature range from 6 °C to 90 °C with 2 °C intervals, equilibration for 2 mins per temperature interval was allowed. The background measurement consisting of blank solvent was collected prior to each measurement, and subsequently subtracted from the sample measurements. The resulting data was smoothed in Origin Pro 2017 using a 5 points adjacent-averaging algorithm.

6

Understanding the Effects of Mutations on Collagen Backbone Molecular Motions

Synthetic collagen triple-helical peptides have been widely used in biochemical studies as toolkits to map out where the collagen receptors and extracellular matrix components bind to collagen, and study protein-collagen interactions.[91,188] Furthermore, they have been invaluable in gaining insight into the triple helical structure, conformational variability, the stability of various sequential and pathological effects via the X-Ray crystallographic analysis,[25,96,189,190] solution-state[100,191–193] and recently solid-state NMR studies.[92,93] These peptides are usually short, ca. 30 amino acids per polypeptide chain consisting of the host sequence of interest in the middle of the triple helix flanked by GPO or GPP triplets to ensure triple helical folding and

stability. The synthetic approach enables various combinations of a native sequence to be obtained in good yields and purity. In this work, the synthetic approach permits isotopic enrichment of particular amino acids at different positions of the three polypeptide chains of the triple helix, that further enables to study the molecular motion at specific sites in a very detailed manner.

In this chapter, we aim to establish a potential scale of the molecular motions possible within the collagen triple helix backbone. As nitrogen is every third atom in the polypeptide backbones, we use ^{15}N T_1 relaxation measurements to study molecular dynamics of the collagen backbone of structurally different model peptides that represent normal and pathological collagen sequences. Additionally, we test ILT as a tool for processing the relaxation data to help resolve T_1 constants and assist in separating the broad overlapping NMR signals in the ^{15}N NMR spectrum.

6.1. Collagen-like Peptides Investigated in This Study

Peptides used in this work are listed in Table 6.1 with [U-^{13}C,^{15}N]-amino acids highlighted in bold within the sequence. These peptides were synthesised by Dr Dominique Bihan, Department of Biochemistry, University of Cambridge. The preparation procedure for these peptides is discussed in Section 5.1. These peptides are homotrimers, rather than heterotrimers as in native collagen I which has two α1 and one α2 chains. This is because it is not possible to control the heterotrimerisation process, in terms of preference for leading, middle or lagging chain for the hetero polypeptide chain, synthetically (see Figure 6.1.). The trimerisation process is spontaneous in this case, and for a heterogeneous mix of polypeptide chains, gives a mixture of different heterotrimer isomers that cannot be separated via HPLC. These peptides are synthesised with GPO repeats in the beginning and end of the sequence in order to maintain triple-helical pitch and stability. The middle of the sequence is varied so as to look at different aspects of the native sequence, namely site mutation, site deletion and an integrin binding site.

Table 6.1. Primary sequences of the peptides used in this project. [U-^{13}C,^{15}N]-isotopic enriched amino acids are highlighted in bold. The sequence length is 33 residues per polypeptide chain, except peptides I1 and I2 where 36 residues were used. The model peptide code is used to refer to the peptides in the text. The mass of each peptide shown indicates the overall mass packed in the MAS insert used for solid-state NMR experiments. The melting temperature, T_m, of each peptide in °C is measured in distilled water using CD spectroscopy. Experimental details of the peptide synthesis and CD spectroscopy are summarised in Section 5.7.

Model peptide	Sequence	Mass / mg	T_m / °C	Comments
P1	$(GPO)_5$**GP**O$(GPO)_5$-NH_2	11.0	78	
P2	$(GPO)_5$GP**PG**PO$(GPO)_4$-NH_2	9.0	78	
P3	$(GPO)_5$**GP**P$(GPO)_5$-NH_2	14.1	76	
P4	$(GPO)_5$**G**PO$(GPO)_5$-NH_2	20.0	80	^{15}N – enrichment only / used for CD analysis
P5	$(GPP)_5$GP**P**$(GPP)_5$-NH_2	10.1	51	^{15}N – enrichment only / used for CD and ^{1}H linewidth analysis
A1	$(GPO)_5$**A**PO$(GPO)_5$-NH_2	8.1	50	
A2	$(GPO)_5$**AP**O$(GPO)_5$-NH_2	10.0	48	
A3	$(GPO)_5$AP**PG**PO$(GPO)_4$-NH_2	21.7	48	
A4	$(GPO)_5$GPP**AP**O$(GPO)_4$-NH_2	20.5	48	
I1	$(GPO)_5$**GF**OGER $(GPO)_5$-NH_2	8.1	66	
I2	$(GPO)_5$GFOGE**RG**PO$(GPO)_4$-NH_2	10.0	66	
D1	$(GPO)_4$G**PG**PO$(GPO)_4$-NH_2	10.1	34	

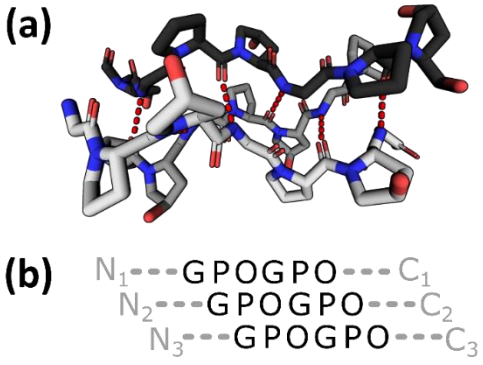

Figure 6.1. (a) Model peptide sequence of $(GPO)_2$ visualised using PyMol,[26] showing staggered leading (white), middle (light grey) and lagging (black) polypeptide chains; derived from PDB code 1V4F.[25] Red dashed lines indicate interchain hydrogen-bonding shown between Gly N-H and opposite chain Pro C=O. (b) Letter representation of the $(GPO)_2$ sequence showing all three polypeptide chains out of registry with leading, middle and lagging strands (from top to bottom).

The four groups of peptides used in this study are: imino rich amino acids (designated P-type peptides here), glycine to alanine mutation (designated A-type peptides), single amino acid deletion (designated D-type peptide) and sequences containing an integrin-binding site (designated I-type peptides). The imino rich P-type peptides contain only glycine, proline and hydroxyproline residues; the GPO triplet contributes about 10% of the GXY triplets in the native sequence.[194] Furthermore, the positions of these triplets are highly conserved in the native collagen type-I sequences between animal species[92] (see Chapter 2.5.4.1, Figure 2.11). The A-type peptides contain a glycine to alanine mutation in the middle of the sequence giving a triplet of APO or APP. These types of mutations in the native sequence are known to be pathological, causing diseases such as osteogenesis imperfecta (OI), Ehlers-Danlos syndrome IV and chondrodysplasia.[39–42] Such mutations also alter collagen fibril formation and their stability, further affecting the primary ligand binding sites with other proteins, such as integrins and metalloproteases (MMPs).[38,97] Peptides with one residue deletion, used in this work, contain an interruption site where hydroxyproline is removed from one of the GPO triplets giving a Hyp^- peptide. Such deletions in X or Y positions in the native collagen sequence are frequent for non-fibrillar collagens, like type-VIII and type-IV, giving a rise to a specific structural features and are required to maintain higher-order structures and proper function.[102–105] However, deletions in the native sequence of collagen type-I are always pathological.[108] For instance, they can lead to impaired triple helix formation, reduced extracellular secretion and α-chains being post-translationally over modified, resulting in several different types of OI.[108] The final group of peptides consist of a specific amino acid sequence (GFOGER) that is recognised by the integrin $\alpha2\beta1$ I domain, found in the $\alpha1$(I) chain of type-I collagen.[87,111]

These four groups of peptides represent various structural conformations available within the collagen triple helix. We hypothesised that the combination of [U-^{13}C,^{15}N]-isotopic enrichment of specific positions and solid-state NMR ^{15}N T_1 relaxation measurements can give a better understanding of molecular motions at nano- to picosecond time-scales and a scale of possible molecular backbone motions within the collagen triple helix, and we explore that hypothesis in the following sections.

6.2. ^{15}N T_1 Relaxation of Synthetic Collagen Peptides

In general, the primary relaxation mechanism operating in collagen and the synthetic model peptides for all nuclei is the dipolar interactions between all NMR active nuclei and in some cases, the shielding anisotropy. Molecular motions cause orientation and magnitude changes

in both dipolar and shielding interactions; these fluctuating local magnetic fields will lead to T_1 relaxation (as discussed in Chapter 4.). Only local motions that are around the Larmor frequency of the ^{15}N nuclei will be effective in inducing T_1 relaxation, that is around 40 MHz. Thus, the ^{15}N T_1 relaxation is sensitive towards fast molecular motions with correlation times of ns to ps.

The ^{15}N relaxation will be affected by the nuclear interactions between ^{15}N-^{15}N, ^{15}N-^{13}C and ^{15}N-^{1}H nuclei, and ^{15}N CSA. The dipolar interactions, as discussed in Section 4.3. are distance-dependent. In the given peptides, the amide to amide ^{15}N dipolar coupling is relatively small, below 45 Hz. Dipolar interaction between heteronuclear nuclei of ^{15}N-^{13}C and ^{15}N-^{1}H is between 0.3 – 1.5 and 9 – 11 kHz, respectively. The dipolar couplings for peptides used in this work were calculated using Equation (3.4). CSA for amide ^{15}N nuclei are usually around 4.0 – 4.5 kHz (at a spectrometer frequency of 400 MHz).[165]

In this section, measurements of ^{15}N T_1 for all the peptides in Table 6.1 are presented. After performing line broadening, zero filling, phase and baseline correction, the 2D experimental data sets were processed by an Inverse Laplace transform (ILT) along the relaxation delay dimension resulting in two-dimensional spectra correlating ^{15}N chemical shift and T_1 relaxation time as described in Section 4.4. Each 1D ^{15}N chemical shift spectrum prior to ILT was processed by combining points across the frequency range of the 1D NMR spectra to give approximately 1 point per 1 ppm to increase the signal-to-noise (S/N) ratio (as described in Section 4.4.). The $G(T_1)$ solutions were evaluated using 55 T_1 points distributed between 10^0 and 10^3 s. The errors of the relaxation constants were measured as the half-width of each T_1 signal in the T_1 dimension of the 2D chemical shift – ^{15}N T_1 relaxation time correlation spectra. Relaxation measurements were carried out at temperatures of 250, 270, 290, 300 and 310 K.

6.2.1. Imino Rich Peptides

The first group of the peptides represent imino-rich triplets within the native sequence, peptides P1-P3. The peptide P1 is a simple GPO repeat, with isotopic enrichment of Gly and Pro in the middle of the sequence, as highlighted in Table 6.1. This enrichment will show environments and dynamics of a very well-ordered triple helix. Peptides P2 and P3 are identical with hydroxyproline in the eighteen-position substituted by Pro in all three polypeptide chains, giving a GPP insertion in the middle of the sequence. This change was introduced partly due to challenges associated with obtaining isotopic enrichment of Hyp. X-Ray diffraction structures of GPO and GPP-rich synthetic model collagen peptides suggest that

changing a single O for P should have at most only a very small effect on the triple helix conformation (PDB: 1V4F[25] and 1A3J[195]).

The two-dimensional ^{15}N chemical shift – T_1 relaxation time correlation contour plots of the P-type peptides were obtained for all imino rich peptides that were further used to extract the relaxation constants. In Figure 6.2. (a) only plots for measurements performed at 290 K are shown as an example; the extracted slices from the ILT contour plot at the highlighted chemical shifts are shown in Figure 6.2. (b).

In the case of the proline ^{15}N relaxation, it was not possible to record spectra at sufficiently long τ delays to obtain full decay of the signal (Appendix B). In all experiments (at all temperatures) the signal intensity of the proline ^{15}N decayed to half the initial intensity, and glycine ^{15}N to almost zero. The longest relaxation delay used in the measurements was 120 s. If there are components that have significantly longer relaxation constants than 120 s, these will not appear in the T_1 relaxation spectra. Similarly, the T_1 spectra will not show components that have faster relaxation than the shortest τ delay used, which was 0.0001s; however, that would imply much faster relaxation than is expected, so this is not a big restriction. The ^{1}H relaxation in these samples is of the order of tenths of seconds, so a recycle delay of 2 s was used in all experiments between scans (i.e. 5 times larger than the proton T_1 relaxation).

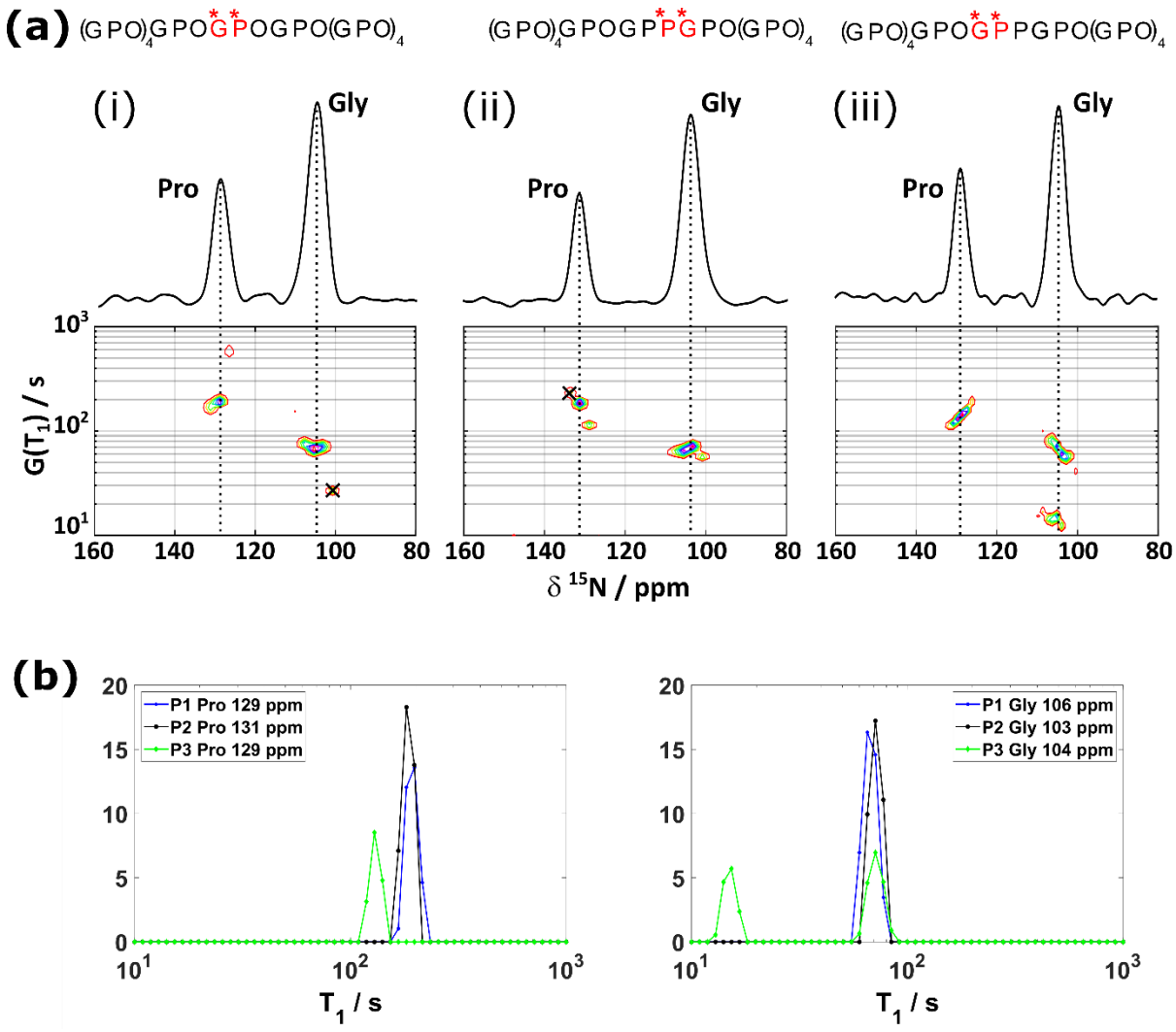

Figure 6.2. (a) Contour plots of relaxation time constant T_1 versus ^{15}N chemical shift of peptides (i) P1, (ii) P2 and (iii) P3 at 290 K. Above each contour plot, 1D ^{15}N NMR spectra of the corresponding peptide are given. Signals that are not corresponding to the chemical shifts for the collagen triple helix are crossed out and are not considered for ^{15}N T_1 analysis. (b) Slice from the contour plot extracted from highlighted cross-section with the dashes in (a); proline (left) and glycine signals (right) compared.

The contour plots show that there is a distribution of T_1 values across each ^{15}N signal, that looks like broadening and stretching of the 2D signals. The broadening in the T_1 dimension can be a result of a sub-optimal S/N ratio in the ^{15}N chemical shift spectra or it can reflect a real distribution of the T_1 values, i.e. T_1 depends on the isotropic ^{15}N chemical shift. The distribution of the relaxation constant values is highly plausible because these samples are lyophilised, and they will show a distribution of the ^{15}N chemical shifts; i.e. line-broadening of the ^{15}N signals for glycine and proline in the chemical shift dimension likely reflect slightly different molecular conformations which may in principle be associated with different relaxation times. However, it is not possible to conclude from the ILT contour plots whether the T_1 distributions are due to heterogeneity or an artefact from poor S/N ratio. Furthermore, the artefacts could arise due to the proton spin diffusion that affects the ^{15}N signal intensity

from the initial NMR experiment (Section 6.3.1.). Thus, the weighted average of the resulting distributions of the ^{15}N T_1 relaxation constants of each separate chemical shift signal (per residue) from the ILT contour plot were calculated and summarised in Table 6.2.

In the contour plots shown in Figure 6.2., there are signals observed with chemical shifts at 100 and 134 ppm, which cannot be assigned to glycine or proline residues being in a triple helical structure, respectively. These signals are crossed out in the contour plots and are not considered for the relaxation of the triple helical peptides. These signals could be either artefacts due to poor S/N ratio, or denatured polypeptide chains. Furthermore, the ^{15}N T_1 relaxation experiments were performed three times in total, and only consistent signals were reported and analysed.

Table 6.2. The table below shows ^{15}N T_1 relaxation constants estimated at temperatures of 250, 270, 290, 300 and 310 K for glycine and proline residues in P-type peptides after ILT processing. Errors of T_1 are taken as a half-width of the signal intensity for the ILT calculated relaxation constants. At some temperatures, the nitrogen nuclei in the glycine residues show two separate signals with significant intensities.

P1: $(GPO)_5$**GP**$O(GPO)_5$
P2: $(GPO)_5GP$**PG**$PO(GPO)_4$
P3: $(GPO)_5$**GP**$PGPO(GPO)_4$

Peptide and Residue		δ_{iso} / ppm	250 K	270 K	290 K	300 K	310 K
P1	**Pro**	129	190 ± 17	167± 16	182 ± 34	124 ± 14	87 ± 8
	Gly	106	95 ± 11	81 ± 9	71 ± 10	47 ± 6	43 ± 7
P2	**Pro**	131	391 ± 83	190± 18	180 ± 18	198 ± 18	167± 16
	Gly	105	-	-	-	-	100 ± 11
		104	74 ± 8	43 ± 5	68 ± 9	67.9 ± 10	14 ± 2
P3	**Pro**	129	157 ± 30	130 ± 11	136 ± 11	121 ± 14	100.0 ± 9.0
	Gly	106	19 ± 3	-	15 ± 2	-	-
		105	68 ± 21	42 ± 9	65 ± 10	51 ± 8	44 ± 9
		103	-	-	-	35 ± 5	-

The most apparent feature from the relaxation data of imino-rich peptides is that the T_1 constants are very different for the nitrogen nuclei in proline and glycine residues, with glycine showing faster relaxation compared to proline. This is expected due to the protonation on the glycine nitrogen that introduces much stronger ^{1}H-^{15}N heteronuclear dipolar coupling that helps to drive the relaxation. Furthermore, a bimodal distribution of the relaxation time constants is observed for glycine residues in peptides P2 and P3. However, this behaviour is not consistent for all temperatures.

Comparing the T_1 constants across all P-peptide samples there are constant patterns observed. Overall, ^{15}N T_1 values decrease with increasing temperature. However, there are outliers observed for glycine nitrogen in peptides P2 and P3 between temperatures of 250 and

290 K. The T_1 value is significantly lower at 270 K than at 250K and 290K, around the freezing temperature of water. The same change in T_1 at ca. 270 K is also observed for the proline nitrogen in the peptide P1. These abrupt changes in the T_1 around the water freezing temperatures will be addressed further in Section 6.5. Significant differences of T_1 values between peptides P1 and P3 are observed at low temperatures, below 290 K. In these two peptides the isotopic enrichment is the same, however, peptide P3 contains a GPP insertion. The difference in the relaxation constants between P1 and P3 suggests that there might be a difference in the molecular motion for the two peptides. They are expected to have highly similar molecular structures and therefore, nuclear spin interaction strengths for their respective ^{15}N. The key differences in molecular motion or possibly the structure of water around the peptides can account for differences in ^{15}N T_1. This will be discussed more fully below.

6.2.2. Peptides with Gly → Ala Mutation

The second group of peptides contain a single amino acid mutation from Gly to Ala in the middle of the sequence, giving a triplet of APO or APP. Gly mutation to Ala causes structural changes in collagen triple helix structure as shown by the crystal structure of this type of peptide (see Section 2.5.4.2.).The Ala side-chains are pointing in to the core of the triple helix resulting in all three polypeptide chains being pushed apart. Therefore, we expect this mutation to have an impact on local conformations and molecular motions. To assess the change in molecular structure and motions, four peptides with different enrichment combinations were selected. Peptides A1 and A2 contain an APO insertion with isotopic enrichment of [U-^{13}C,^{15}N]-alanine residues in A1, and Ala and the following Pro in A2. Peptide A3 contains an APP insertion, where [U-^{13}C,^{15}N]-Proline in the Y position of APP triplet and the [U-^{13}C,^{15}N]-Glycine of the next triplet are enriched. Peptide A4 has an GPPAPO insertion where Ala and the following Pro are labelled. The ^{15}N chemical shifts of the Pro and Ala are close to one another resulting in a signal overlap (Figure 6.3.), thus we tested ILT on the ^{15}N relaxation data to see if overlapping resonances can be separated by their T_1, and furthermore to examine what structural and dynamic information can be extracted.

The resulting ILT contour plots for A-type peptides at temperature of 290K are shown in Figure 6.3. The plots of peptides A2 and A4 show two distinct T_1 signals correlated with the Ala/ Pro ^{15}N signal which are assigned to the relaxation constants for the Pro and Ala residue ^{15}N. The separation of T_1 signals for Ala and Pro ^{15}N was observed through all temperatures at which data was collected. The ^{15}N T_1 relaxation constants were estimated as described in

previous section by taking an average in case of the signal broadening. The T_1 constants obtained as a function of temperature are summarised in Table 6.3.

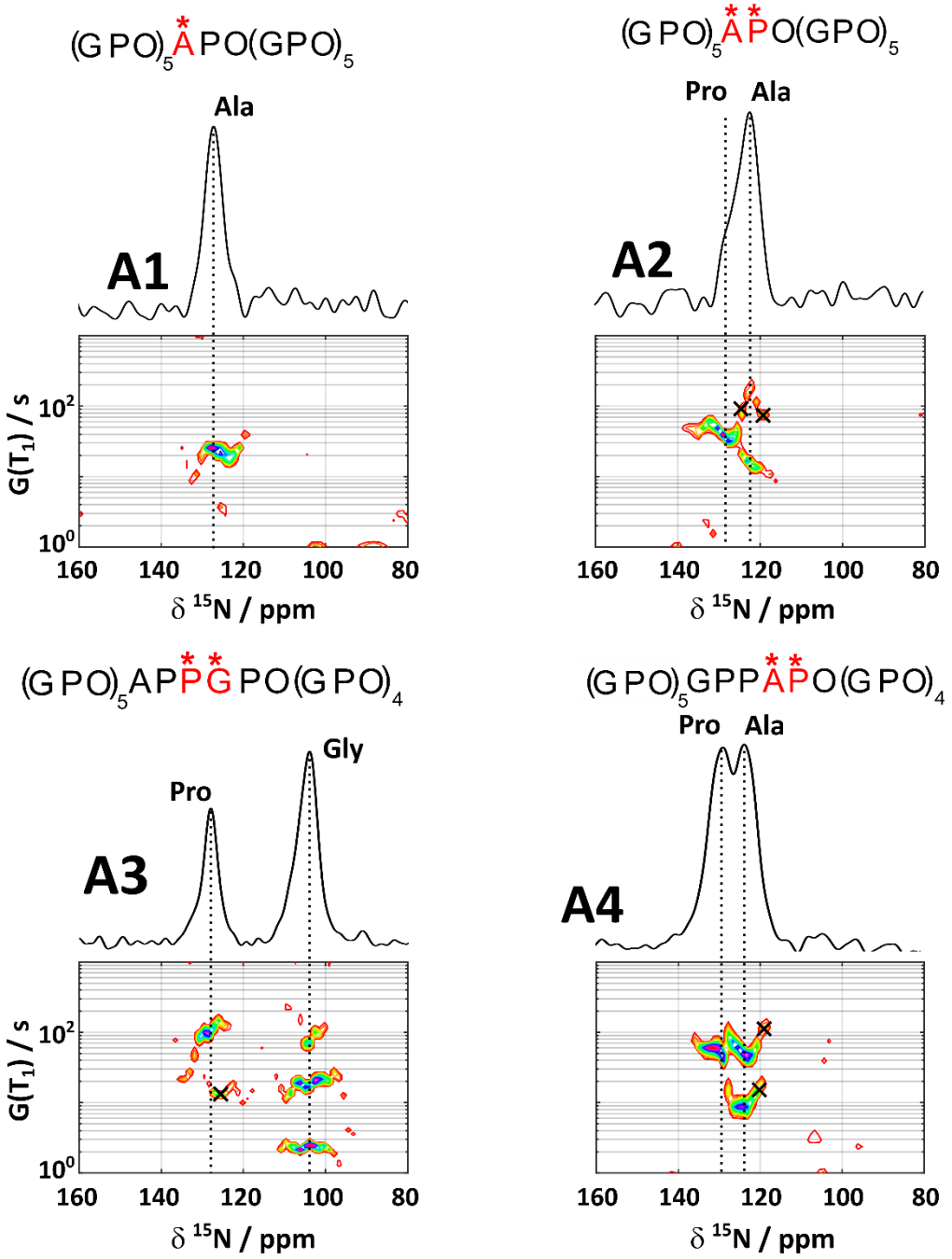

Figure 6.3. Contour plots of relaxation time constant T_1 versus ^{15}N chemical shift of peptides A1, A2, A3 and A4 at 290 K. Above each contour plot, 1D ^{15}N NMR spectra of the corresponding peptide are given. Signals that are not corresponding to the chemical shifts for the collagen triple helix are crossed out and are not considered for ^{15}N T_1 analysis. The average ^{15}N chemical shifts for the Ala and Pro residues are 123 and 135 ppm, respectively, according to the BMRB database.[196] These chemical shifts are an average of mainly α-helix and β-sheet secondary structures, not a triple helix.

Table 6.3. The table below shows ^{15}N T_1 relaxation constants estimated at temperatures of 250, 270, 290, 300 and 310 K for alanine, glycine and proline residues in A-type peptides after ILT processing. Errors of T_1 are taken as a half-width of the signal intensity for the ILT calculated relaxation constants.

A1: $(GPO)_5$**A**$PO(GPO)_5$
A2: $(GPO)_5$**AP**$O(GPO)_5$
A3: $(GPO)_5AP$**PG**$PO(GPO)_4$
A4: $(GPO)_5GPP$**AP**$O(GPO)_4$

Peptide and Residue		δ_{iso} / ppm	250 K	270 K	290 K	300 K	310 K
A1	**Ala**	124	32 ± 9	25 ± 8	23 ± 8	21 ± 6	16 ± 5
A2	**Pro**	128	60 ± 14	49 ± 13	40 ± 12	32 ± 9	48 ± 11
	Ala	124	32 ± 7	39 ± 9	17 ± 4	26 ± 8	26 ± 6
A3	**Pro**	130	116 ± 18	93 ± 16	93 ± 14	93 ± 14	62 ± 10
	Gly	106	-	-	9	-	-
		104	44 ± 8	38 ± 9	48 ± 11	32 ± 7	11 ± 5
A4	**Pro**	129	84 ± 15	48 ± 7	60 ± 11	60 ± 13	60 ± 12
	Ala	124	13 ± 3	32 ± 8	9 ± 2	12 ± 4	35 ± 7
		124	-	-	60 ± 12	-	-

Only peptides A1 and A2 are identical in terms of their polypeptide sequence; however, the relaxation constants for the Ala nitrogen nuclei are not the same. Furthermore, the ^{15}N T_1 values of Ala in the peptide A2 fluctuate with respect to the temperature, where increase at 270 K is observed. The same behaviour of ^{15}N T_1 with temperature is observed for the Ala nitrogen in the peptide A4, where Pro nitrogen at the same temperature has a lower T_1 value compared to the 250 and 290 K temperatures. Proline nitrogen relaxation constants in peptides A2 – A4 are lower in general compared to the imino rich peptides. Furthermore, the change of the T_1 across the temperature range is smaller for the A-type peptides compared to the P-type peptides, especially for the proline residues.

The decrease in the overall observed T_1 relaxation constants of all ^{15}N nuclei in A-type peptides could be related to the Ala methyl group rotations that introduce a rapid change in local fluctuating magnetic fields felt by the nuclei nearby. These methyl groups in the proteins are known to sometimes act as so-called 'relaxation sinks', where the spin diffusion to more rapidly relaxing nuclei can be a dominant mechanism of the T_1 relaxation for the less mobile groups. Relaxation sinks in these samples could potentially homogenise the measured T_1 relaxation constants, resulting in the T_1 constants that are closely similar for the isotopically enriched residues in the peptides in which Ala is ^{13}C labelled. It would require extra NMR experiments to determine the actual effect of the methyl group rotations on the T_1 constants in these samples; it would require synthesising peptides with a deuterated Ala methyl group and repeating the measurements and comparing the results.

6.2.3. Peptide with One-Residue Deletion

Peptide D1 contains an interupted sequence, where the Hyp residue is removed from the fouth GPO triplet giving a peptide with sequence of $(GPO)_3GPGPO(GPO)_4$ (or Hyp⁻ peptide). [U-^{13}C,^{15}N]-Proline was introduced in the predeletion and U-[^{13}C,^{15}N]-Glycine postdeletion of the Hyp residue.

The relaxation measurements were not performed above 310 K so as to avoid denaturation of the peptide (see Table 6.1. T_m of peptides). The ^{15}N chemical shift-T_1 correlation contour plot obtained after relaxation data recorded at 290 K processed using ILT is shown in Figure 6.4. The relaxation constants across the temperature range are summarised for both nitrogens of the Pro and Gly in table 6.4. The resulting T_1 values of the same isotopically enriched residues are overall low compared to the imino-rich peptides and the Ala mutated peptides. ^{15}N T_1 relaxation constants for Pro and Gly nitrogens do not change significantly across the temperature range for which measurements were made, but there are fluctuations observed in the T_1 of the Gly nitrogen across the temperature range.

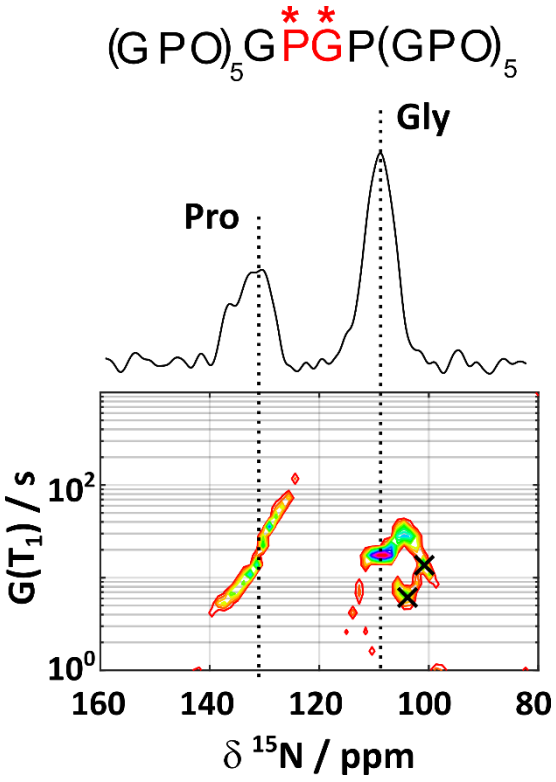

Figure 6.4. Contour plot of relaxation time constant T_1 versus ^{15}N chemical shift of peptide D1 at 290 K. Above the contour plot, 1D ^{15}N NMR spectra of the corresponding peptide is shown. Signals that are not corresponding to the chemical shifts for the collagen triple helix are crossed out and are not considered for ^{15}N T_1 analysis.

Table 6.4. The table below shows ^{15}N T_1 relaxation constants estimated at temperatures of 250, 270, 290, 300 and 310 K for glycine and proline residues in D1 peptide after ILT processing. Errors of T_1 are taken as a half-width of the signal intensity for the ILT calculated relaxation constants.

D1: $(GPO)_5G\mathbf{PG}PO(GPO)_5$

Peptide and Residue		δ_{iso} / ppm	250 K	270 K	290 K	300 K
D1	Pro	131	28 ± 7	28 ± 6	24 ± 6	28 ± 7
	Gly	108	16 ± 3	14 ± 5	23 ± 7	10 ± 2

6.2.4. Peptides Containing Integrin Binding Site

The last group of peptides studied contain a recognition site for the integrin $\alpha 2\beta 1$ binding site. The peptides I1 and I2 are identical except for their isotopic labelling pattern. The integrin binding site is incorporated in the middle of the polypeptide sequence of $(GPO)_5GFOGER(GPO)_5$, where GPO triplets are at each end are definded as N- and C-terminal zones, and the GFOGER sequence as a central zone of this peptide.

Contour plots are shown in Figure 6.5 for the measurements carried out at 290 K. The relaxation constants derived from them (weighted signal averages) are summarised in Table 6.5. In general, ^{15}N T_1 relaxation constants are higher for the peptide I1. Additionally, the relaxation constants are estimated from the data for the Arg sidechain nitrogens, which shows them to have to have very short T_1 constants compared to the backbone nitrogen.

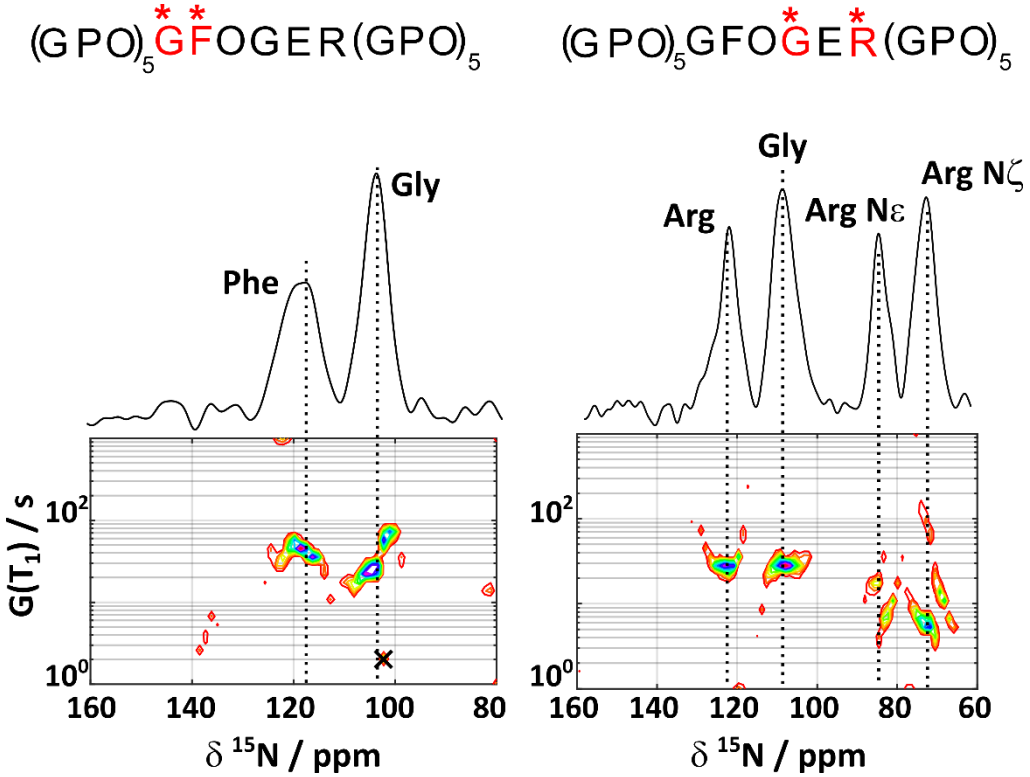

Figure 6.5. Contour plots of relaxation time constant T_1 versus ^{15}N chemical shift of peptide I1 and I2 at 290 K. Above each contour plot, 1D ^{15}N NMR spectra of the corresponding peptide are shown. Signals that are not corresponding to the chemical shifts for the collagen triple helix are crossed out and are not considered for ^{15}N T_1 analysis.

Table 6.5. The table below shows ^{15}N T_1 relaxation constants estimated at temperatures of 250, 270, 290, 300 and 310 K for phenylalanine, glycine and arginine residues in I-type peptides after ILT processing. Errors of T_1 are taken as a half-width of the signal intensity for the ILT calculated relaxation constants.

I1: $(GPO)_5$**GF**$OGER(GPO)_5$
I2: $(GPO)_5GFO$**G**E**R**$(GPO)_5$

Peptide and Residue		δ_{iso} / ppm	250 K	270 K	290 K	300 K	310 K
I1	**Phe**	118	117 ± 15	45 ± 10	46 ± 11	46 ± 10	49 ± 12
	Gly	104	73 ± 14	65 ± 13	25 ± 5	46 ± 9	36 ± 8
I2	**Arg N**	123	51 ± 11	32 ± 8	28 ± 5	25 ± 4	20 ± 3
	Gly	109	36 ± 7	32 ±	28 ± 7	36 ± 8	32 ± 6
	Arg N_ε	83	29 ± 5	10 ± 2	7 ± 1	7 ± 2	5 ± 1
	Arg N_ζ	72	16 ± 3	11 ± 3	6 ± 1	10 ± 2	5 ± 1

6.3. Mechanisms Perturbing the Analysis of ^{15}N T_1 Relaxation

As it was discussed in previous section, the main interactions affecting the ^{15}N relaxation in these peptides will be due to the dipolar interactions between nitrogen and its bound proton and CSA. Additionally, there are other interactions in the NMR that can further influence the interpretation of the relaxation constants, and one needs to be aware of.

6.3.1. ^{15}N Proton Spin Diffusion

Peptides used in this work are fully protonated; thus, the ^{15}N proton spin diffusion (PSD) can contribute towards the nitrogen-15 magnetisation decay. Spin diffusion is a process in which magnetisation is exchanged spontaneously between spins driven by the dipolar interaction and depends on the internuclear distances. Even though the ^{15}N spin pairs are far away to be directly coupled to one another strongly, the PSD will enhance the exchange of the magnetisation between the given spin pairs in these peptides. In other words, the ^{15}N T_1 relaxation in these peptides is not purely due to the molecular motion and the resulting fluctuations in the magnetic fields, but it also affected by the surrounding protons and their fluctuating magnetic fields via PSD process which is not mediated (in a simple way) by the molecular motion. The overall effect of the PSD will depend upon the sample that is investigated. There are examples in the literature where the overall PSD effects are considered to be negligible, e.g. for the corticotropin-releasing hormone (Crh), where an induced error in the measured ^{15}N T_1 by the PSD is likely to be less than the experimental error.[197] On the other hand, it has been shown that by fully deuterating the SRC homology 3 (SH3) protein, the resulting ^{15}N T_1 values are different and thus the PSD has a significant effect on the absolute T_1 relaxation time in fully protonated samples.[198] Additionally, by increasing the MAS frequency above 15 kHz, the PSD effect can be reduced significantly.[199] In this work MAS of 6 kHz was selected to obtain a good S/N ratio in reasonable time; higher spinning frequencies would result in longer experimental times, especially for the relaxation measurements. Moreover, the MAS frequencies high MAS frequencies can result in water loss and potential structural changes of the collagen molecules or fibrils. Additionally, the work done here focuses on testing the ILT performance rather extracting exact ^{15}N T_1 relaxation constants and on comparing T_1 values between molecules of similar structure, where PSD effects can be assumed to be similar in each case and thus not be the cause of any differences in ^{15}N T_1 between samples.

6.3.2. Chemical Exchange

Another mechanism that can perturb the analysis of the ^{15}N T_1 relaxation is the chemical exchange between the amide protons and the solvent, in this case, with the water. The rate of the chemical exchange is on the order of 0.5 – 50 s^{-1} and depends on the pH.[200] In the solid-state, the relaxation time constants are long; thus, the chemical exchange can be a significant contribution to the T_1 values measured in these samples. However, this is not the case for the peptides where the stabilising interchain hydrogen bonding between glycine N-H and the carbonyls (Figure 6.1.), where the N-H exchange of the glycine residue in the GPO_{10} peptide in the solution state is very slow at around 8.3 x 10^{-7} s^{-1}.[201] Thus, it is unlikely that the amide nitrogen atoms in the peptide bonds of the triple-helical collagen peptides will be affected by the chemical exchange. Furthermore, the proline residue ^{15}N will not be affected by chemical exchange at all due to the lack of the protonation on its nitrogen. Only ^{15}N in the arginine sidechain in peptide I2 and the ones participating in the water-mediated hydrogen bonding in the A-type peptides of the alanine nitrogen will have T_1 likely to be affected by chemical exchange.

6.3.3. Summary

The measured values of the ^{15}N T_1 relaxation constants for the peptides studied here will be primarily affected by the dipolar interaction between nitrogen and its bound proton, and the CSA. In the case of proline nitrogen atoms, CSA will be the dominant interaction causing relaxation because of the lack of a strong ^{1}H-^{15}N dipolar interaction. Chemical exchange is unlikely to contribute towards the relaxation, whereas PSD will influence the magnetisation decay; however, the effects can be assumed to be very similar across all peptides. Thus, only nitrogen nuclei from the same amino acids will be compared between different peptides, as the PSD effects discussed here and dipolar nuclear interactions (Section 3.1.) will be similar between the residues.

6.4. Discussion

6.4.1. Performance and Accuracy of the ILT method

The ^{15}N relaxation data were processed by fitting a single exponential to the integrated intensity of each signal of the imino rich peptides (P-type) to compare the obtained values to

the ILT processed results. The single exponential fit plots with the data are shown in Appendix B. The results show that in the case of the proline nitrogen, the T_1 constants are similar between both methods and are within their error estimates across all three peptides. Furthermore, the relaxation time constants are quite long for the ^{15}N proline in these peptides, longer than the longest τ delay used in the relaxation experiment (120 s); thus, the experiments do not give an accurate measure of T_1 in this case. Nevertheless, the data is enough to conclude that the relaxation time for proline ^{15}N is very long and too long to be useful as a probe to monitor changes in molecular motion.

In case of the glycine nitrogen relaxation, the single exponential decay shows different relaxation constants compared to the ILT estimated ones. This indicates that the Gly ^{15}N T_1 relaxation might indeed contain two or more relaxing components and suggests that the range of T_1 values that the ILT returns may in part be reflecting a genuine distribution of T_1 values.

Application of the ILT method to the Gly to Ala mutation peptides showed that the method has the potential of separating overlapping NMR signals. The Pro and Ala ^{15}N signals for that peptide overlap but are resolved in the 2D ILT spectrum according to their respective T_1 relaxation constants and chemical shifts. The chemical shifts for the Ala and Pro nitrogen atoms according to the Biological Magnetic Resonance Bank (BMRB) are expected to be at 124 and 134 ppm, respectively. These reported chemical shifts in the BMRB are an average taken from amino acids in a 3_{10}-helix secondary structure. This databank does not contain exact chemical shifts for collagen like structures. Therefore, variations in the observed chemical shifts are expected to be observed here. Nevertheless, it gives an indication that the Pro nitrogen chemical shift is higher than the Ala nitrogen, as it was observed here.

6.4.2. Gly → Ala Mutated Peptides and Deletion Peptide

The crystal structure of the Gly → Ala mutated peptide is shown in Figure 6.6. (b), where there is a disruptive site upon the mutation. All three polypeptide chains are slightly pushed apart that enables the water-mediated hydrogen bonding between all three polypeptide chains. As a result, the backbone nitrogen atoms in the Ala mutation sites have a greater amplitude of motion compared to the more tightly packed GPO peptides. This is reflected in the shorter ^{15}N T_1 relaxation constants of the Pro and Gly residues within the A-type peptides compared against imino rich peptides.

The disrupted hydrogen bonding is also observed for the D-type peptide from the crystal structure of this peptide (Figure 6.6.). The crystal structure of the Hyp^{-} peptide shows a twin

crystal structure, i.e. two molecules in an asymmetric unit, (Figure 6.6.(a)) where the two triple helical molecules are not identical. The peptide used in this work was not crystallised, but lyophilised, and thus, there can be even more variations in the given conformations of these triple helices. From the X-ray diffraction data, the interchain stabilising hydrogen bonding is only maintained by one Gly residue pre and post the deletion site giving a disrupted hydrogen bonding pattern (Figure 6.6. (a) (i) and (ii)). The relaxation constants estimated for the Gly and Pro ^{15}N in the deletion peptide are the lowest of all the peptides. This suggests that the deletion of the amino acid has a more significant effect in terms of possible conformations the peptide can have or move through than the Gly to Ala mutation. Such deletions are common for network collagen types, such as collagen IV where such flexibility and variability in structure is required.

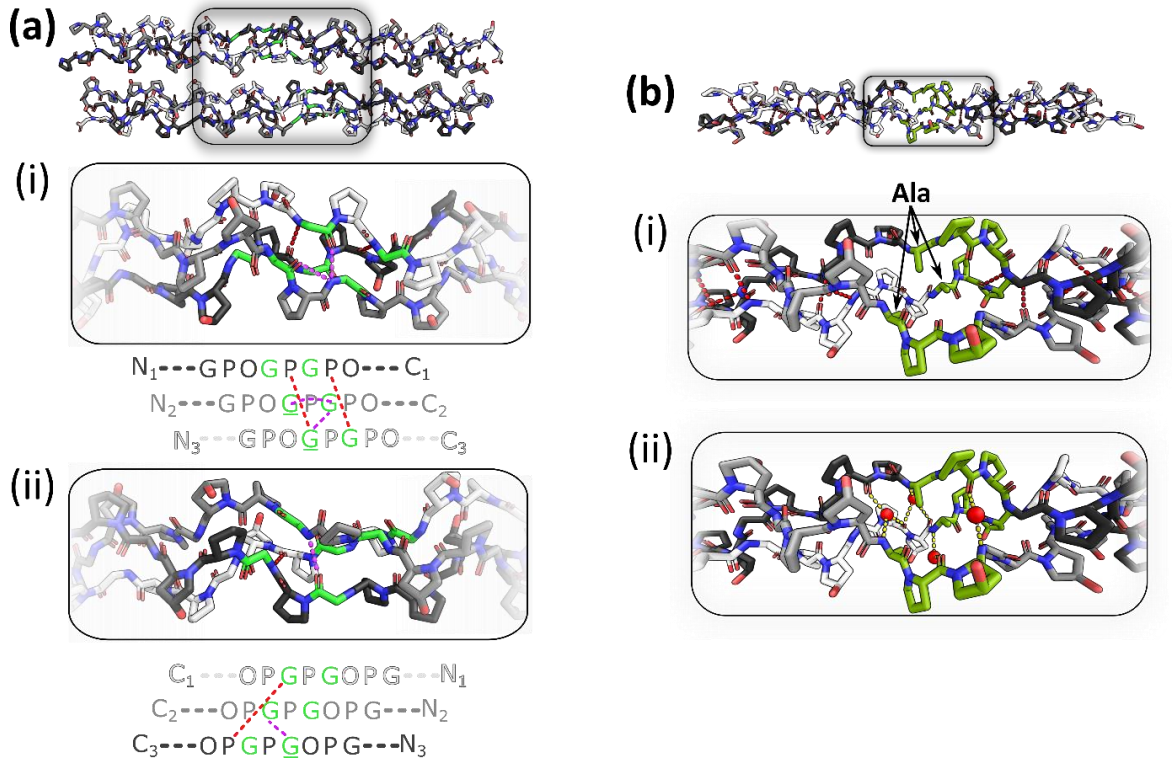

Figure 6.6. (a) The asymmetric unit of the (GPO)4GPGPO(GPO)4 peptide crystal structure; PDB code 1EI8.[109] Leading chains are coloured in white, middle in grey and lagging in dark grey. Gly in the deletion and post-deletion triplets are highlighted in green. Hydrogen bonding is shown with red dashes in the crystal structures. (i) and (ii) Top: detail of the hydrogen bonding in the deletion site of the top peptide. Below: schematic representation of the sequence showing hydrogen bonding. Hydrogen bonds between Gly NH and Pro CO are shown in red, between Gly NH and Gly CO in purple. Gly highlighted with green underscores are hydrogen bond acceptors. (b) Crystal structure of $(GPO)_5APO(GPO)_5$, with APO triplets coloured in green, PDB code 1CGD.[101] (i) Close-up view of the APO substitution site in the middle of the sequence; Ala is highlighted with black arrows. Interchain hydrogen bonding is represented with red dashed lines. Note that there is no interchain hydrogen bonding in the APO substitution site. (ii) Water-mediated hydrogen bonding shown with water molecules represented as red spheres.

6.4.4. Integrin Binding Collagen-like Peptides

The ^{15}N T_1 relaxation constants show that the end of the integrin-binding site sequence is more flexible compared to the beginning, as shown by the lower relaxation constants for both residues in peptide I2. Furthermore, the ^{15}N chemical shift of the Gly at the end of the integrin-binding site is much higher than the usual triple-helical chemical shift. This suggests that the Gly might have a different conformation in this part of the triple helix. Structural flexibility of the integrin binding sequence is implied by the X-ray diffraction studies of this peptide with and without binding of the integrin (Figure 6.7). The X-ray crystal structure shows small kinks or bends at the edges of the central zone of the GFOGER that manifest as deviations from the geometrically straight $(GPO)_{11}$ peptide (Section 2.5.4.4. Figure 2.14.).[112,113] With bound integrin, this deviation becomes even greater at the edges from the GFOGER insertion. These ^{15}N T_1 measurements suggests that backbone has a conformational flexibility on the timescale of ns, that might be important for the ligand-binding.

T_1 values obtained for the Arg sidechain ^{15}N nuclei are much lower than those of the backbone. This is likely to be in part because the sidechains have greater amplitudes of motion contributing to the nanosecond correlation times that affect T_1 and in part because of greater exposure to surrounding water and the resulting chemical exchange of sidechain N-bound ^{1}H and water ^{1}H.

6.5. Hydroxyproline to Proline Mutation in the Imino Rich peptide

Insertion of a GPP triplet (in P2 and P3) in the middle of the $(GPO)_{11}$ peptide (P1) shows a significant effect on the Gly ^{15}N T_1 constants (Table 6.2.). Relaxation values observed for the peptide P1 are decreasing with increasing temperature in a linear fashion. However, this is not the case for the peptides P2 and P3, where the Gly nitrogen T_1 decreases significantly at 270 K followed by an increase in relaxation constant at 290 K. Peptides P1 and P3 have identical isotopic enrichment, the only difference is the Pro → Hyp mutation at the eighteen position in the middle of the sequence. The similar pattern in the relaxation constants are observed between in the Pro ^{15}N T_1 constants of the P1 and P3 peptides. Although the proline ^{15}N T_1 values measured can only be considered as approximate, the proline in the X-position (P_X) ^{15}N T_1 relaxation constant in the GPP-substituted peptide P3 is significantly lower than the equivalent ^{15}N T_1 proline in the GPO_{11} peptide P1. This clearly shows that small changes

in sequence can have a significant local effect around the substitution site. We infer from the relaxation time differences that the O to P mutation results in a greater range of accessible conformations for the Gly and potentially for the P_X backbone to move between.

Furthermore, such small mutations can influence the overall stability of the triple helix as shown by the CD spectroscopy data in Figure 6.9., where a drop in CD response was shown between 60 – 70 °C, indicating that the triple helix may be dissociating or that there is a change in the triple helical structure compared to the pure GPO_{11} peptide.

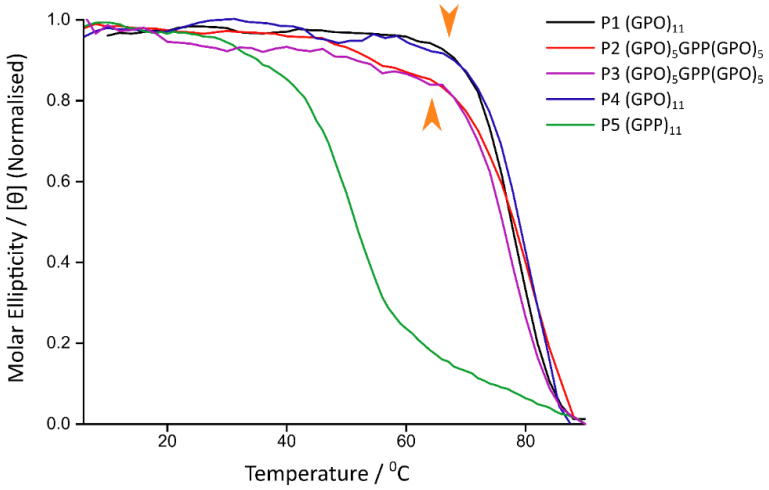

Figure 6.9. (a) Temperature dependence of molar ellipticity at 225 nm for peptides P1-P4 and at 227 nm for P5 solutions. Data were recorded at 2 °C intervals after 2 min equilibration. Plots are normalised to maximum intensity = 1 and lowest, 0 to aid comparison. The melting temperature (T_m) is measured as the temperature where the initial signal intensity has decreased by half. Estimated T_m are summarised from this plot in Table 6.1. Orange arrow heads highlight slight differences in thermal stability for the given peptides; similar trends are observed for peptides with the same sequence: P1 and P4, P2 and P3.

6.5.2. Hydration and Hydrogen-bonding of the Imino Rich Peptides

The relaxation constants show an unexpected temperature dependence behaviour for the imino-rich peptides in that the ^{15}N T_1 for glycine residues of the mutated peptides P2 and P3 the T_1 constants shows a dip at 270 K. Additionally, a similar dip in the Pro T_1 constants are also observed for $(GPO)_{11}$ P1 and mutated P3 at the same temperature. These fluctuations in the relaxation constants occur close to the expected freezing temperature of water and so we speculate that they may result from changes in bound and bulk water dynamics surrounding the model peptides. The freezing temperatures of the bound and bulk water are expected to be different. Bulk, free water freezes at around 273 K, whereas bound water freezes at a lower temperature.[202] We performed X-ray diffraction studies on $(GPO)_{11}$ and $(GPP)_{10}$ to better

understand the water network arrangement around these two peptide structures and which residues are involved in the hydrogen bonding with the surrounding water.

The water content in $(GPO)_{11}$ and $(GPP)_{10}$ peptides in crystal structures takes up almost 50% of the unit cell density; however, the arrangement of the water molecules are different between these two peptides (Figure 6.10.).[25,189] From further X-ray diffraction studies, it was found that the hydroxyproline and glycine residues participate in the hydrogen-bonding with the surrounding water molecules. Glycine and hydroxyproline carbonyls each are hydrogen-bonded with one and two water molecules, respectively. The hydroxyl group on the hydroxyproline is further hydrogen-bonded to water molecules. Additionally, there is a hydrogen bonding between hydroxyprolines in adjacent triple helices (Figure 6.11.); this hydrogen bond occurs every 20 Å along the lateral length of the peptide (Figure 6.11. (b)). This hydrogen bonding is present for the second, sixth and tenth triplets in the leading and lagging chains, and in fourth and eighth triplets in the middle chain.

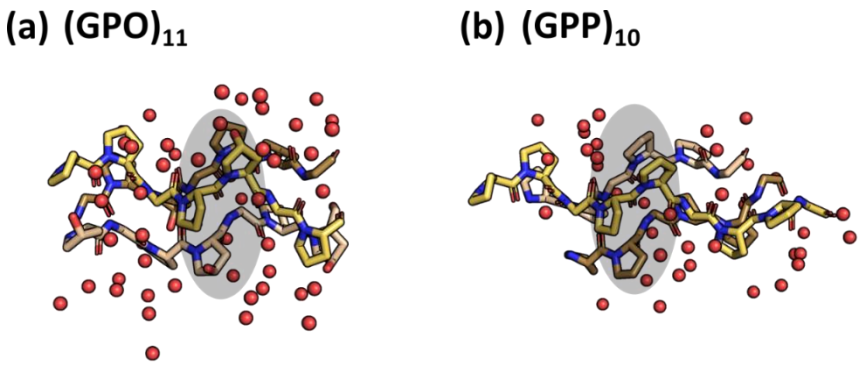

Figure 6.10. The crystal structures of peptides (a) $(GPO)_{11}$ and (b) $(GPP)_{10}$ shown with water molecules present in the unit cell consisting of two triplet repeats in both examples. PDB: (a) 1V4F[25] and (b) 1A3J.[195] In grey are highlighted (a) GPO and (b) GPP regions.

In mutated peptides P2 and P3, the GPP insertion is in the sixth triplet, meaning that the hydrogen bonding between the adjacent triple helices is lost at the centres of the helices. Furthermore, the arrangement of water molecules around the insertion site has changed due to loss of the hydrogen bonding between water and the hydroxyl group of the hydroxyproline. The resulting molecular water ordering around the insertion site can be expected to be different compared to the rest of the peptide and the $(GPO)_{11}$ peptide.

Changes of water dynamics with temperature and differences in water dynamics between peptides were confirmed by ^{1}H one-dimensional NMR spectra obtained at the same temperature range as the ^{15}N T_1 relaxation measurements (Figure 6.12.). The ^{1}H spectra in the solid-state NMR, in general, are broad and contain ^{1}H signals from the peptide ^{1}H as well

as water. Glycine Cα and Cβ ^{1}H are at 4.3 and 2.0 ppm, respectively. Proline Cα ^{1}H is at 4.4 ppm, and the rest of the Pro ring protons will be in the range of 1.9 – 3.6 ppm. Unless there is considerable molecular motion, the signals from these ^{1}H under 6 – 12 kHz MAS will be too broad to be observed above the much narrower water ^{1}H signal (narrower due to the water molecules tumbling/reorienting fast on the NMR timescale). Thus, the dominant signal in the ^{1}H NMR spectra corresponds to the water in the sample. Even though the samples are lyophilised, they still contain a high content of the structural water that cannot be removed by freeze-drying. Signals centred at 4.6, 4.7 and 4.8 ppm are observed for peptides P1, P2 (P3) and P4 (Figure 6.12.), respectively and no signals corresponding to ^{1}H in amino acids can be observed, as expected. The linewidths of the ^{1}H NMR spectra will be affected by the mobility of the water molecules. For free water, the overall resonance lines are narrow due to increased molecular motion and tumbling; however, when the water molecules are interacting with solid, the reorientation amplitudes of the water molecules are reduced or even non-existent on NMR timescale, and thus broadening of the water ^{1}H signal is observed.[202]

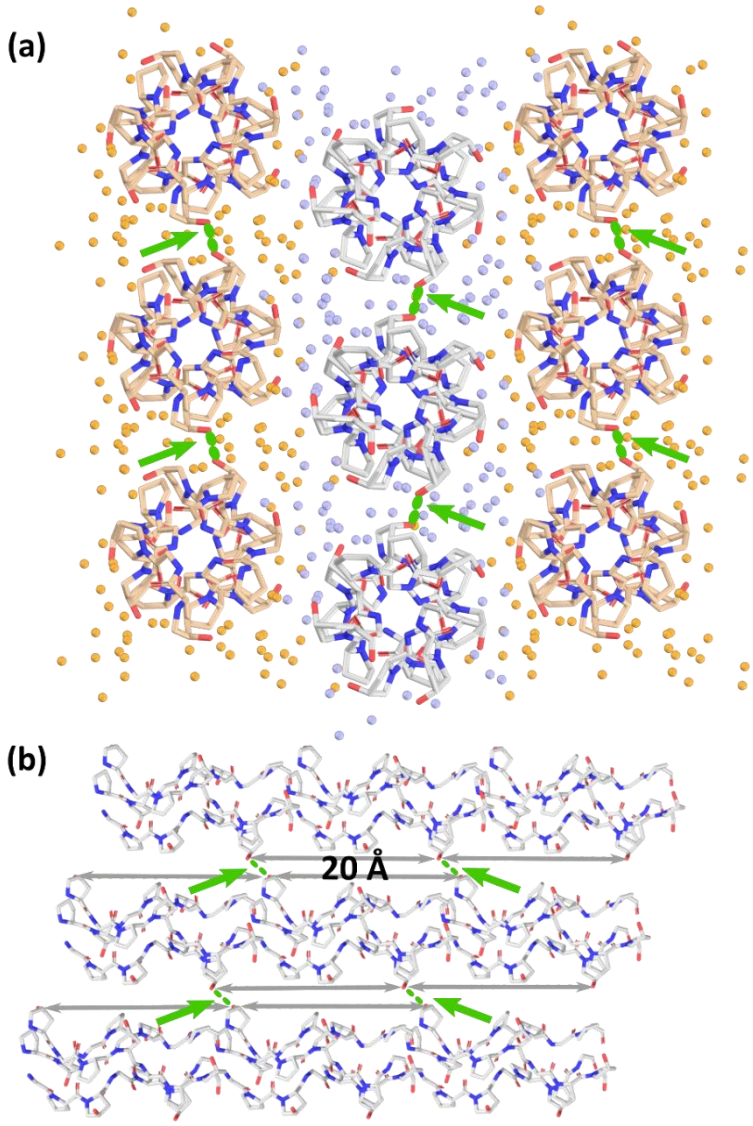

Figure 6.11. (a) Lateral arrangement of the GPO_{11} peptide with water molecule oxygen atoms shown as orange and grey spheres (PDB: 1V4F). Dashed green lines represent a hydrogen bonding between hydroxyl groups of hydroxyprolines in the adjacent triple helices. The figure was adapted from reference [25].[25] (b) The hydrogen-bonding of adjacent light grey triple helices from (a) between hydroxy groups of hydroxyproline shown with dashed lines in green colour shown. Hydroxyproline residues that are involved in the hydrogen bonding are separated by 20 Å apart from one another. Only six triplets of the total 11 are shown for clarity.

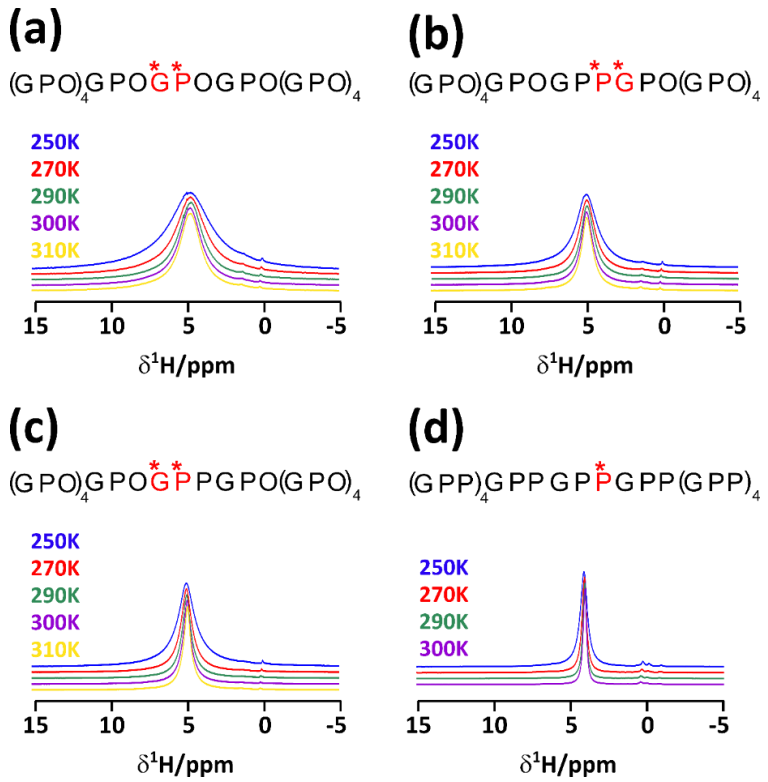

Figure 6.12. ^{1}H 1D NMR spectra of peptides (a) P1, (b) P2, (c) P3 and (d) P5 shown. The signal intensities are matched so that the linewidths of the proton signals can be compared. Each spectrum was acquired using 16 scans at the indicated temperatures under MAS of 6 kHz. Slight chemical shift changes at different temperatures of the ^{1}H signal maxima are observed in the (b), (c) and (d) spectra.

The ^{1}H linewidths are different between the $(GPO)_{11}$ peptide P1 (Figure 6.12 (a)) and the GPP-mutated peptides P2 and P3 (Figure 6.12 (b) and (c)); the width of the proton signal of GPO_{11} peptide is almost twice as broad compared to the other peptides (Table 6.6.). This suggests that the $(GPO)_{11}$ peptide P1 has less mobile water, and therefore, we hypothesise that the water in this peptide sample is more strongly involved in the binding to the peptide triple helix. To further confirm that, we recorded ^{1}H NMR spectra of peptide consisting only of GPP triplets, abbreviated as peptide P5. The X-ray diffraction studies on the $(GPP)_{11}$ peptide show reduced hydrogen bonding to the surrounding water (Figure 6.10.).[25,189] The ^{1}H NMR spectrum of peptide P5 shows even narrower linewidth compared to the rest of the P-type peptides suggesting that the water around this peptide is highly mobile, with mobility close to that of free water. The ^{1}H linewidths of the GPP-substituted peptides P2 and P3 are in between those of the peptides pure $(GPO)_{11}$ P1 and pure $(GPP)_{10}$ P5. We hypothesise that by having the GPP insertion in the peptides P2 and P3 promotes a water network at the substitution site that is observed in the $(GPP)_{11}$ peptide. The exchange between these two water networks will result in a ^{1}H linewidths that reflect the average of the nominal ^{1}H water linewidth in each network around the peptides P1-P3.

Table 6.6. The ^{1}H linewidths of the peptides used in this work shown below. The linewidth is a half-height of the signal intensity. The ^{1}H linewidth of the free water is 55 Hz at 290 K.

Temperature / K	^{1}H linewidth of P1 / Hz	^{1}H linewidth of P2 / Hz	^{1}H linewidth of P3 / Hz	^{1}H linewidth of P5 / Hz
250	1302	657	645	222
270	926	495	452	139
290	755	418	376	111
300	689	366	340	105
310	657	348	322	-[a]

[a] *Data was not acquired at this temperature to avoid possible denaturation of the triple helix.*

The change in the ^{15}N T_1 relaxation around the 270 K is not only observed for the imino rich peptides but also for the Ala substituted residues. The crystal structures and the water network around the Ala is not discussed in a detailed manner as was done for the P-type peptides, as the crystal structure is only available for the $(GPO)_5APO(GPO)_5$ (A1 and A2), but not for other type of variations that would include GPP or APP triplets. Nevertheless, the ^{1}H results show that there are different water networks around the triple helices with different mobilities. In general, when water freezes, molecules rearrange themselves, and energy is released that will be absorbed by the triple helix and the molecular motions. Such processes can be expected to cause a shortening of T_1 at the temperature at which water locally freezes around the residue for which T_1 is being measured at. In the case of GPP mutation (in P2 and P3) where the local water network is more mobile compared to around the GPO triplets, water freezing is expected to occur at relatively higher temperatures compared to that for water around GPO triplets. The observed further variations in the ^{15}N T_1 of the A-type peptides suggest that the freezing might take place over a range of the temperatures depending on the sequence and the resulting water network around it.

6.6. Summary

Overall, the T_1 constants estimated using the ILT shows that this method can be used to process ^{15}N relaxation data and gain useful information on the molecular motions on the backbone of synthetic peptides that model sections of the triple helix. Additionally, this method demonstrated that it can be used to separate overlapping signals and give approximate relaxation constants for overlapped components. The resulting ^{15}N T_1 values obtained for imino-rich peptides showed that a simple substitution from hydroxyproline to proline in the middle of the polypeptide chain results in a molecular motion change around the substitution

site. The relaxation results show that substitution of GPO by GPP in the middle of the sequence might result in a greater variety of conformational structures available to move through. Decreased ^{1}H linewidths of peptides containing GPP compared to those with only GPO showed that there is more mobile water surrounding the triple helices at the GPP insertion site than around GPO. A decrease in relaxation constants was observed for the glycine residues for GPP-substituted peptides at the water freezing temperatures. This observation can be attributed to the energy absorbed by the triple helix upon the 'free' like water freezing surrounding the GPP insertion site. Additionally, the loss of some intermolecular hydrogen-bonding between the adjacent triple-helical centres of the GPP-substituted peptide molecules can give greater conformational freedom.

Based on these results, we hypothesise that ^{15}N T_1 relaxation constants can serve as a scale for variations of the molecular motions of the backbone in the collagen from extracellular matrix. We now have indicators of ^{15}N T_1 values for regions of tightly packed triple helices, for more flexible and disrupted triple helices, by reference to the model peptides studied here.

7

Effects of Disease in Collagen Fibrils and ECM

The aim of this chapter is to gain insights into the dynamic structural changes that take place in collagen following glycation chemistry and to explore the corresponding biological consequences. As was discussed previously in Section 2.4., glycation is a spontaneous non-enzymatic reaction that takes place between ambient sugars and free amine groups that are present in proteins, lipids or nucleic acids.[29,203] Collagen fibrils are readily accessible to sugars in the extracellular fluid. Glycation generates many complex intermediate products while further reactions result in so called advanced glycation end products (AGEs), some of which are intermolecular cross-links. Glycation leads to chemical and physical changes of the collagen fibrils that in turn alters the mechanical and biological properties of the extracellular

matrix and tissues. It has been shown that upon glycation, for example, collagen-proteoglycan binding is decreased and cell adhesion is weakened.[204]

Studies of glycated collagens have focused on the changes at a macromolecular scale using TEM, SEM and AFM microscopy; however, there are fewer studies examining detailed molecular structure and dynamics. Solid-state NMR spectroscopy has proven useful in identifying many glycation products and has even detected new, previously unreported species, including oxalamido and hemiaminal substructures.[16,17] Conventional studies, such as mass spectrometry (MS) and fluorescence spectroscopy, of glycation products have been unable to detect these species due to the sample processing involved prior to application of the specific technique. For instance, the applicability of MS to study these systems relies on hydrolysis of insoluble proteins to extract glycation-modified residues where the initial materials are subjected to harsh conditions, including acidic conditions and high temperatures. Therefore, detection of the glycated residues relies on the stability of these products under these pre-treatment conditions whereas NMR does not.

The formation of the glycation products can affect the molecular dynamics of the collagen fibrils. For example, more tightly packed overlap zones are likely to have different modes of molecular motion compared to the more loosely packed gap zone of the collagen fibrils, and the cross-linked collagen will exhibit different molecular dynamics compared to the normal, non-crosslinked collagen fibrils. Therefore, molecular dynamics studies of collagen can probe structural changes accompanying glycation, as well as revealing dynamical properties important for ligand binding and ultimately cell function.

As discussed in the previous chapter, NMR relaxation is caused by fluctuating local magnetic fields, the main source of which is molecular motion. Therefore, the collagen dynamics probed by solid-state NMR relaxation can be related to changes in the structural and molecular packing caused by glycation chemistry.

This study uses commercially-available collagen type-I and [U-^{13}C,^{15}N]-isotopically enriched in-vitro extracellular collagen, further reacted with ribose-5-phosphate (R5P) resulting in glycated collagen (see Materials and Methods chapter for details of sample preparation). Glycation with glucose can be rather slow thus we use R5P to maximise the extent of glycation so as to clearly observe the effects of the chemistry. R5P is an important cellular sugar that is an intermediate in nucleic acid metabolism.[205,206] We use various isotopic enrichments of collagen to target different parts of the fibrils, allowing us to locate changes in fibrillar molecular motions. Amino acids that are not directly involved in chemistry with the reducing sugar can be regarded as inert reporters on the overall fibrillary molecular dynamics. For this we

incorporated [U-^{13}C,^{15}N]-proline and -glycine to assess changes in backbone motions. Isotopic enrichment of phenylalanine and leucine are used to investigate the changes in side chain motions as these will reflect the changes of the intermolecular packing in the hydrophobic regions. Finally, we use isotopic enrichment of amino acids that are targeted by the glycation chemistry directly: lysine and arginine.

7.1. Glycation of Collagen Type-I

First, we investigate the overall glycation effect on the 'pure' collagen type-I fibrils reconstituted from acetic acid-soluble collagen obtained from bovine Achilles tendon (see Materials and Methods Chapter 5.). In Figure 7.1 ^{13}C CP-MAS spectra of 'pure' collagen, and collagen after incubation with R5P for six weeks, are compared.

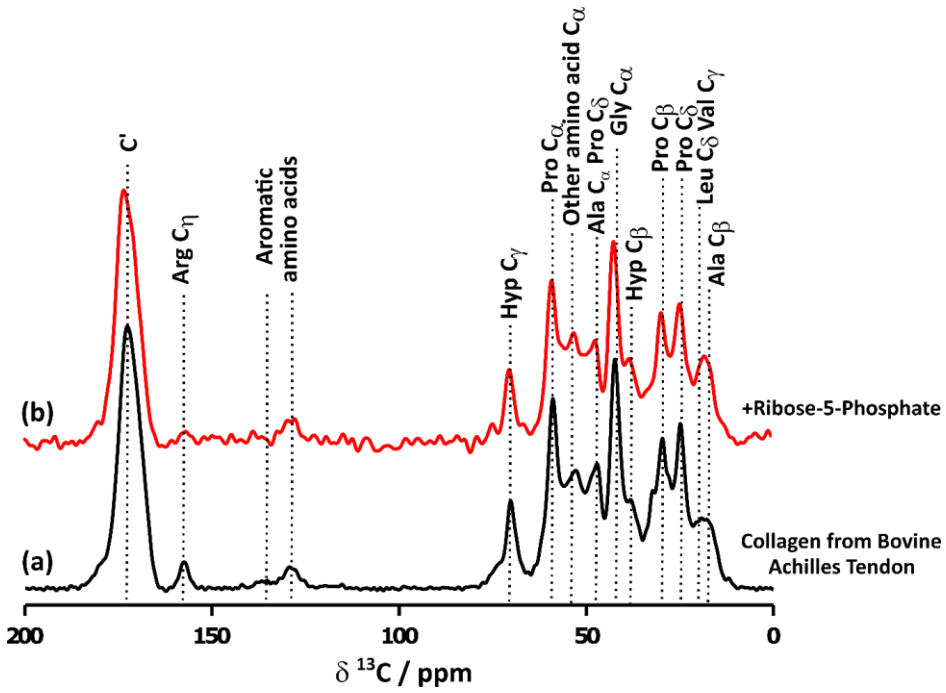

Figure 7.1. 1D ^{13}C spectra of collagen reconstituted from bovine Achilles tendon (a) before, and (b) after glycation with ribose-5-phosphate. Most prominent NMR signals are assigned.

The spectra in Figure 7.1 are similar. The ^{13}C 1D NMR spectrum of glycated collagen shows slight broadening in the carbonyl, aliphatic and hydroxyproline Cγ regions. Most of the glycation intermediate products generated between collagen and R5P are expected to give signals in these regions, accounting for much of the apparent line broadening.[16,207] The glycated sample does not show sharp signals from glycation products due to their

heterogeneity, resulting in many overlapping signals, and low isotopic abundance. Glycation on collagen fibrils is expected to be mainly observed on lysine, hydroxylysine and arginine sidechains. However, not all these residues will be glycated.[17] The total fraction of residues that can be targeted by glycation (Lys, Arg and Hyl) is around 13 % in the Achilles bovine tendon collagen (see Section 5.4.2.2. Table 5.5.). Therefore, overall NMR signals corresponding to Lys and Arg residues even in unglycated fibrils will be buried under the NMR signals of higher abundance residues like proline, glycine and hydroxyproline.

The ^{13}C T_1 relaxation data of ^{13}C natural abundance unglycated and glycated collagen samples were processed using ILT. The contour plots are shown in Figure 7.3. together with the calculated ^{13}C T_1 relaxation constants. In these examples, we give a range of the T_1 values for specific spectral regions, as we cannot assign all the residues in the natural abundance sample.

The contour plot of the reconstituted bovine tendon collagen shows a separation of three differently relaxing regions as expected (Figure 7.2. (top)), with the highest T_1's observed for the carbonyl C' region (176 – 169 ppm), followed by Cα region between 40 – 60 ppm together with Pro sidechains at 23 – 30 ppm, and the shortest relaxation constants observed for the amino acid side-chains in range from 10 to 40 ppm.

The ILT plot of the glycated collagen show a similar separation of the relaxing regions as observed in the unglycated sample (Figure 7.2. (bottom)). In most cases, the T_1 relaxation constants for the glycated sample are lower. The Cα's resolve into two regions, where the majority of Cα's and Pro sidechains have T_1 constants covering a smaller T_1 range compared to the unglycated sample between 4 – 9 s. Further, no distinctive signal is observed for the Pro Cα as it was in the unglycated sample. The decrease in the signal intensity can be attributed to the overlap of the chemical shifts corresponding to the glycation products. These glycation products have strong broad NMR signals observed in the range of 50 – 110 ppm (when isotopic enrichment for sugar is used).[16,17] These signals arising from the glycation products happen to overlap with the Pro Cα, that is around 59 ppm. Furthermore, the side-chain region signals at ca. 40 – 30 ppm in the ILT plot in are absent. The signals at ca. 40 ppm in the unglycated sample may correspond to the Arg Cδ, Cε of Lys, Hyl, and Cβ of Asn, Asp, Ile and Leu atoms; signals at the lower frequency range towards 30 ppm correspond to most of the other Cβ atoms from rest of the amino acids. These signals in the glycated sample could be overlapping together with the Cα region and Pro side-chain signals. Furthermore, the loss of the signal intensities of the Arg Cδ, Cε of Lys, Hyl carbon atoms is

unsurprising, as these residues are expected to be involved in the glycation and the resulting new species will change the chemical shifts of these atoms.

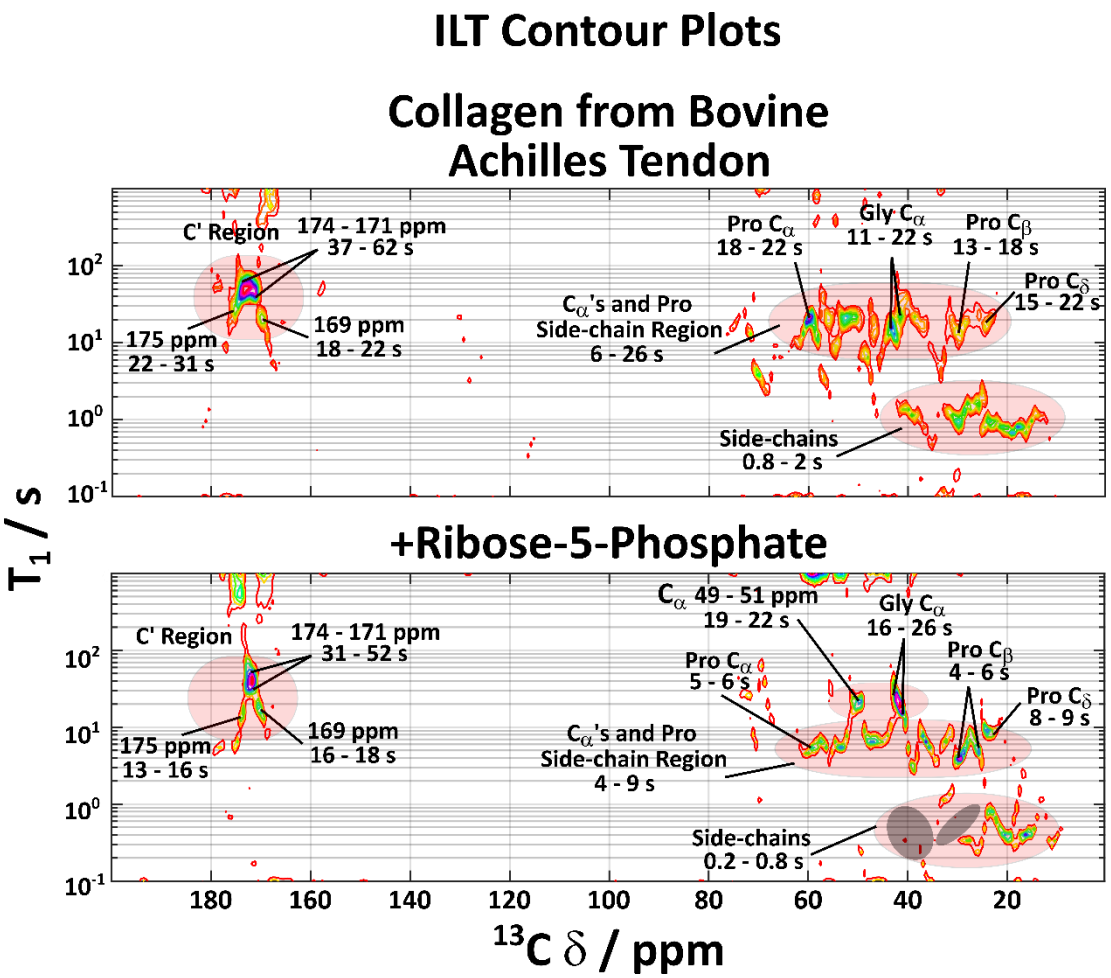

Figure 7.2. ILT contour plot of the ^{13}C T_1 relaxation time distribution plotted against ^{13}C chemical shift for unglycated (top) and glycated collagen with ribose-5-phosphate (bottom). Regions with similar relaxion are highlighted in pale pink. The dark grey circles highlight missing side-chain relaxation regimes in the glycated sample.

There are two signals in the contour plots observed at ca. 42 ppm corresponding to Gly Cα, and between 51 – 49 ppm attributed to other amino acid Cα's chemical shifts. The observed T_1 relaxation constant for both of these regions is higher than the rest of the Cα's, in the glycated sample. Gly is the smallest amino acid with no side-chains and is buried in the middle of the triple helix, and thus less accessible to the solvent. Therefore, both higher Cα T_1 regions may correspond to the hydrophobic regions of the collagen fibrils.

Overall, the ^{13}C T_1 relaxation results of the unlabelled reconstituted collagen from the bovine Achilles tendon show a general decrease in the observed T_1 relaxation constants which indicates a general increase in the molecular mobility. Further explanation for observed relaxation behaviour is discussed in Section 7.4. The next sections examine relaxation results

of isotopically enriched ECM collagen that is subjected to the same glycation chemistry as these reconstituted collagen fibrils.

7.2. Inert amino acids – overall collagen molecular flexibility reporters

7.2.1. U-^{13}C,^{15}N-Glycine and Proline Enriched in vitro Collagen

The glycine and proline amino acids make a significant contribution to the collagen molecular structure[92] (see Section 2.4.5.1.). Collagen fibrils [U-^{13}C,^{15}N]-enriched at Gly, Pro and Hyp were obtained by culturing primary foetal sheep osteoblasts and extracting the ECM that they produce. The extracted ECM was subjected to multiple detergent washes and enzymatic digestion using α-chymotrypsin to remove the extracellular proteins. This afforded a collagen rich matrix. Experimental details describing the cleaning process and the total amounts of the collagen left in the samples is presented and discussed in Chapter 5. These appropriately isotopically enriched collagen rich materials were further glycated with R5P (6 weeks) in the same way as the reconstituted collagen described in the previous section.

Gly, Pro and Hyp are not directly involved in glycation chemistry. Thus, they can be used to study indirect effects that the glycation chemistry has on the fibril structure, and in particular, the dynamics of the collagen backbone. High abundance of these amino acids in collagen type-I enables successful investigation of ^{15}N relaxation as well as ^{13}C.

In Figure 7.3. 1D ^{13}C and ^{15}N CP-MAS NMR of the [U-^{13}C,^{15}N]-enriched of normal and glycated in vitro collagen are shown. No difference in the ^{13}C NMR signals of Pro and Gly was observed between the glycated and unglycated samples. However, the 1D ^{15}N NMR signal of the Pro shows higher intensity under the same cross-polarization conditions between the samples. This could be attributed to many factors, including different CP efficiencies. The CP efficiency will depend on the spin dynamics, whereby the rate of signal build-up is dictated by the heteronuclear dipolar couplings between the spins that are participating in the CP transfer (see Section 3.2.3.). The molecular motions that are responsible for the decay of the CP signal intensity will have correlation times of the order of the inverse of the spin-lock nutation frequency, which in this case is motions with timescales of ms. In this particular set of samples, it was observed that the ^{15}N signal of the Pro exhibits a more rapid build-up of the signal intensity across different contact times for the glycated sample. The observed differences of

the signal build-up might indicate that the extent of molecular motions at the ms time-scale are different for the glycated sample. Further differences in the signal intensities could result from varying concentrations of the isotopic enrichment in the final sample. This is because isotopic enrichment is subjected to cell metabolism which might be changed depending on the passage of the cells used (cells can modulate their metabolomic status and differentiation). During these experiments it was ensured that the samples used here are from passages close to one another (± 1 to 2 passages maximum).

The Pro ^{15}N chemical shift is slightly greater, by 0.8 ppm, for the glycated sample compared with unglycated, centred at 129.6 ppm, and the Gly signal is 0.1 ppm higher in the glycated sample.

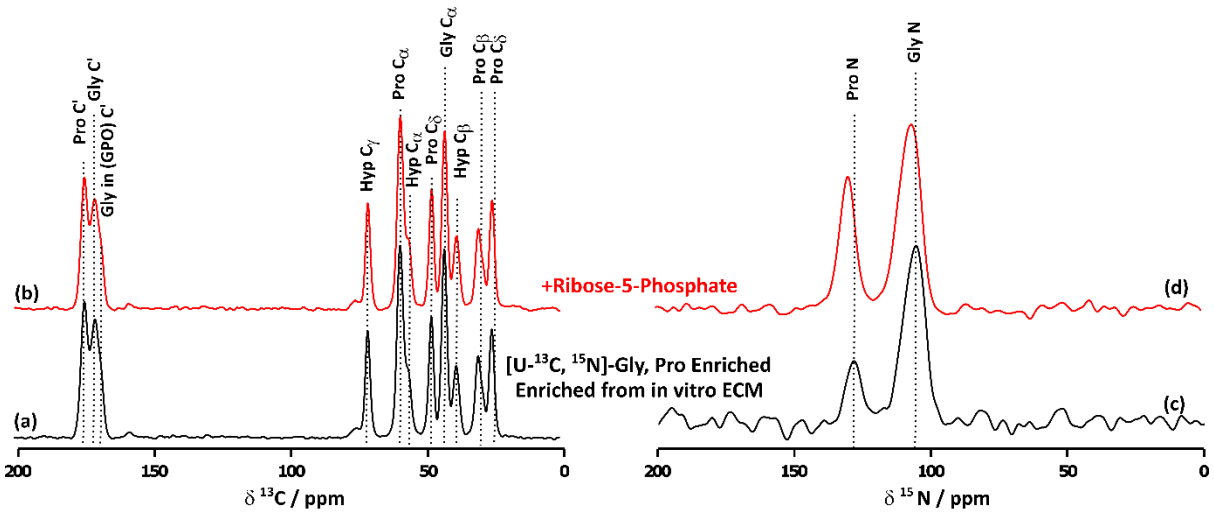

Figure 7.3. (a) – (b) 1D ^{13}C and (c) – (d) 1D ^{15}N CP MAS NMR spectra of normal (black) and glycated with ribose-5-phosphate (red) collagen enriched with [U-^{13}C,^{15}N]-Gly, Pro from in vitro sheep osteoblasts. The NMR spectra is dominated by signals from Pro, Hyp and Gly (assigned). All spectra were recorded with 1024 scans, with contact times of 2.5 ms and 3.5 ms for the ^{13}C and ^{15}N NMR, respectively.

T_1 Relaxation measurements were carried out for the ^{13}C and ^{15}N nuclei for both samples, and the resulting ILT is shown in Figure 7.4. and 7.5., respectively. The observed ^{13}C T_1 relaxation constants are not divided in to the same groups as it was observed for the unlabelled collagen samples in the previous section, where the carbonyls were observed with higher T_1 values compared to the Cα carbons. In these isotopically enriched ECM collagen samples, the relaxation constants for each residue between the carbon atoms are similar, i.e. Gly C' and Cα have the same T_1 relaxation range of 8 – 12 s. This result could be due to the additional homonuclear dipolar interaction between the ^{13}C–^{13}C and heteronuclear ^{13}C–^{15}N nuclei added due to the isotopic enrichment.

The ^{13}C T_1 relaxation constants of the glycated [U-^{13}C,^{15}N]-Gly, Pro enriched collagen ECM samples are about 1 – 3 s lower than for the unglycated sample (Figure 7.4.). In both contour plots, Gly carbons show longer T_1 values. This difference is attributed to the Gly being buried in the middle of the triple helix, and the further lack of sidechains that could reduce T_1 through dipolar interactions between the ^{13}C–^{13}C and ^{13}C–^{1}H nuclei, compared to Pro residue. Additionally, the signals observed in the ILT contour plot corresponding to the Pro carbons are broader in the T_1 dimension, showing a wider distribution of T_1 values especially in the side-chain region (Cβ and Cγ). Gly carbon T_1 distribution, on the contrary, is shorter by 1 s compared to the unglycated sample.

Contour plots of the ^{15}N T_1 relaxation results of the glycated sample shows higher and broader signals for the Pro and Hyp nitrogen nuclei (Figure 7.5.). The Gly distribution of the ^{15}N T_1 values is more homogeneous for the glycated sample, where in the ILT plot of unglycated sample shows a separation in at least two major regions giving a broader distribution of T_1 constants. The chemical shifts for these two Gly nitrogen regions are 110 – 108 and 107 – 104 ppm, where these may be corresponding to the Gly residues in the collagen and other proteins present in the ECM, e.g. proteoglycans; or it could be corresponding to the Gly residues being in different parts of the collagen fibrils, i.e., gap and overlap zones.

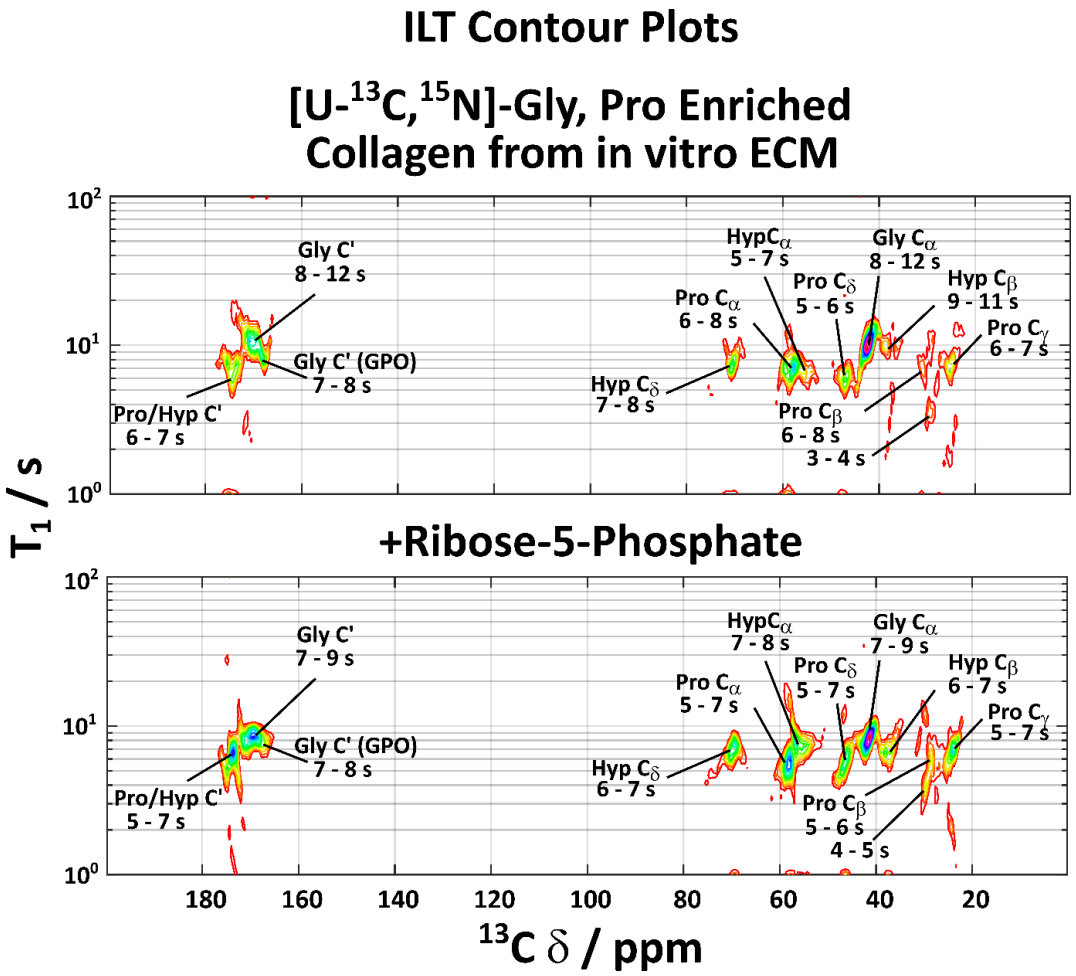

Figure 7.4. ILT contour plots of ^{13}C T_1 normal (top) and glycated (bottom) of collagen enriched with [U-^{13}C,^{15}N]-Gly, Pro from in vitro sheep osteoblasts. Each nuclei type relaxation data were processed by ILT using the same parameters in order to be comparable to one another.

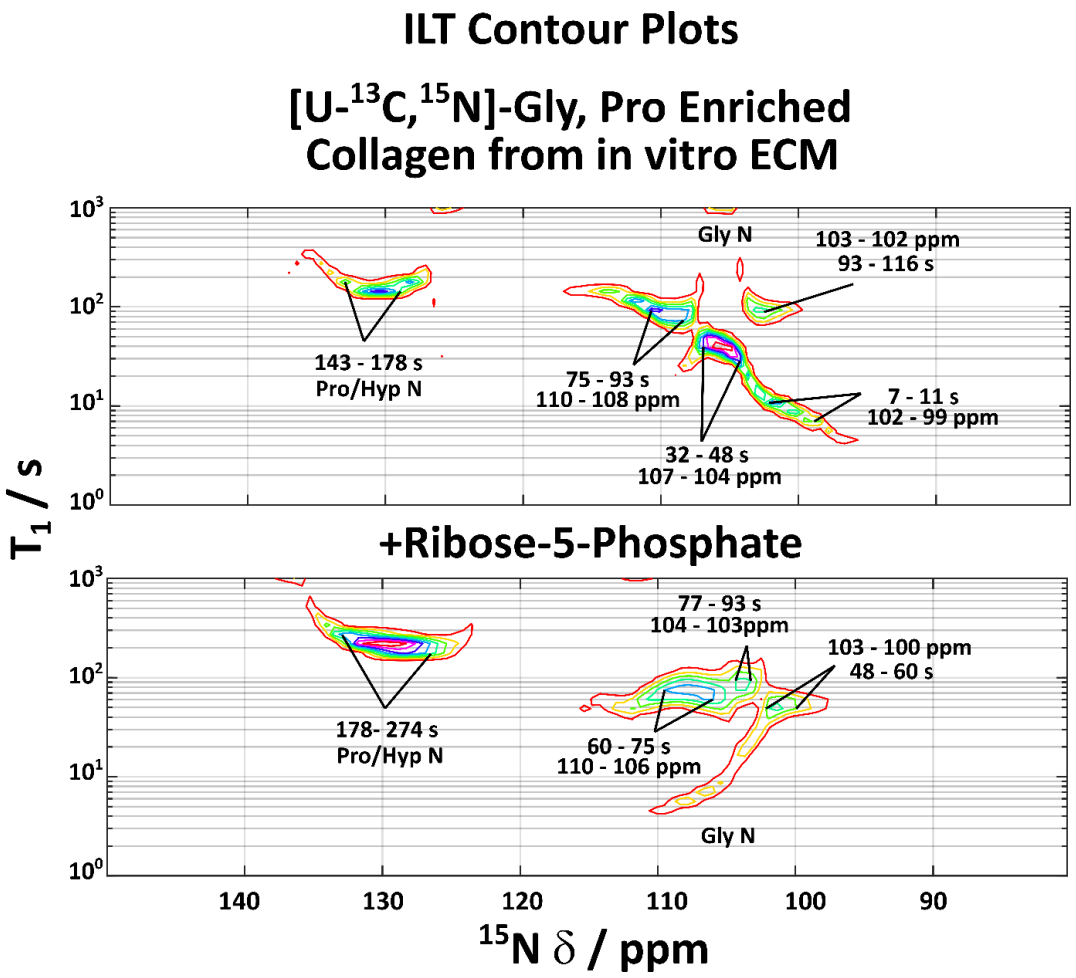

Figure 7.5. ILT contour plots of ^{15}N T_1 normal (top) and glycated (bottom) of collagen enriched with [U-^{13}C,^{15}N]-Gly, Pro from in vitro sheep osteoblasts. Each nuclei type relaxation data were processed by ILT using the same parameters in order to be comparable to one another.

7.2.2. U-^{13}C,^{15}N-Leucine Enriched in vitro Collagen

Leucine residues are identified as one of the amino acids that can undergo structural reorganisation upon fibril formation.[137] The Leu sidechains in the collagen nanofibrils are folded along the triple-helical axis, minimising the hydrophobic surface of the collagen molecule. Furthermore, they are predominantly distributed around the collagen C and N-termini of the collagen fibrils, on the interfaces between gap and overlap zones and in one band in the hole zone (Figure 7.6.). In this section we investigate the molecular dynamics changes upon the glycation of [U-^{13}C,^{15}N]-Leu isotopically enriched in vitro collagen in order to assess hydrophobic regions of the collagen fibrils in much closer detail.

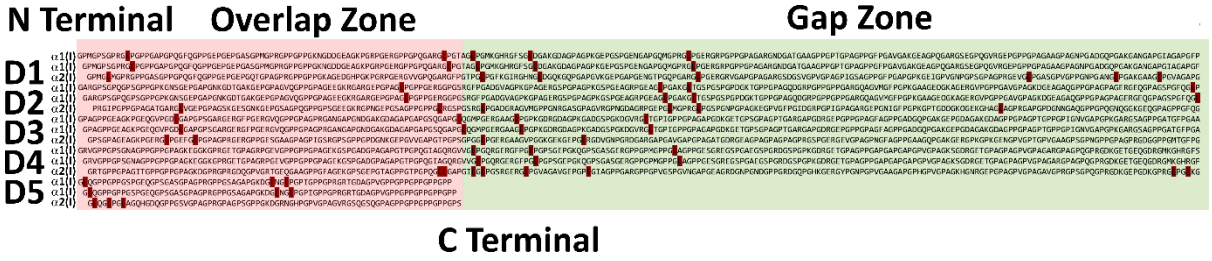

Figure 7.6. Consensus sequence for collagen type-I with Leu residues highlighted in dark red. The gap and overlap zones are highlighted in light green and red colours, respectively.

The 1D ^{13}C POST C7 and ^{15}N CP-MAS NMR spectra of the [U-^{13}C,^{15}N]-Leu enriched normal and glycated collagen are shown in Figure 7.7. For the ^{13}C spectrum, the POST C7-filtered NMR experiment was used to selectively observe signals from isotopically-enriched residues so as to avoid overlap with the natural abundance signals from the Pro, Gly and Hyp residues. The 1D ^{15}N NMR spectra was recorded under DNP conditions due to low abundance of the Leu residue in collagen. The DNP experiments were performed at a temperature of 100 K; as a result, the observed signals might be broader due to the decrease in the molecular motions at this temperature and the consequent “freezing in” of conformational heterogeneity. The ^{13}C spectra of normal and glycated materials appear similar. On the other hand, in the ^{15}N 1D NMR spectra the glycated materials nitrogen nuclei is shifted by 0.6 ppm higher compared to the unglycated sample.

The T_1 relaxation measurements were carried out for the ^{13}C nuclei, it was not possible to obtain a sufficient signal-to-noise ratio for the ^{15}N nuclei that could be further used for the T_1 relaxation studies. The resulting ILT contour plots are shown in Figure 7.8.

The ILT contour plots show a separation of T_1 signals from isotopically-enriched amino acids and the residues containing natural abundance ^{13}C nuclei of Pro, Gly and Hyp (signals

highlighted in Figure 7.8.). ^{13}C NMR signals corresponding to these latter three amino acids are observed due to their high abundancies in collagen type-I. The observed relaxation constants for the natural abundance nuclei are higher due to the lack of the ^{13}C-^{13}C dipolar interaction.

The ^{13}C T_1 relaxation constant distribution of Leu side chains (Cδ, Cγ) for the unglycated collagen fibrils are uniform and similar to the backbone (Cα, C') carbonyl carbon nuclei T_1 values. After glycation, the ^{13}C T_1 relaxation constants of the Leu carbons have a broader distribution, especially for the side chain carbons (Cδ, Cγ) where relaxation constants are in range of 0.3 – 9 s. The ^{13}C relaxation results show that the carbon nuclei of the Leu side chains in the glycated sample are experiencing various environments, with different conformational degrees of freedom, i.e. parts within the fibrils with increased and decreased molecular mobility.

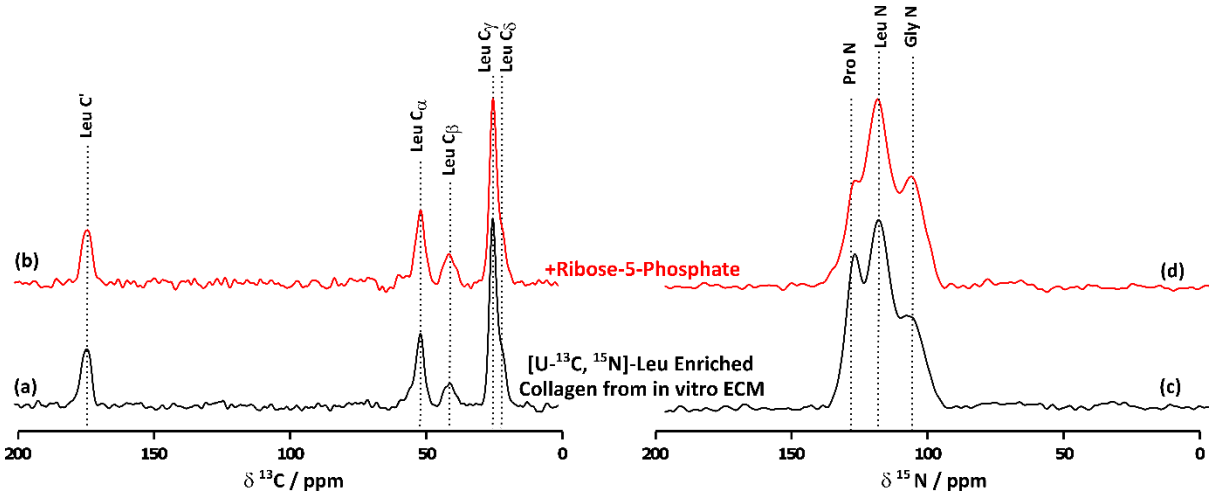

Figure 7.7. 1D ^{13}C POST C7 filtered CP MAS NMR spectra of (a) normal and (b) glycated with ribose-5-phosphate collagen enriched with [U-^{13}C,^{15}N]-Leu from in vitro sheep osteoblasts. POST C7 filtering was used to only detect isotopically-enriched ^{13}C nuclei, i.e. U-^{13}C-Leu. In (c) – (d) 1D ^{15}N CP MAS DNP spectra of the same samples shown in (a) – (b), respectively. ^{13}C and ^{15}N spectra were recorded with 4096 and 1024 number of scans, respectively. Contact times of 2.5 and 3 ms for the ^{13}C and ^{15}N NMR experiments, respectively, were adopted. The ^{15}N 1D DNP experiments were recorded at temperature of 100 K, whereas the ^{13}C 1D experiments were recorded at 293 K.

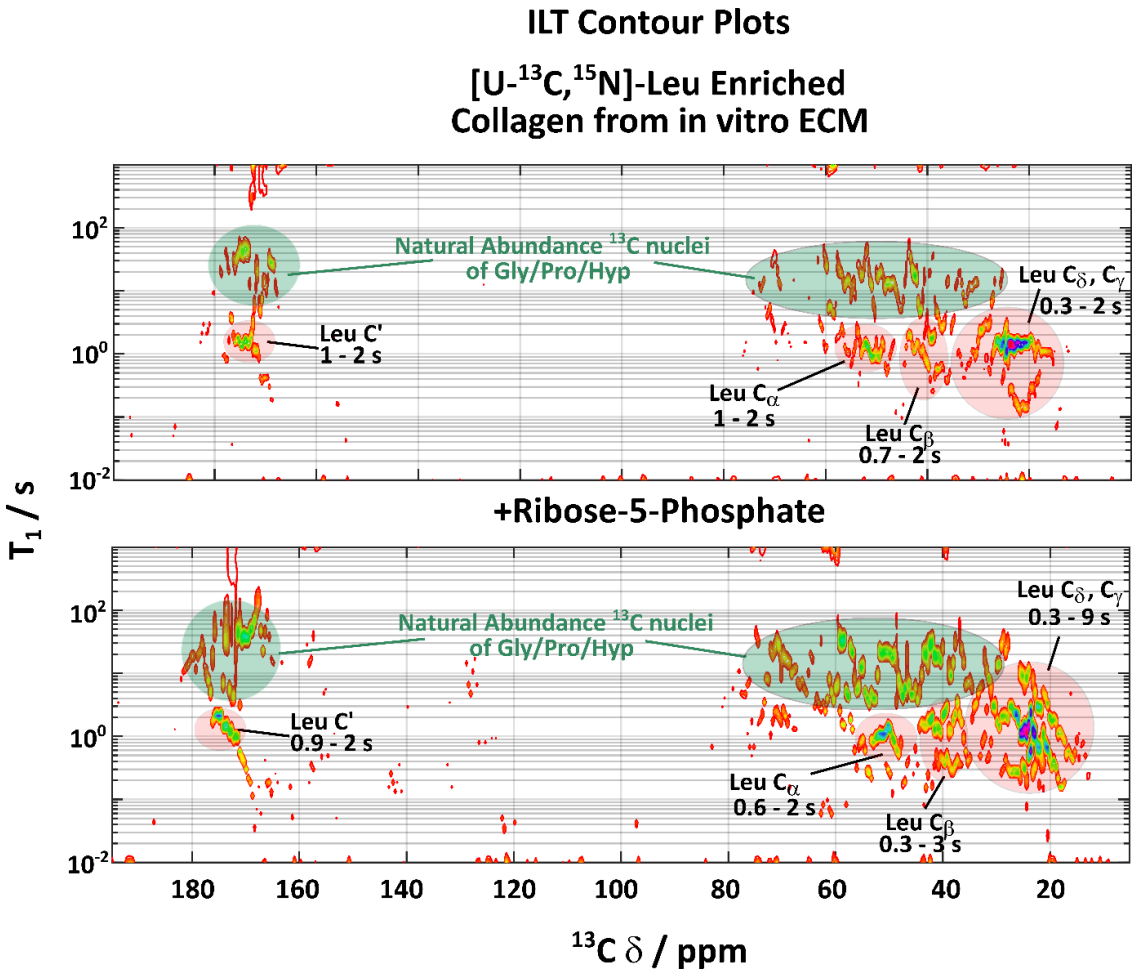

Figure 7.8. ILT contour plots of ^{13}C Leucine T_1 for normal (top) and glycated (bottom) [U-^{13}C, ^{15}N]-Leu enriched collagen obtained from in vitro sheep osteoblasts. Chemical shift regions corresponding to the natural abundance ^{13}C nuclei are highlighted in light green. Both relaxation data sets were processed using the same ILT parameters in order to be comparable to one another.

7.2.3. U-^{13}C,^{15}N-Phenylalanine Enriched in vitro Collagen

The relative amounts of the aromatic amino acids within the collagen type-I triple helix are low in general, contributing only 1.2% of total amino acid content per collagen triple helix unit (calculated based on the sequence shown in Figure 7.9.). These residues have an important role in the formation of the higher-order structures and in performing biological functions. The phenylalanine residues aligned in the gap region are suggested to be conferring the rigidity and stability via Phe-Phe interactions.[208] Studies on synthetic model peptides have shown a presence of proline-aromatic CH···π interactions between adjacent triple helices;[140] this interaction has been further observed in fibrillar collagen type-I via DNP NMR.[139] The GFO triplet has been recognised as a crucial component of the sequence recognition site for the integrin α2β1 A-domain.[209] Distribution of the Phe residues in the mammalian consensus collagen type-I sequence are highlighted in Figure 7.9.

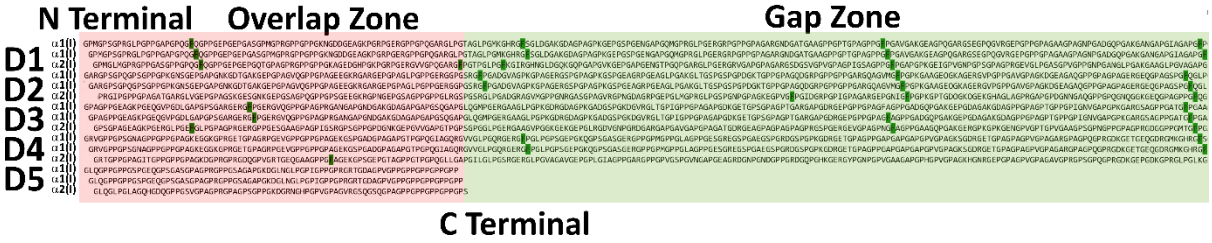

Figure 7.9. Consensus sequence for mammalian collagen type-I with Phe residues highlighted in the green. The gap and overlap zones are highlighted in light green and red colours, respectively.

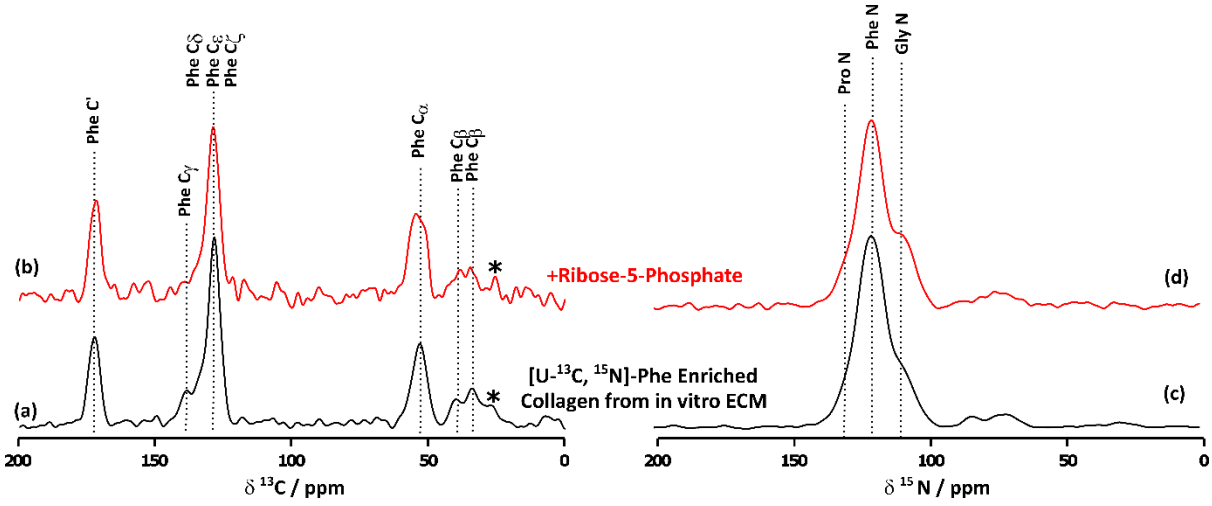

Figure 7.10. 1D ^{13}C POST C7 filtered CP MAS NMR spectra of (a) normal and (b) glycated with ribose-5-phosphate collagen enriched with [U-^{13}C,^{15}N]-Phe from in vitro sheep osteoblasts. POST C7 filtering was used as to selectively detect isotopically-enriched ^{13}C nuclei, i.e. U-^{13}C-Phe. Spinning side bands are highlighted with asterisk. In (c) – (d) 1D ^{15}N CP MAS DNP spectra of the same samples shown in (a) – (b), respectively. ^{13}C and ^{15}N spectra were recorded with 8192 and 1024 number of scans, respectively. Contact times used were 2.5 and 3 ms for ^{13}C and ^{15}N NMR experiments, respectively. The ^{15}N 1D DNP experiments were recorded at temperature of 100 K, whereas the ^{13}C 1D experiments were recorded at 293 K.

The 1D ^{13}C POST C7 and ^{15}N CP-MAS NMR spectra of the U-^{13}C,^{15}N]-Phe enriched normal and glycated in vitro sheep osteoblast collagen are shown in Figure 7.10. The 1D ^{13}C NMR shows small differences in the chemical shifts observed between the unglycated and glycated materials. In the glycated sample, the Phe C' chemical shift is lower by 0.9 ppm towards lower frequency, where the chemical shift of aromatic carbons (Cδ, Cε, Cζ) has increased by 0.5 ppm compared to the unglycated sample. Broadening is observed for the Cα signal with the maxima shifted by approximately 1.5 ppm to higher frequency for the glycated sample. The Cβ signal region shows separation of two signals, which is more clearly observed in the 2D NMR spectra (Figure 7.11.). After glycation, the resolution of the two Cβ signals is lost (Figure 7.10.), potentially indicating a broader range of possible conformations accessible for the Phe residues that results in overall decrease in the signal-to-noise ratio. The 1D ^{15}N spectra show little difference between unglycated and glycated materials.

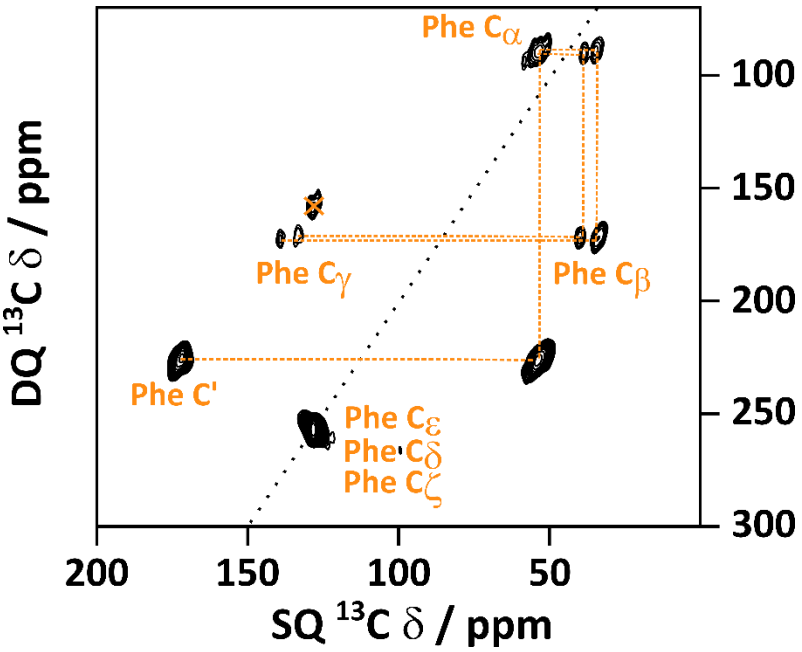

Figure 7.11. 2D ^{13}C-^{13}C SQ-DQ POST C7 correlation spectrum of [U-^{13}C,^{15}N]-Phe enriched collagen from in vitro sheep osteoblasts. The correlations of the ^{13}C-Phe carbon atoms are shown in orange. Phe Cβ has two ^{13}C signals centred at 38.6 and 34.6 ppm. The SQ-DQ diagonal is highlighted with black dashes through the 2D spectra. The spinning sideband of the aromatic carbons (Cδ, Cε, Cζ) is crossed out.

The ILT contour plots of ^{13}C T_1 relaxation experiments of Phe labelled samples are presented in Figure 7.12. The observed signals show a strong overlap with the natural abundance ^{13}C nuclei of other amino acids. The ^{13}C T_1 relaxation measurements are performed using CP rather than the POST C7 filtered experiments. The resulting 1D CP NMR spectra of the both materials are shown in Figure 7.13. The spectra show strong signals of the natural abundance nuclei of the Gly, Pro, Hyp and other amino acid side chains. Therefore, the relaxation constant estimation in this example might be hindered by the low abundance of the phenylalanine in the collagen and the consequential low signal-to-noise ratio observed in the NMR spectra for both glycated and unglycated materials (Figure 7.13.), despite the isotopic enrichment of the Phe amino acid. Furthermore, the relaxation measurements were performed using CP based experiment rather the double quantum filtered experiment, and therefore signals corresponding to the natural abundance nuclei are observed.

In the contour plot of the unglycated sample, the signals are separated into similar regions as it was observed for the reconstituted collagen of the Achilles bovine tendon: into the carbonyl C', Cα's and side chain regions (Figure 7.2.). However, good separation of the aromatic carbon nuclei corresponding to Phe is observed (125 – 140 ppm). The calculated ^{13}C T_1 relaxation constants for the glycated sample are higher and with a greater dispersion compared to the unglycated sample. The ^{13}C T_1 signals corresponding to the natural

abundance side chain carbons (25 – 15 ppm) observed in the unglycated ILT contour plot show increased T_1 values for the same chemical shift range in the glycated sample. The signals in the region of the 15 – 70 ppm show more homogeneous distribution of the observed signals in the ILT contour plot for the glycated sample.

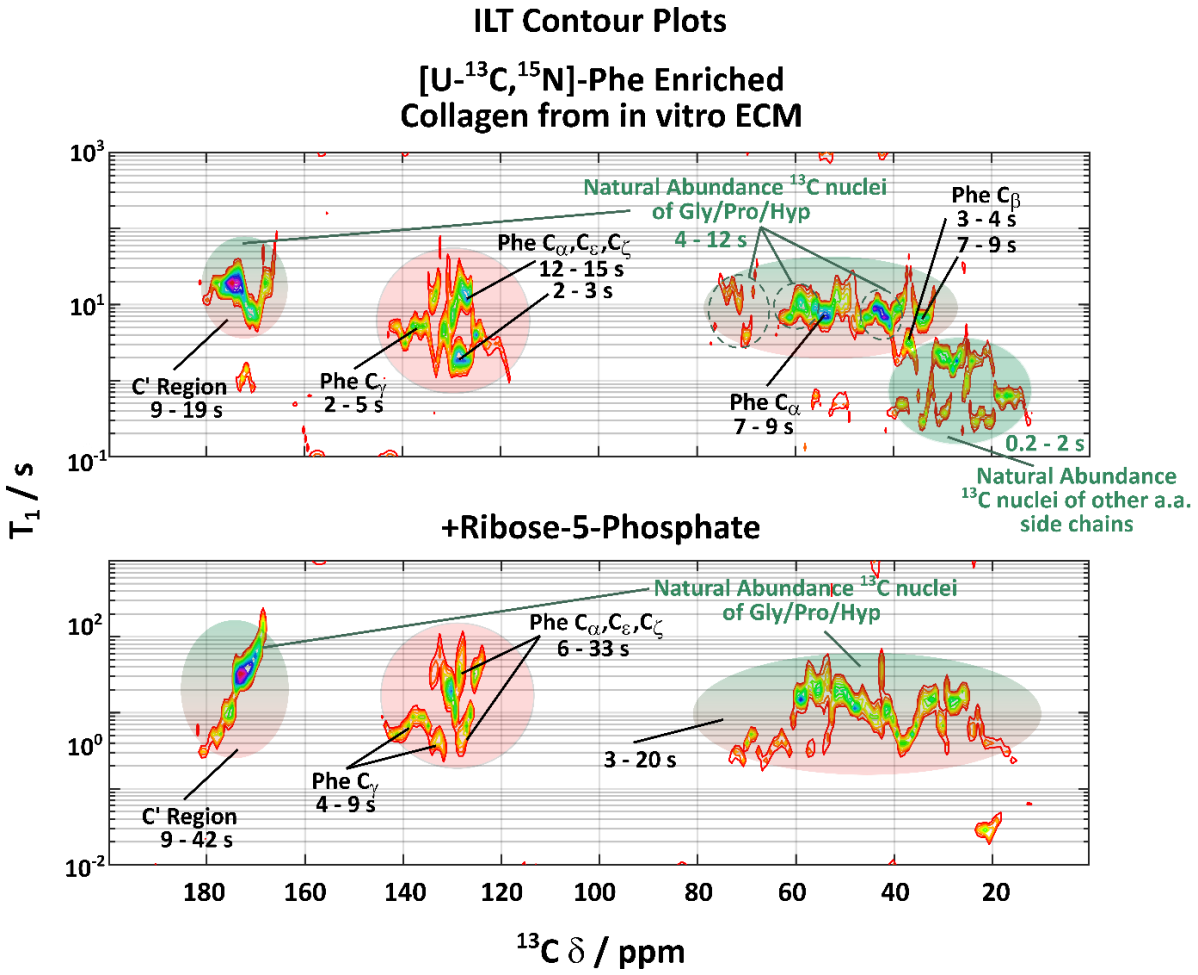

Figure 7.12. ILT contour plots of ^{13}C Phenylalanine T_1 for normal (top) and glycated (bottom) [U-^{13}C,^{15}N]-Phe enriched collagen from in vitro sheep osteoblasts. Chemical shift regions corresponding to the natural abundance ^{13}C nuclei are highlighted in light green. Regions coloured in both orange and green contain a mix of ^{13}C-Phe carbon atoms and nuclei corresponding to the natural abundance signals (see Figure 7.13. of ^{13}C 1D CP NMR spectra of these both samples). Both relaxation data sets were processed using the same ILT parameters in order to be comparable to one another. *(a.a. – amino acids).

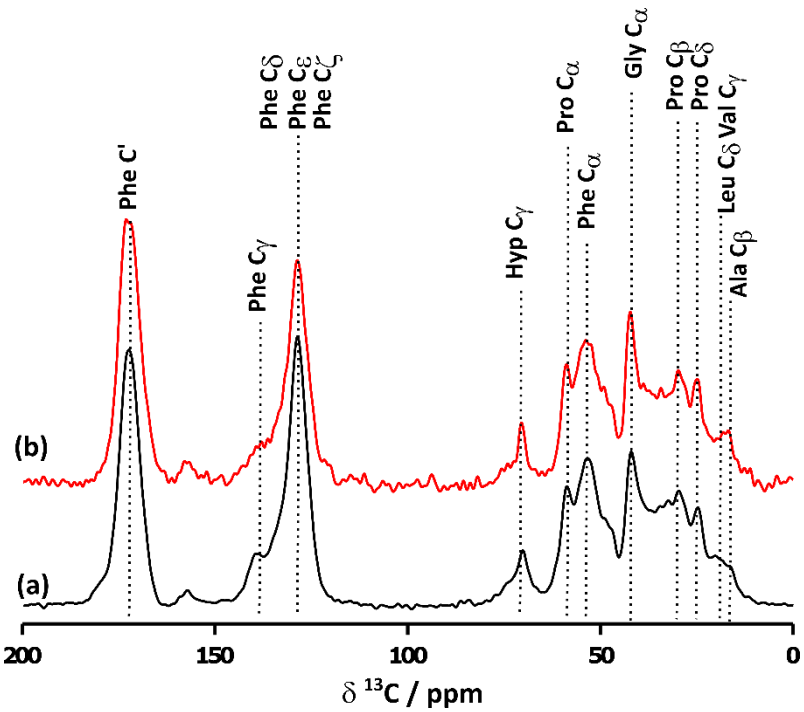

Figure 7.13. 1D ^{13}C CP MAS of (a) normal and (b) glycated ECM collagen, enriched with [U-^{13}C,^{15}N]-Phenylalanine. Signals corresponding to leucine, glycine, proline and hydroxyproline are assigned. Both spectra were acquired with 4096 number of scans.

7.3. Amino Acids that are Directly Involved in Glycation Chemistry

7.3.1. U-^{13}C, ^{15}N-Lysine Labelled in vitro Collagen

Lysine residues are highly conserved in collagen type-I.[210] Hydroxylation of this residue takes place during the collagen biosynthesis. However, the extent of the lysine hydroxylation can vary from 15 – 90 % depending on the tissue type.[211] Both lysine and hydroxylysine are direct targets of glycation due to their sidechain amine group (more details in Section 2.4.). Figure 7.14. shows lysine amino acids highlighted in the collagen type-I mammalian consensus sequence.

The 1D ^{13}C POST C7 and ^{15}N CP-MAS NMR spectra of the [U-^{13}C,^{15}N]-Lys enriched normal and glycated in vitro collagen from sheep osteoblasts are shown in Figure 7.15. The chemical shifts of the ^{13}C Lys nuclei do not change after glycation; however, Hyl Cε and Cδ are not that well resolved after glycation. There is a new nitrogen signal observed in the ^{15}N 1D NMR spectra at 44 ppm, that comes from glycation products on the Lys sidechain. The chemical

shift of the Lys backbone nitrogen and the sidechain has not changed with glycation; however, broadening of the signal corresponding to the backbone nitrogen is observed.

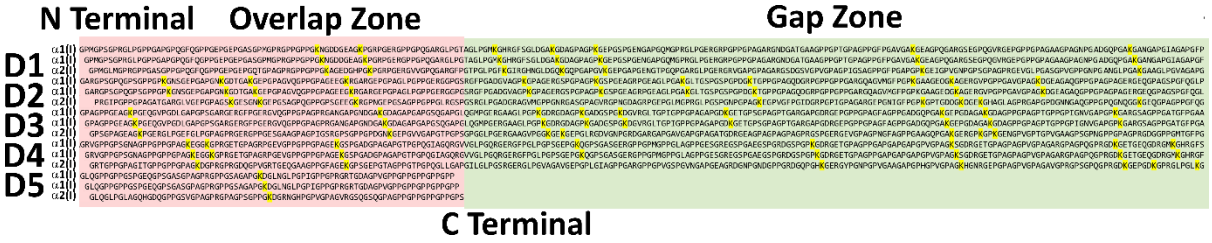

Figure 7.14. The consensus sequence for mammalian collagen type-I with Lys residues is highlighted in yellow. The gap and overlap zones are highlighted in light green and red, respectively.

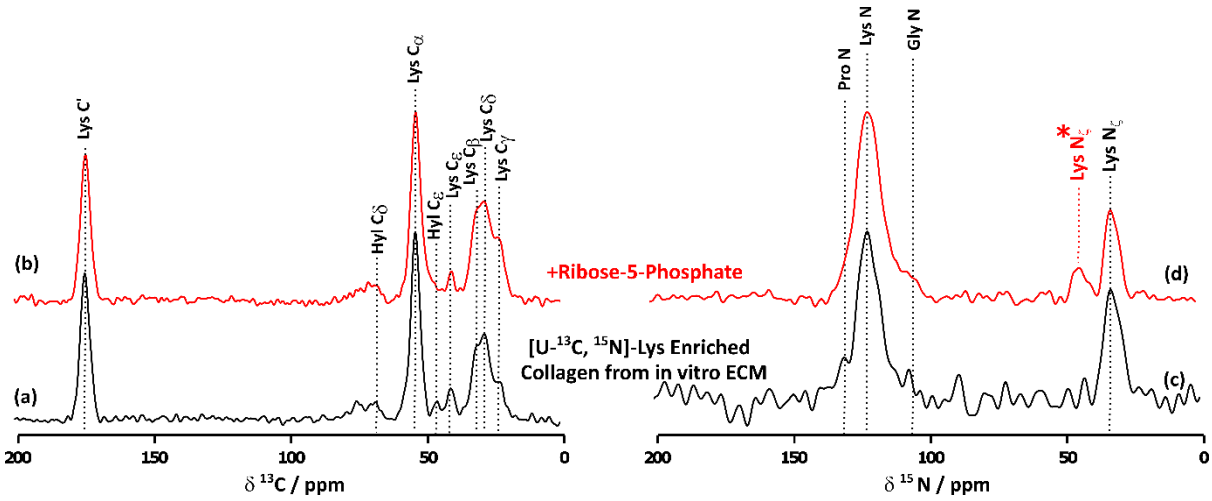

Figure 7.15. 1D ^{13}C POST C7 MAS NMR spectrum of (a) normal and (b) glycated in vitro osteoblast collagen enriched with [U-^{13}C,^{15}N]-Lys. (c) – (d) ^{15}N CP MAS NMR spectra of the same materials. In (d) a new ^{15}N signal (in red, Lys *Nζ) is observed at 44 ppm corresponding to glycation products involving the Lys Nζ side chain nitrogen. ^{13}C and ^{15}N spectra were recorded with 2046 and 5120 number of scans, respectively. Contact times of 2.5 and 3.5 ms for ^{13}C and ^{15}N NMR experiments, respectively, were employed.

The contour plots show a separation of two T_1 regions across the whole chemical shift range for both unglycated and glycated samples highlighted in Figure 7.16. We speculate that the two regions correspond to two different Lys residue types present the collagen fibrils: either different Lys in the gap and overlap zones, or different Lys between the interior and exterior of the collagen fibrils. As Lys is a charged hydrophilic residue, they are expected to be strongly bound to surrounding water. Additionally, Lys residues are also known to form stabilising salt bridges with negatively charged amino acids (Asp and Glu).[78,212] The charge-charge interactions are more likely to be present in tightly packed regions, i.e. in overlap zones, as well as within the interior of the collagen fibrils. The gap zones have lower molecular packing where the space around the collagen helices can be filled with more water compared to the overlap zones.[22,43,213] As a result, the Lys on the surface and gap zones are exposed to more

water. The resulting ^{13}C T_1 relaxation of the carbon nuclei will be affected by the mobility of the residues as well as its surroundings. The Lys on the surface and in the gap zones are likely to be more mobile, therefore the resulting ^{13}C T_1 of these nuclei are shorter. Furthermore, ^{1}H-^{13}C heteronuclear dipolar interaction between the Lys and the surrounding water will influence the relaxation rate of the observed ^{13}C nuclei resulting in a lower T_1 constant.

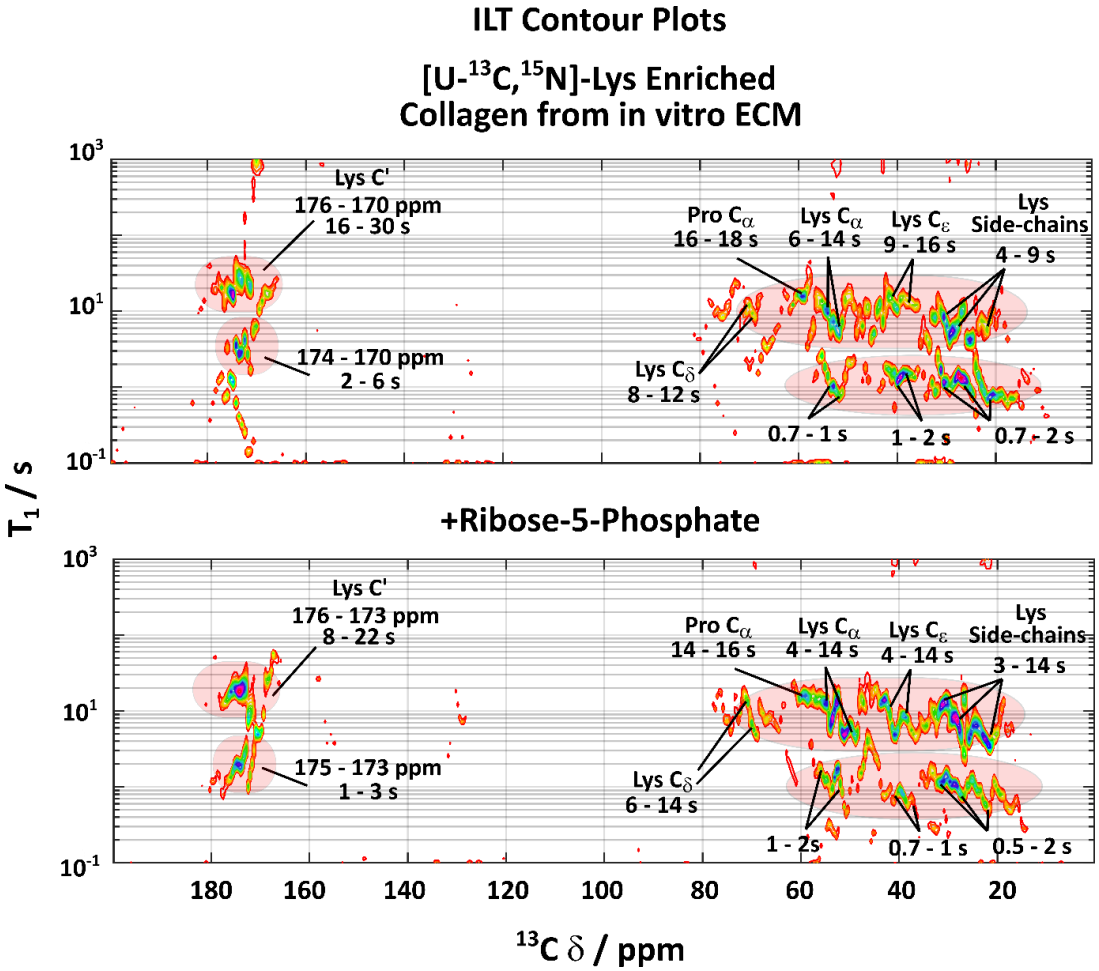

Figure 7.16. ILT contour plots of ^{13}C Lysine T_1 for normal (top) and glycated (bottom) [U-^{13}C,^{15}N]-Lys enriched collagen from in vitro sheep osteoblasts. Both relaxation data sets were processed using the same ILT parameters in order to be comparable to one another.

The ^{13}C T_1 relaxation constants for the glycated sample, in general, are slightly lower and show a greater dispersion of T_1 values compared to the unglycated sample (Figure 7.16.). The greater dispersion of the relaxation constants observed for the glycated sample can be attributed to the effect of molecular disordering, similar as it was observed in the Leu ^{13}C T_1 relaxation results of the side chain carbons (Cδ, Cγ, Figure 7.8.). Alternatively, the dispersion of these Lys ^{13}C T_1 constants could be related to the formation of many Lys-glycation products that each have slightly different relaxation behaviours. Due to the low signal-to-noise ratio in the ^{15}N NMR spectrum, relaxation measurements on ^{15}N were not carried out.

7.3.2. U-^{13}C, ^{15}N-Arginine Labelled in vitro Collagen

Arg is also known to undergo glycation reactions in an analogous manner to Lys residues. Figure 7.17. exhibits a distribution of the Arg amino acids highlighted in the collagen type-I mammalian consensus sequence.

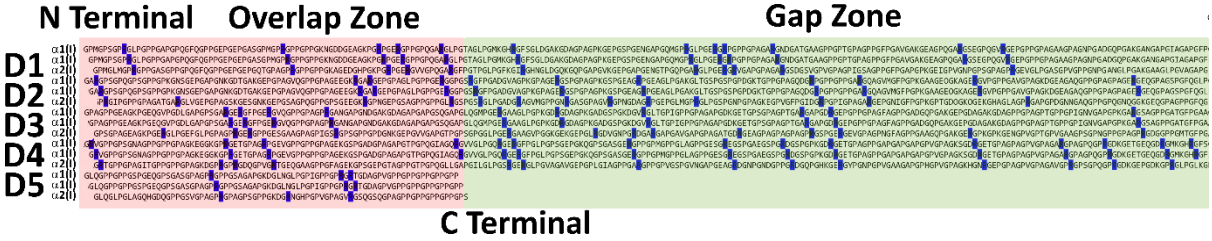

Figure 7.17. The consensus sequence for collagen type-I with Arg residues highlighted in the blue. The gap and overlap zones are highlighted in light green and red, respectively.

The ^{13}C NMR spectra of unglycated and glycated [U-^{13}C,^{15}N]-Arg labelled in vitro collagen from sheep osteoblasts samples shows minimal changes of the recorded chemical shifts (Figure 7.18.). Broadening of the Cδ and Cζ are observed in the glycated sample, accompanied by a low-field shift of 0.5 ppm for the Cδ carbon in the glycated sample. Chemical shift changes are also observed in the 1D ^{15}N NMR spectrum, where the backbone nitrogen has a higher chemical shift by 2 ppm for the glycated sample. Additionally, a new broad, small signal is observed in the ^{15}N spectrum for the glycated sample between 97 – 102 ppm, corresponding to glycation products involving the Arg *Nε and *Nη side chain nitrogen nuclei.

The ^{13}C T_1 ILT contour plots of U-^{13}C-Arg enriched samples in Figure 7.19. show similar separation of two differently relaxing regions of the carbons in range of 20 – 160 ppm, as it was observed in the U-^{13}C-Lys labelled samples (Figure 7.16.). These differently relaxing regions could be assigned to Arg being in either the gap and overlap zones, or interior and exterior of the collagen fibrils. Arg, the same way as the Lys, is a positively charged hydrophilic residue therefore the same discussion regarding the possible separation of the T_1 constants applies here as in previous section. The carbonyl region in the ILT contour plots are partially resolved. The signal regions observed in the contour plots of the ^{13}C-Arg labelled samples show to be more homogeneous and broader compared to the ^{13}C-Lys signals in the ILT plots. Overall, the ^{13}C T_1 relaxation constants are lower in the glycated sample and with slightly broader distribution, as it was observed for the glycated samples of the ^{13}U-Lys and Leu samples.

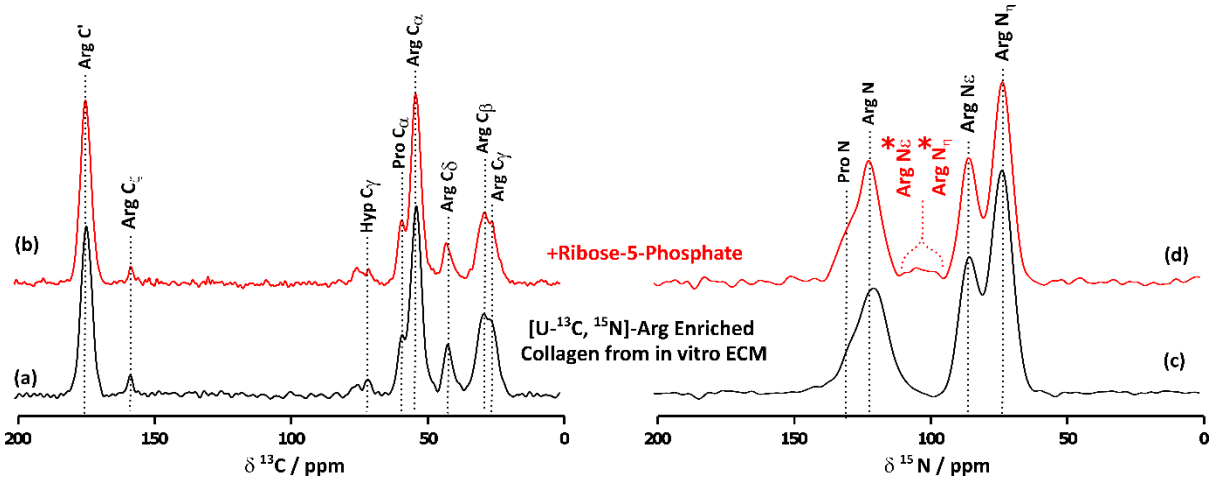

Figure 7.18. 1D ^{13}C POST C7 filtered CP MAS NMR spectra of (a) normal and (b) glycated with ribose-5-phosphate collagen enriched with [U-^{13}C,^{15}N]-Arg from in vitro sheep osteoblasts. In (c) – (d) 1D ^{15}N CP MAS spectra of the same samples shown in (a) – (b), respectively. In (d), a new ^{15}N signal (in red, Arg *Nε and *Nη) is observed between 97 – 102 ppm corresponding to glycation products involving the Arg *Nε and *Nη side chain nitrogen nuclei. ^{13}C and ^{15}N spectra were recorded with 4096 and 5120 number of scans, respectively. Contact times of 2.5 and 3.5 ms for ^{13}C and ^{15}N NMR experiments, respectively, were used. The ^{15}N NMR spectrum of glycated sample in (d) was acquired by Rui Li (University of Cambridge, Chemistry Department).

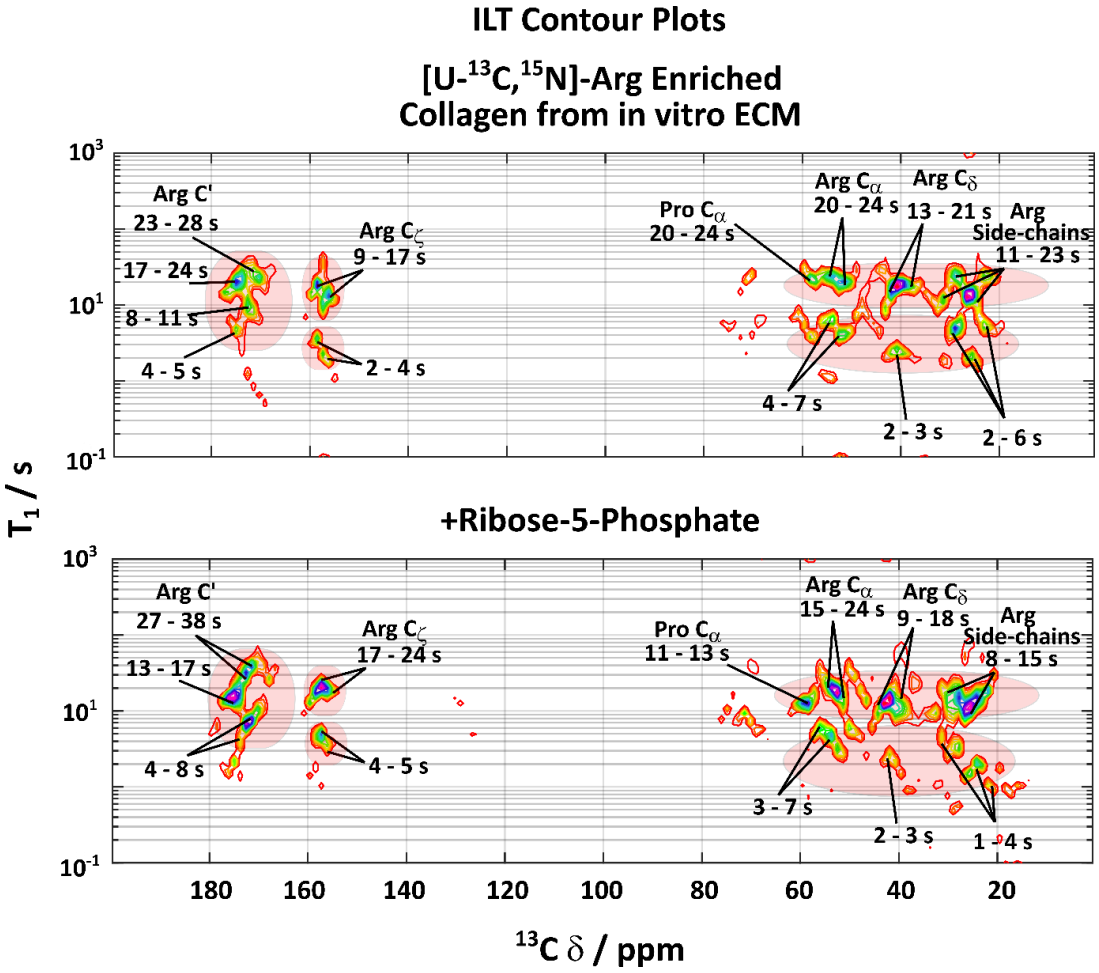

Figure 7.19. ILT contour plots of ^{13}C Lysine T_1 for normal (top) and glycated (bottom) [U-^{13}C,^{15}N]-Arg enriched collagen from in vitro sheep osteoblasts. Both relaxation data sets were processed using the same ILT parameters in order to be comparable to one another.

7.4. Structural and Dynamic Changes in the Glycated Fibrillar Collagen

The observed variations of the ^{13}C and ^{15}N (Pro/Gly samples only) T_1 relaxation results in the glycated samples show that the collagen molecular dynamics changes are likely due to alterations of the fibrillar structure. The residues directly affected by the glycation chemistry (Lys and Arg) show an overall decrease in the T_1 relaxation constants of ^{13}C, as well as a and greater dispersion in these values, compared to the unglycated samples (Figures 7.16. and 7.19.). We infer that the broader distribution of the ^{13}C T_1 values of the Lys and Arg side chains result from the formation of the glycation products that each have slightly different conformations. The ^{15}N NMR spectra of these samples revealed new signals corresponding to the glycation products involving the Lys and Arg side chains (Figure 7.15. and 7.18.). However, the relative intensities of the new ^{15}N signals show that the overall amounts of the glycation products are low. No chemical shift changes were observed for the ^{13}C and ^{15}N nuclei of the backbone atoms for Lys labelled samples, indicating that the glycation has little effect on the local conformational changes at Lys. In contrast, the ^{15}N chemical shift of the Arg residue in the backbone increased to a higher frequency range, towards values that are consistent with a random coil conformation,[214] indicating more extended time-averaged side chain conformations.[137] Additionally, an increase in the chemical shift of the backbone nitrogen was observed for Pro, Gly and Leu labelled samples (Figures 7.3. and 7.7.), showing that the molecular conformation changes with glycation, even at sites that are not involved in the glycation. The most significant change in the ^{13}C NMR chemical shifts were observed for the Phe carbons in the glycated sample, where the Cα chemical shifts were shifted to a higher frequency range accompanied with a signal broadening, and the carbonyl C' shifted to a lower chemical shift (Figure 7.10.). Phe residues in the native sequence were typically found in triplets together with Arg residues that are recognition sites for the integrin binding (G**F**OGE**R**). This result suggests that the integrin binding site after glycation changes its conformation, where this further can weaken the cell adhesion.[204] The relaxation results and the chemical shift changes suggest that the glycation influences overall molecular conformation, even when the extent of glycation is low. In addition, molecular disordering is also observed under TEM imaging of the glycated collagen fibrils, where the loss of the definition of the sub-bands is observed (Figure 7.20.). Signs of degradation of the collagen fibrils were observed (Figure 7.20. red arrows). We infer this to be due to the unravelling of the collagen fibrils.

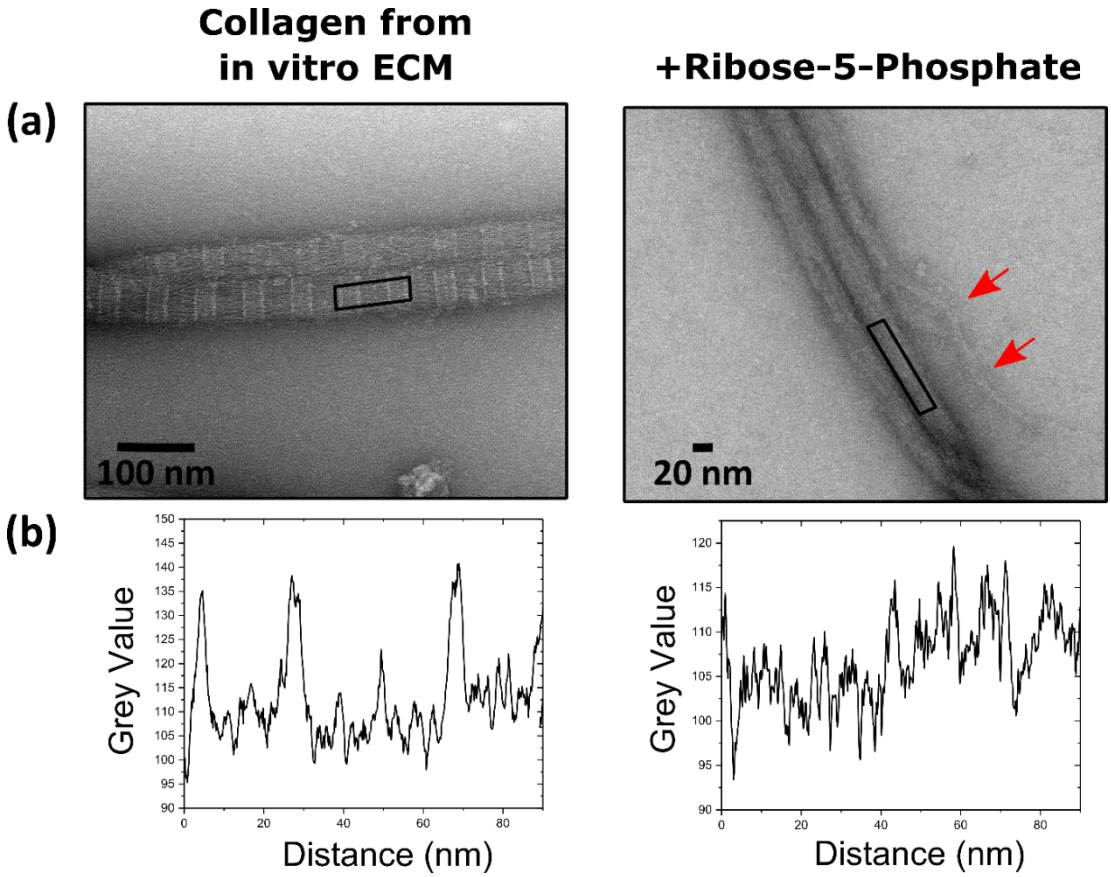

Figure 7.20. (a) Bright-field TEM images of negatively-stained (uranyl acetate) collagen from in vitro sheep osteoblasts. Left: TEM image of unglycated; right: glycated collagen fibrils. The characteristic gap and overlap zones are clearly visible in both examples, where gap zones appear dark grey in colour. Sub-banding is visible on the unglycated (left) collagen fibrils. The sub-banding definition is lost on the glycated fibrils. (b) The densitometry plots for the periods with highlighted with black rectangles for the unglycated (left) and glycated (right) fibrils. These plots represent intensity profiles across the fibril where the maximum of the signal corresponds to the whitest part of the TEM image, and the minimum to the darkest part. These plots also show the collagen fibril sub-bands. The red arrows indicate potential 'shredding' of the collagen fibrils after glycation.

The ^{13}C T_1 relaxation constants of the amino acids that are not directly involved in the glycation chemistry showed a great dispersion of the T_1 values, especially for the carbon nuclei in the Leu residue corresponding to the side chains (Cδ, Cγ, Figure 7.8.). This shows that the side chains acquire a greater variety of environments after glycation and that many of them are much less closely packed (more mobile) than in unglycated collagen fibrils. Overall, the dominant effect of glycation on collagen is not the formation of new collagen cross-links but rather substantial changes in collagen protein backbone and side-chain conformations. We hypothesise that glycation attracts more water into the fibril due to the disrupted molecular ordering and the addition of highly hydrophilic glycation species. The increase of the water content within the fibrils is expected to make them stiffer. The increased level of water in glycated collagen fibrils is confirmed by ^{1}H NMR spectra of glycated samples compared to those for unglycated samples (Figure 7.21.). The observed water signal is much stronger in

intensity for the glycated sample suggesting a higher content of water, despite the sample being freeze dried for at least 48 hours. The water ^{1}H signal also has a narrower component as well as a broad component from protein-bound water. The narrower water ^{1}H signal comes from more mobile water molecules and this signal suggests that glycated collagen fibrils contain a substantial amount of more mobile water, that nevertheless, is not readily removed by freeze drying.

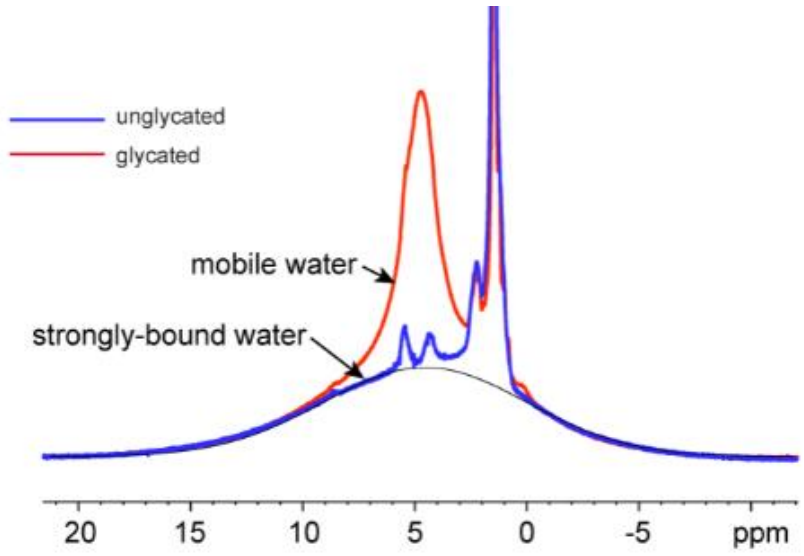

Figure 7.21. The 1D ^{1}H NMR spectra of the normal (blue) and glycated (red) reconstituted type-I collagen. Water signals highlighted in both. Spectra were recorded by Rui Li (Chemistry Department, University of Cambridge).

The observed increase in water content further explain why most of the observed T_1 relaxation constants are lower. The surrounding water molecules can facilitate the relaxation of the ^{13}C via proton spin diffusion (PSD). The ^{13}C nuclei are strongly coupled to their bound protons, and these protons can be further affected by the surrounding water molecules through PSD (same concepts of PSD are observed on the ^{13}C as on the ^{15}N nuclei, see previous Chapter, Section 6.3.1.).

To summarise, the ^{13}C relaxation studies have shown that upon glycation, the collagen fibrils become more heterogeneous as evidenced by the great dispersion of the T_1 constants not only in the amino acids that are involved in the glycation chemistry, but also on other residues. These results highlight that the dynamic studies employed here can be used to assess the molecular changes of the normal and diseased collagen fibrils.

8

Towards Spatial Resolution of Collagen Fibril NMR Spectra

Collagen fibrils, despite their representation as long rod-like structures, have complex tertiary architectures. The molecular arrangement within collagen microfibrils defines their surface chemical functionality that is important for interactions with other ECM macromolecules and ligand binding that is vital for biological processes underlying proper cell function.[24,31,32] The fibril surface must be highly functionalised to enable recognition and binding by various ECM proteins, such as integrins, osteonectin, fibronectin, fibrillin and other proteoglycans, which confer specific properties on different tissues. However, the arrangement of functional groups on collagen fibril surfaces remains unknown. X-ray diffraction, AFM and TEM studies provide a static structural snapshot of collagen fibrils, though they cannot provide information on dynamic processes such as protein binding. Work in the previous chapters has shown that NMR spectroscopy can give structural and dynamic insights at atomic-scale resolution of the collagen molecules. Nevertheless, even these studies have limitations. Synthetic collagen

model peptides, with sequence-specific enrichment, enables one to make detailed inferences about conformations of specific residues, but only at the triple-helical, not the fibrillar level. Isotopic enrichment of in vitro produced ECM collagen can give highly-uniform isotopic enrichment of fibrils. This method however is under the control of cell metabolism, and site-specific enrichment is not possible, limiting more detailed structural and dynamic studies. Therefore, there is still a need to develop methods that would, allow us to differentiate between interior and exterior, or gap and overlap zones or even target specific binding sites of the fibrils. This chapter investigates the use of paramagnetic probes to differentiate different fibrillar regions and environments.

The unpaired electron(s) of paramagnetic metal ions interact strongly with nuclear spins of collagen which are close to their binding site(s) which enhances the relaxation rates of these nearby nuclei.[23,215,216] The regions of the fibrils which these probes target can be visualised by TEM imaging; we hypothesised that knowing the distribution of each paramagnetic species may enable the relaxation enhancements to be ascribed to specific fibril regions.

8.1. Paramagnetic Targeting of the Collagen Fibrils

The use of paramagnetic probes has increased in the last two decades due to the advantages they provide in structural and dynamic studies of proteins.[23,215,216] Paramagnetic species possess one or more unpaired electrons, the magnetic moments of these species are around three orders of magnitude greater than typical nuclear magnetic moments. Consequently, these paramagnetic species generate large local magnetic fields resulting in strong electron-nuclear dipole-dipole interactions. The paramagnetic ions may significantly change both the chemical shifts and relaxation rates of nuclei in their proximity. The mechanism of interaction is similar to nuclear spin-spin interactions; however, due to the size of the electron spin magnetic moment, the interactions between the unpaired electron and nuclear spins will dominate over nuclear spin-spin interactions. Paramagnetic ions fall roughly into two groups: paramagnetic relaxation enhancement (PREs) probes and paramagnetic chemical shift (PSCs) probes according to the NMR property they mainly affect. In solid-state NMR spectroscopy, paramagnetic tags have been used to study the structures of proteins and their molecular interactions, as well as providing fast acquisition of NMR data, through shortened relaxation times to help increase signal-to-noise (S/N) ratio in spectra.[23,215–217]

Paramagnetic ions can be introduced to proteins in various ways. These include substituting a diamagnetic ion coordinating to the protein with a paramagnetic ion with similar properties in metalloproteins,[218,219] or synthetically by covalently attaching a paramagnetic tag to

cysteine[220] or lysine[221] residues. Such tags for the latter method include nitroxide, EDTA-M, TETAC-M (M = paramagnetic metal centre) and others. However, in the case of collagen fibrils, the lack of cysteine residues limits this approach; lysine sidechain modification will not generate a regionally specific and reproducible tagging strategy, as the distribution of the Lys residues is more or less uniform across the collagen fibril surface (see Figure 7.14.). An alternative method used in material studies is paramagnetic "doping", where the sample is directly mixed with the paramagnetic complex affording sensitivity gains in natural abundance ^{13}C and ^{15}N NMR.[138,222,223] In this work, a "doping" approach is used in situ, combined with TEM imaging to visualise where the paramagnetic complexes have accumulated within the sample. Moreover, paramagnetic metal ions have found their way into DNP-NMR in recent years where Gd(III) and Mn(II) complexes have been used as DNP polarizing agents.[224]

The main objective of this work is to establish a new methodology which can be used to identify and differentiate between the chemical functionalities of the surface and the interior of collagen fibrils, as well as gap and overlap zones. Here, we explore the effects of various paramagnetic probes on the ^{13}C nuclei NMR relaxation of collagen fibrils. We test paramagnetic probes that are known to cause relaxation enhancement, such as Cu (II) and Gd (III) based complexes and we investigate if paramagnetic complexes of the lanthanide ions Pr (III), Eu (III) and Dy (III) can be used as chemical shift reagents. All these paramagnetic probes were chelated with either pentetic acid or diethylenetriaminepentaacetic acid (DTPA) ligands, except Cu (II) which was employed as the chloride salt. The general structure of the lanthanide complexes are shown in Figure 8.1. The chelated complexes were adopted to avoid unwanted interactions that would lead to precipitation of the lanthanides as hydroxides or phosphates, and to avoid strong interactions between the free metal ions and collagen fibrils that would lead to unwanted modifications or even degradation of the fibrils. The chelation of the metal centres reduces the overall paramagnetic effect due to increased spatial separation between the observed nuclei and the paramagnetic centre. Chelation further decreases chemical exchange with the surrounding water molecules that would introduce strong line broadening that could result in no observable signals from the protein (see Section 7.1.1.).[225]

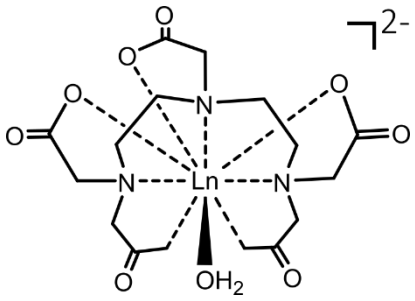

Figure 8.1. General structure of the lanthanide (III) DTPA dihydrogen salt complexes. In this work, Ln = Pr (III), Eu (III), Gd (III) and Dy (III). The DTPA moderates the relaxation enhancement of the lanthanide (III) ion, because the organic shell prevents the close approach and thus direct binding of the lanthanide ion to the collagen fibril. The minus two charge carried by the DTPA acts as a probe for positively charged residues on the collagen fibril surface.

8.1.1. Paramagnetic Enhancement Relaxation Probes

The relaxation of the nuclear spin arises due to fluctuating magnetic fields from surrounding nuclei and electrons interacting with the observed spin. When a paramagnetic ion is introduced into the system, this additional large fluctuating magnetic field will influence further nuclear interactions significantly, and therefore the T_1 and T_2 relaxation times of the observed nuclei will be affected as well. The contribution from the paramagnetic centre depends on three terms: the electron relaxation (T_{1e}, T_{2e}) , molecular tumbling, and chemical exchange between the paramagnetic probe and the surrounding water molecules.[23,226,227] For solids, the molecular tumbling and chemical exchange terms can be neglected; therefore, the electron relaxation will dominate and will be the strongest contributor to the T_1 and T_2 nuclear relaxation. The electronic relaxation times of the paramagnetic ions depend on their electron shell configurations, occupancy and symmetry of the atomic orbitals that the unpaired electrons occupy. T_{1e} and T_{2e} are in the range of 10^{-8} to 10^{-13} s. T_1 and T_2 of the surrounding nuclei will depend inversely on the sixth power of the distance between the observed nuclei and paramagnetic ion, where we have assumed that the electron-nuclei interactions dominate the nuclear relaxation.[228–230] Due to this distance dependency, these paramagnetic probes can provide information about the metal ion – nuclear distances within a spherical shell of limited radius around the paramagnetic centre, as shown in Figure 8.2. The ‘effect radius’ will depend on the relaxation capability of the metal ion, which depends on the number and orbital configurations of the unpaired electrons, T_{1e}, T_{2e} values, and on the rotational correlation time of the molecule.[218] The understanding of the electron relaxation and its effects on the nuclear spin relaxation is a challenging field, especially in the solid-state.[23,218]

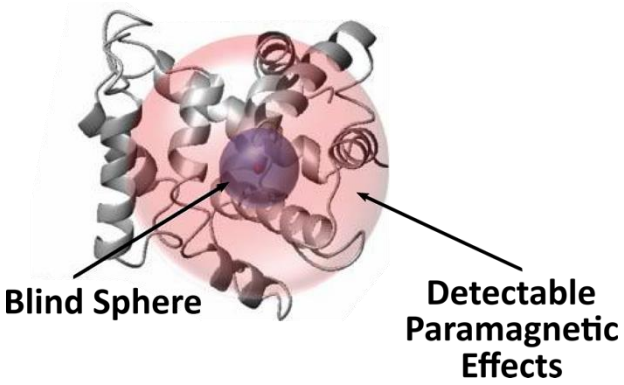

Figure 8.2. A representation of the effects of PREs effects on a protein. In the 'blind-zone', shown in purple, proton NMR signals will be too broad to be detected. Outside the 'blind-sphere' there is another sphere, coloured in light pink, where paramagnetic effects will be detected on all NMR-active nuclei. These effects decrease with increasing distance. The figure was adapted from reference [215].

One of the advantages of using the PRE probes is reduction of the nuclear longitudinal relaxation time leading to more rapid experimental repetition rates and increased S/N ratio for the NMR experiment. The NMR linewidth broadening is related to T_2:

$$\Delta\nu_{1/2} = 1/(T_2\pi) \tag{8.1.}$$

Therefore, shortening of T_2 through the paramagnetic effect will lead to broader linewidths of the signals that are close to the paramagnetic metal ion. PREs can be used for spectral editing if the paramagnetic tag location is known, and the extent of the line broadening can indicate relative distances between the electron and the nuclei.

In this work a number of paramagnetic probes are investigated. Cu (II) and Gd (III) are regarded as PREs probes; they have low magnetic susceptibility anisotropy and electron relaxation times around 10^{-8} – 10^{-9} s of the order of the nuclear Larmor frequency and thus will induce relatively large PREs on the surrounding nuclei. Other paramagnetic probes used in this work are generally regarded as PCSs (vide infra) but they do all exert some PRE effects that will be explored here as well.

8.1.2. Paramagnetic Chemical Shift Probes

The PCSs occur due to non-zero averaging of the dipolar coupling between the nuclear magnetic and the average electron magnetic moments. The resulting chemical shifts are separated into two contributions: the through-bond contact shift, which is due to the delocalisation of unpaired electron density from the paramagnetic centre to the directly bonded nucleus, and through-space pseudo-contact shift, due to the anisotropic dipolar interaction

between the nuclei and the unpaired electron.[231,232] In this work, we use DTPA as a chelating ligand that will prevent direct bonding to the free amino acid side chains on the collagen fibrils; therefore, only pseudo-contact shifts will be observed. The induced chemical shift depends on the inverse of the third power of the distance between the nuclei and the paramagnetic metal centre. For example, a ^{1}H chemical shift can be shifted in the range of 1.4 to 38 ppm with distances between the proton and paramagnetic Dy (III) centres of 30 and 10 Å, respectively.[215] Pr (III), Eu (III) and Dy (III) containing complexes are used here as PCSs and tested on the [U-^{13}C,^{15}N]-Lysine collagen grown in in vitro cell culture. Lysine was selected because its positively charged side chain is likely to favour the interaction between the negatively charged paramagnetic lanthanide DTPA complexes and because lysine is of biological interest in collagens due to its involvement in intermolecular crosslinking and glycation.

8.2. Paramagnetic Doping of Collagen Fibrils

TEM imaging of collagen fibrils commonly uses negative staining with uranyl acetate (UA), which gives good image contrast and resolution of the fibrils and fine structure, such as gap and overlap zones and sub-banding regions (Figure 8.3.). The characteristic banding pattern observed under negative staining conditions in TEM images of collagen fibrils is partially due to charged amino acid interactions with the heavy metal uranyl ion, and partially from the spatial distribution favouring less over more bulky residues.[75,76] Higher electron density areas of the TEM image correspond to zones of high metal salt concentration. Negative staining with UA was performed for two types of collagen materials used in this work: commercial ‘pure’ collagen obtained from reconstituted fibrils from bovine Achilles tendon collagen, and in vitro ECM collagen extracted as described in Section 5.5. Figure 8.3. (top) compares TEM images from the two materials. The in vitro collagen fibrils show clear gap (darker) and overlap (lighter) zones; however, further sub-banding is not as clearly observed as for the reconstituted collagen. The same staining protocol was carried out for these samples using $Ln(DPTA)^{2-}$ and $CuCl_2$ solutions (Ln = Pr (III), Eu (III), Gd (III) and Dy (III)) with 2 mM concentration proving optimal. A selection of results with staining using $Pr(DTPA)^{2-}$ and $CuCl_2$ is shown in Figure 8.3. and compared against conventional staining with UA. Staining with other lanthanide complexes resulted in similar images (not shown).

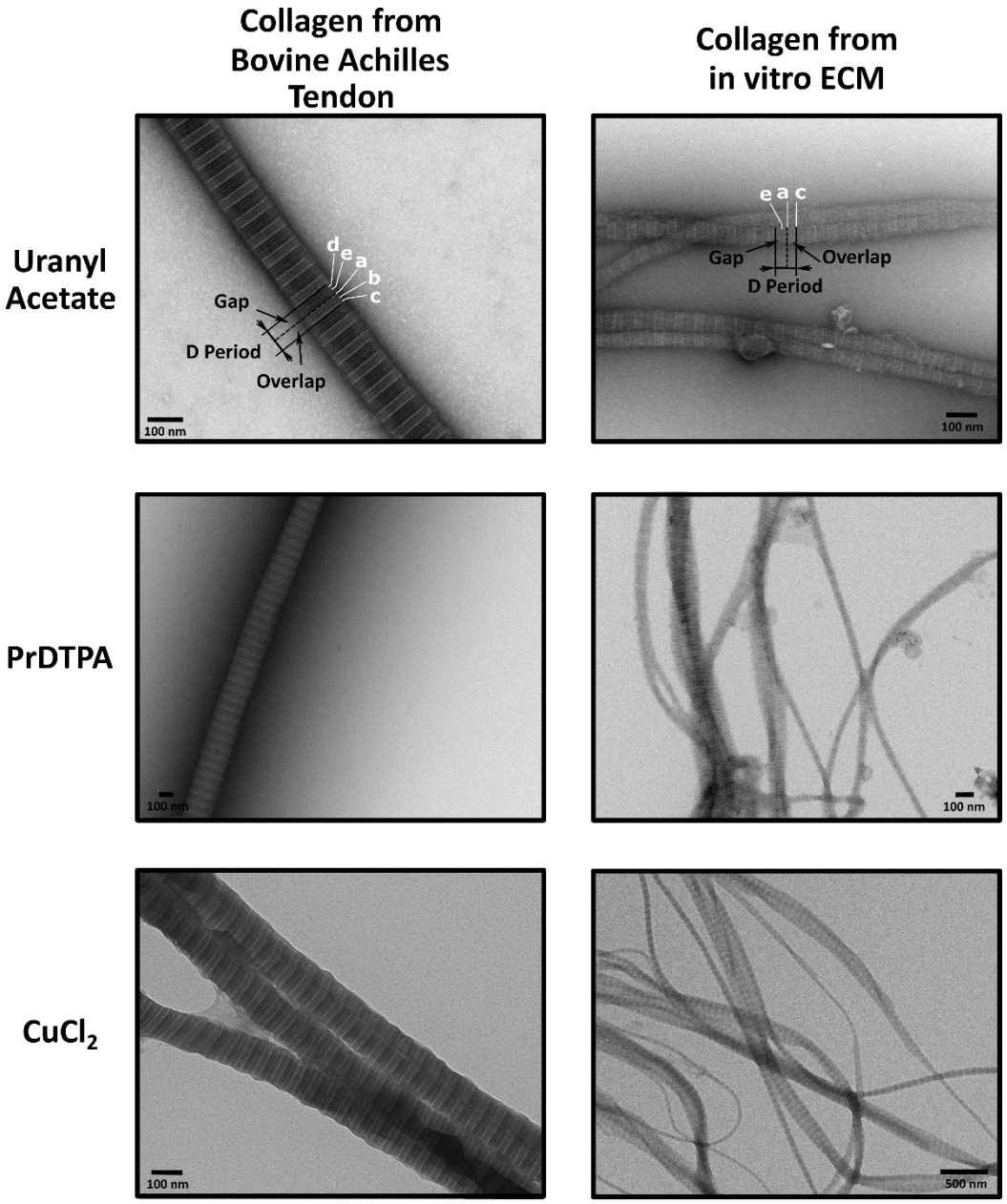

Figure 8.3. Bright-field TEM images of negatively stained collagen from bovine Achilles tendon collagen (reconstituted fibrils) (on the left) and from in vitro sheep osteoblast ECM (on the right). Samples were prepared as described in Section 5.5. Negative staining with uranyl acetate, $Pr(DTPA)^{2-}$ and $CuCl_2$ shown. Concentrations of the $Pr(DTPA)^{2-}$ and $CuCl_2$ were 2mM in water. In the images at the top, sub-banding of the D period are highlighted.

These paramagnetic reagents generate similar staining patterns as UA. The paramagnetic complexes accumulate in the gap zones (higher electron density in the wider, 40nm, band than in the 27 nm band, see Section 2.4.). The UA and Cu (II) complexes are expected to interact with negative surface charges while the $Ln(DTPA)^{2-}$ complexes are hypothesised to interact with positive surface charges. The staining patterns of both charge type probes are very similar indicating no clear segregation of positive and negative charges, consistent with the distribution of the charged residues shown in Figure 8.4.

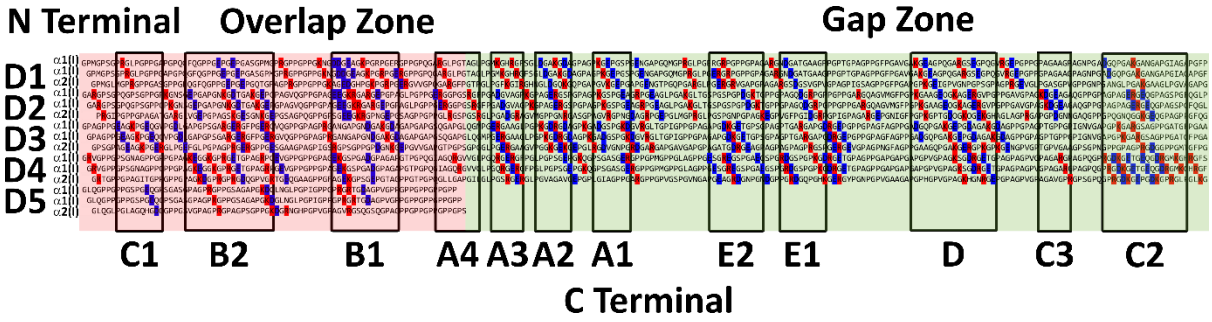

Figure 8.4. The consensus sequence of mammalian collagen type I with the predicted fibrillar sequence alignment. Each DN (N = 1, 2, 3, 4, 5) period consists of three polypeptide chains, two α2(I) and one α1(I) chain, each staggered by one residue with respect to one another. The image here is provided to schematically show positive (red) and negative (blue) charge distribution between the gap and overlap zones, that are highlighted in light green and red, respectively. Further sub-banding of the collagen is indicated by the rectangular boxes as determined by TEM.[231] The sub-banding was derived by Anna M. Puszkarska (Department of Chemistry, University of Cambridge).

Reconstituted bovine Achilles tendon collagen fibrils were incubated in 2 mM paramagnetic solutions in D_2O overnight at 4 °C, washed twice using D_2O, and centrifuged to remove excess paramagnetic species. Samples were then freeze-dried overnight and ^{13}C CP-MAS NMR experiments were performed on these samples, the resulting 1D NMR spectra are shown in Figure 8.5. Samples were prepared using D_2O rather H_2O. The samples prepared with H_2O resulted in excessively broad and unresolvable NMR signals due to the strong proton spin diffusion (especially for $CuCl_2$, $Gd(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$. Cross-polarization conditions for each sample were optimised for S/N ratio, and PRE and PSC effectiveness evaluated.

The ^{13}C 1D CP-MAS NMR signals of collagen incubated with $CuCl_2$, Pr and $Eu(DTPA)^{2-}$ (Figure 8.5. (b) – (d)) appear better defined with sharper signals compared to the reference spectra (Figure 8.5. (a)). On the other hand, the spectra from samples treated with Gd and $Dy(DTPA)^{2-}$ complexes show broader NMR signals as well as a and poorer resolution when the same number of scans are used, with the poorest S/N ratio obtained for sample incubated with $Gd(DTPA)^{2-}$. This observation is because Gd (III) has an isotropic environment of seven unpaired electrons, other paramagnetic metal centres have less, and the distribution is more anisotropic. In other words, for the more symmetrical electronic arrangement the paramagnetic effect will be induced evenly to its surroundings, compared to metal ions with anisotropic populations of the molecular orbitals where the paramagnetic effect will depend on the orientation of the molecular orbitals with respect to its surroundings, i.e. surrounding nuclei and the applied magnetic field of the NMR experiment

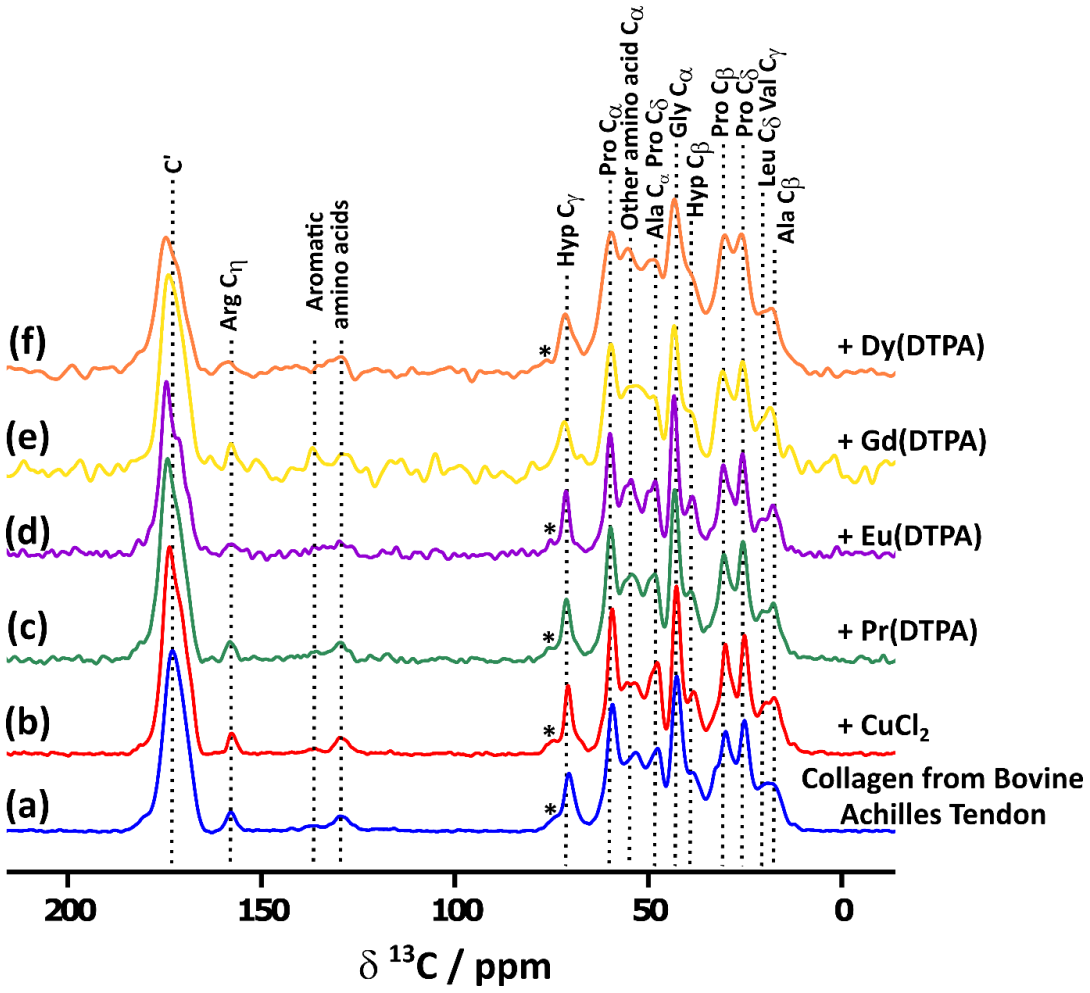

Figure 8.5. 1D ^{13}C CP-MAS proton decoupled NMR spectra of reconstituted bovine Achilles tendon collagen fibrils incubated with (a) D_2O, (b) $CuCl_2$, (c) $Pr(DTPA)^{2-}$, (d) $Eu(DTPA)^{2-}$, (e) $Gd(DTPA)^{2-}$ and (f) $Dy(DTPA)^{2-}$. The concentration of all paramagnetic solutions was 2 mM prepared in D_2O. NMR spectra were recorded using the most optimal CP conditions; acquisition delays were the same across all the samples, which was 2 seconds. The most prominent signals are assigned. Spinning sidebands are highlighted with an asterisk.

Other lanthanide ions used in this work have shorter T_{1e} and T_{2e} relaxation times compared to the Gd (III) and Cu (II) ions. For Dy (III), the T_{1e} is ca. two orders of magnitude shorter, whereas Pr (III) and Eu (III) both have three orders of magnitude shorter electron spin relaxation times. All these electron spin relaxation times are much faster than the nuclear Larmor precession frequency, and as a result these ions will have smaller PRE effects on the surrounding nuclei; and no loss of the resulting ^{13}C nuclei signal intensity in the NMR spectra. As a result, the 1D NMR spectra of the sample incubated with $Pr/Eu(DTPA)^{2-}$ show smallest effect on the ^{13}C NMR signals compared to the samples treated with $Gd/Dy(DTPA)^{2-}$, where the S/N is the poorest from all the samples. Interestingly, there is no apparent effect on the ^{13}C nuclei for the sample treated with $CuCl_2$; the only difference is that the signals in the spectrum of the Cu-treated sample are slightly sharper compared to the reference spectra where there is no paramagnetic doping.

Comparing the chemical shifts of the collagen samples treated with Pr (III), Eu (III) and Dy (III) against those for the reference sample, no significant chemical shift changes were observed (Figure 8.5.). Lack of PCSs effects could be related to the chelating of the paramagnetic ions by DTPA meaning that the distance between the paramagnetic metal centre and the observed nuclei is too large for significant PCS effects to be observed. Additionally, the nuclei that are affected by the Ln (III) ion contribute only low relative signal intensities. This is because of a combination of low concentrations of Ln ions used as well as the loss of the signal intensity due to the strong dipolar interaction between the observed nuclei and the Ln (III) ion. This results in the affected signals being buried under the other non-affected signals.

8.2.1. ^{13}C T_1 Relaxation of Collagen with Paramagnetic Doping

^{13}C T_1 relaxation measurements were performed on the samples discussed in the previous section. The paramagnetic ions will influence the relaxation of the observed ^{13}C and the resulting ^{1}C T_1 relaxation constants will depend upon internuclear distances between ^{13}C nuclei and paramagnetic metal centre as discussed previously. The effects of the different paramagnetic complexes were tested on the collagen samples to see if any separation in the relaxation constants can be observed that would discriminate between different regions of the collagen fibrils. The relaxation data were processed using ILTs to help separate differently-relaxing components (if present); results for all samples are summarised in Figure 8.6. in which T_1 with relaxation constants are indicated.

The ILT contour plot for the reference sample with no paramagnetic doping shows three families of differently relaxing atom types: carbonyls, Cαs together with Pro sidechains, and other amino acid sidechains as highlighted in the Figure 8.6. The paramagnetic probes are expected to be present on the surface of the collagen fibrils, and more concentrated in the gap zones, as observed in the TEM images previously. Thus, one would expect to see two relaxation constants per each differently relaxing region, one corresponding to nuclei in the overlap zone and one corresponding to nuclei in the gap zone. However, this was not observed in these ILT contour plots. The obtained relaxation constants, in general, are shorter for all relaxation components of all the ^{13}C signals indicating that the dipolar interaction with the paramagnetic metal centres is more or less uniform across the whole fibril.

The ILT contour plots of the samples treated with Pr, Eu and Dy(DTPA)$^{2-}$ complexes show the three relaxation regions as observed for the reference sample, with the exception that the proline side-chain carbons no longer have similar relaxation constants as the Cα carbons. The Pro relaxation constants are now lower and similar to those for the other amino acid

side-chains. The relaxation values for these three samples are similar with some small variations.

Separation of relaxation times in the carbonyl region was observed for the sample treated with $CuCl_2$ salt. The carbonyl and Pro C_α carbon signals show a bi-modal relaxation, that could be assigned to differently-relaxing regions as a result of differential Cu^{2+} binding. However, no bi-modal relaxation is observed for the other signals. Samples which contained the $CuCl_2$ salt showed the lowest T_1 values compared to the other lanthanide containing samples. The resulting ILT contour plot of the Cu (II) treated sample shows the greater distribution of the relaxation constants compared to the other samples. The Cu^{2+} ions in contrast to the paramagnetic complexes are not chelated. These free metal ions are therefore expected to penetrate the collagen fibrils and directly interact with the amino acids, potentially throughout the fibril, compared to the 3.5 times larger diameter lanthanide complexes that are more likely to be present on the surface of the fibrils. As a result, the overall observed ^{13}C T_1 relaxation constants are lower for the samples treated with the $CuCl_2$ solutions. The T_1 relaxation data of the sample incubated with $Gd(DTPA)^{2-}$ shows the broadest signals across the T_1 range compared to the other samples, which is due to the poor S/N ratio which result from the strong paramagnetic effect shown in the previous section.

The relaxation results of the lanthanide treated samples show that the proton spin diffusion (PSD) might be affecting the separation of differently relaxing regions. The ^{13}C nuclei are strongly coupled to their bound protons and further these protons can be affected by the surrounding protons within the sample. The proton source can be either collagen fibrils themselves due to low mobility, and further surrounding water molecules. Strong PSD effect can homogenise the observed relaxation, where the observed relaxation constant for different residues can be the same across different protein regions. Furthermore, the PSD effects can be so strong that the overall T_1's of the ^{13}C nuclei appear to be the same across for all carbon types in the sample. Even though the samples were prepared using D_2O, and the samples were lyophilised using freeze-drying, the collagen fibrils still retain high quantities of the strongly bound water within the collagen fibrils (See Figure 7.21.) where the resulting proton network will facilitate the spin diffusion. One way to reduce the spin diffusion would be by performing the experiments at higher MAS frequencies, or reconstitute the collagen using all deuterated solvents; however, such an approach will be excessively expensive and time-consuming.

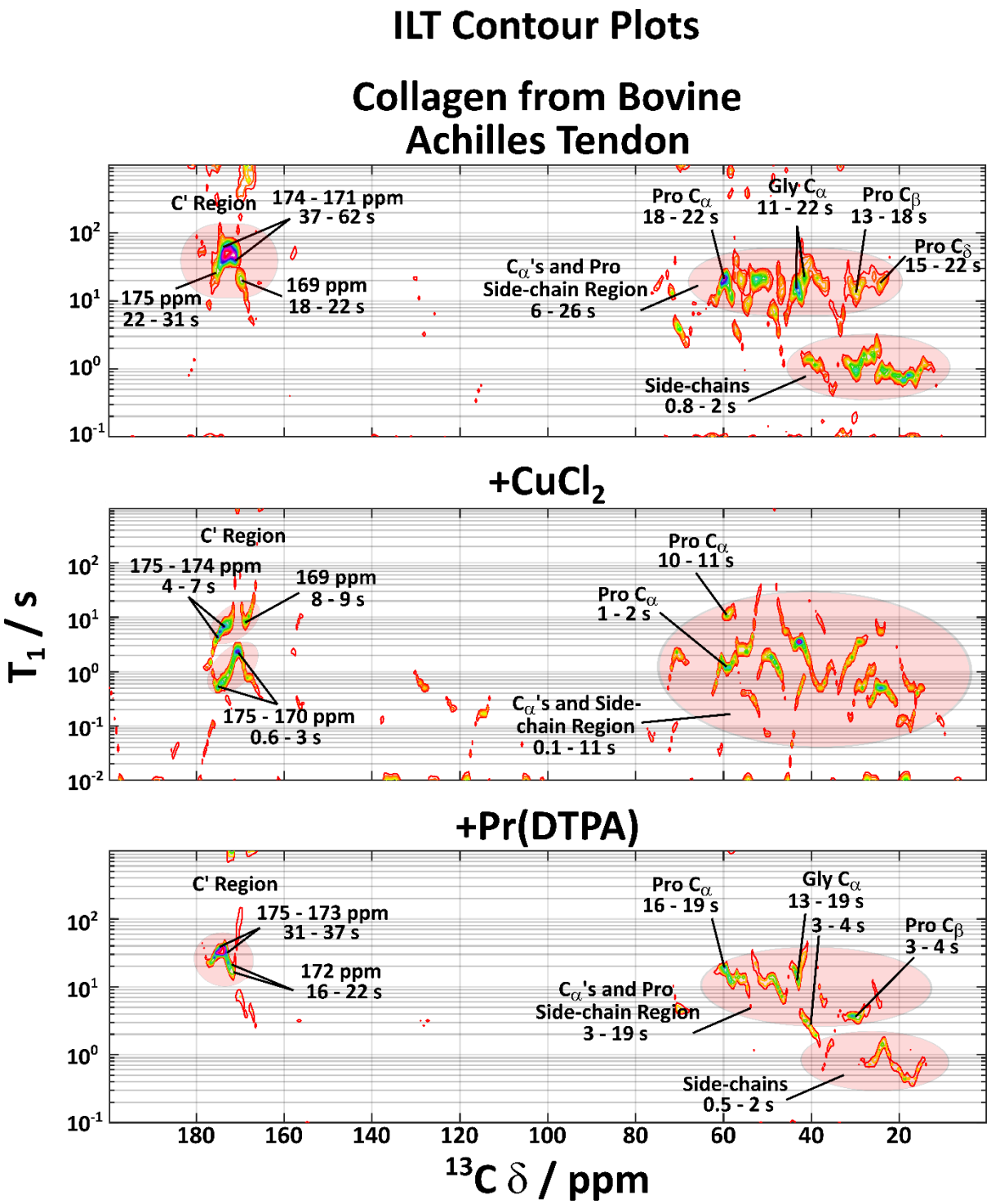

Figure 8.6. ILT contour plots of ^{13}C T_1 relaxation results from top to bottom of collagen, collagen incubated with $CuCl_2$, $Pr(DTPA)^{2-}$, $Eu(DTPA)^{2-}$, $Gd(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$. The number of scans per sample was varied to gain best signal-to-noise ratio possible.

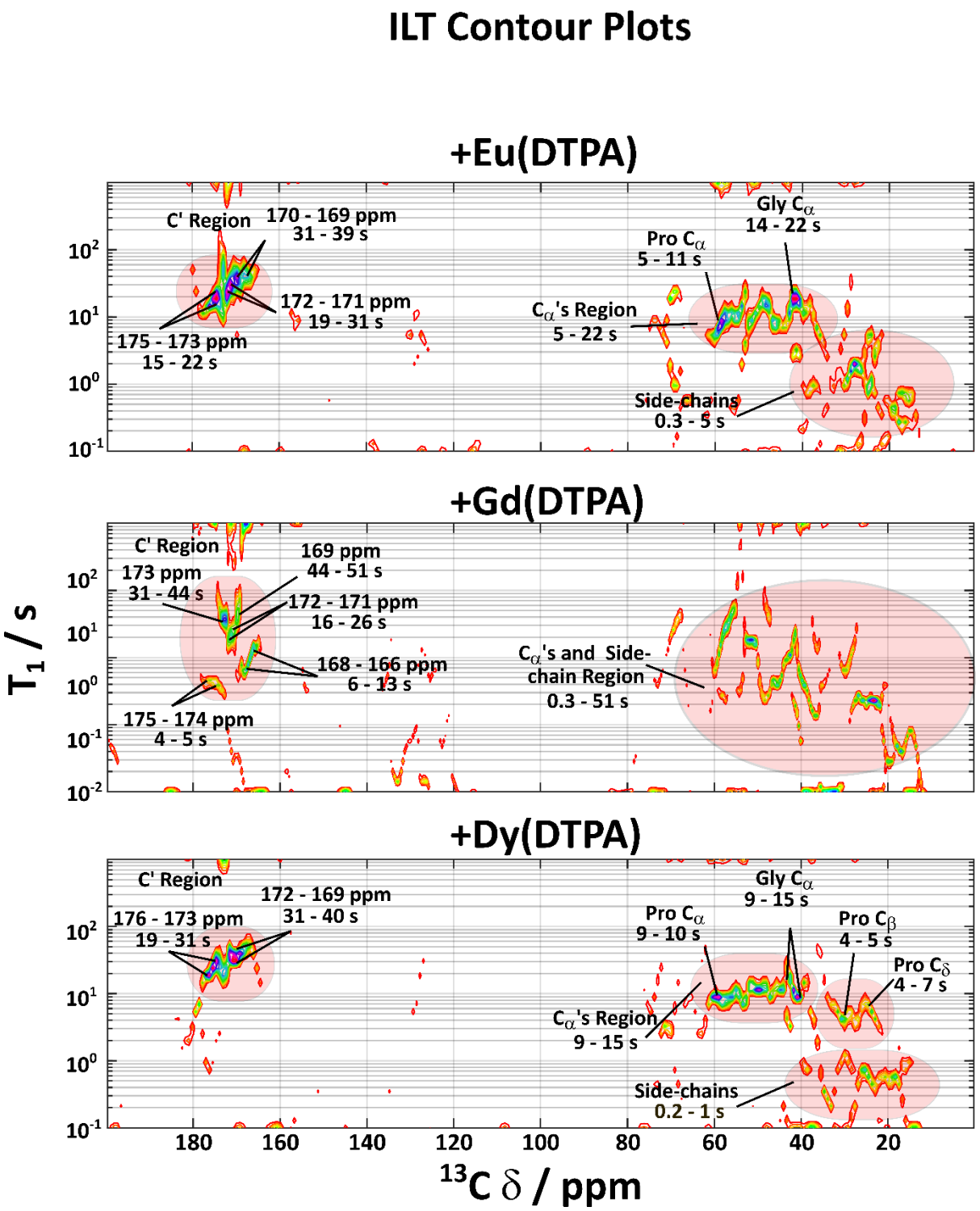

Figure 8.6. (Continued) ILT contour plots of ^{13}C T_1 relaxation results from top to bottom of collagen, collagen incubated with $CuCl_2$, $Pr(DTPA)^{2-}$, $Eu(DTPA)^{2-}$, $Gd(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$. The number of scans per sample was varied to gain best signal-to-noise ratio possible.

8.3. [U-^{13}C,^{15}N]-Lysine Enriched Collagen with Paramagnetic Doping

The [U-^{13}C,^{15}N]-Lysine enriched collagen produced in the ECM was prepared and cleaned up as discussed in Section 5.2.4. The samples were further gently homogenised using a probe sonicator. Samples for the TEM and NMR analysis were prepared in the same way as the reconstituted collagen described in the previous section. The 1D ^{13}C NMR spectra with various paramagnetic complexes is shown in Figure 8.7.

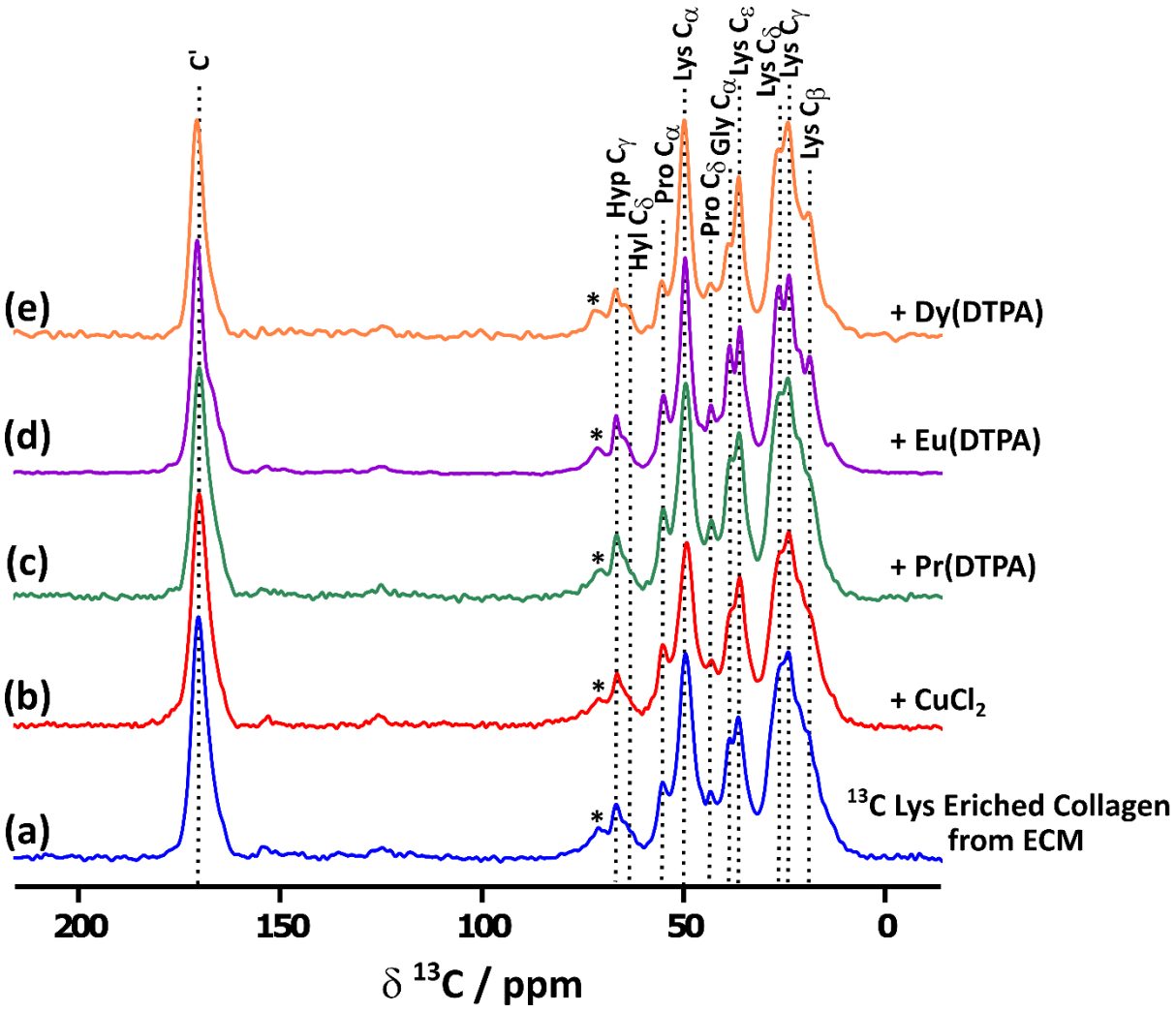

Figure 8.7. 1D ^{13}C CP proton decoupled NMR spectra of [U-^{13}C,^{15}N]-Lysine enriched in vitro grown ECM collagen incubated with (a) D_2O, (b) $CuCl_2$, (c) $Pr(DTPA)^{2-}$, (d) $Eu(DTPA)^{2-}$ and (e) $Dy(DTPA)^{2-}$. The spectrum of the $Gd(DTPA)^{2-}$ sample contained no observable signals, and thus is excluded from this figure. The concentrations of all paramagnetic solutions were 2 mM prepared in D_2O. NMR spectra were recorded using most optimal CP conditions; acquisition delays were the same across all the samples, which was 2 seconds. All samples contained equal mass of sample. The most prominent signals are assigned. Spinning sidebands are highlighted with an asterisk.

No NMR signals were observed for the ECM collagen sample treated with $Gd(DTPA)^{2-}$ with various contact times used, which is likely due to the very strong paramagnetic effect of the Gd(III) ions. Little to no spectral effects are observed for the samples treated with $CuCl_2$ and $Pr(DTPA)^{2-}$. The 1D NMR spectra of the ECM collagen treated with $Eu(DTPA)^{2-}$ shows sharpening of the observed ^{13}C signals and better S/N ratio compared to the other spectra obtained with the same number of scans and on the same mass of material. The sample treated with $Dy(DTPA)^{2-}$ shows a good resolution of the NMR signals as well, despite the severe line broadening observed in the previous examples with the bovine tendon collagen. As in the previous examples with bovine tendon collagen. No significant chemical shift changes were observed, even for the sample incubated with the $Dy(DTPA)^{2-}$ complex, which is expected to be a PCS agent. The PCS tags are known to work better on ^{1}H nuclei; therefore, heteronuclear experiments were carried out where ^{1}H and ^{13}C chemical shifts were correlated to one another (Figure 8.8.). However, no change in the chemical shifts in the proton dimension was observed either.

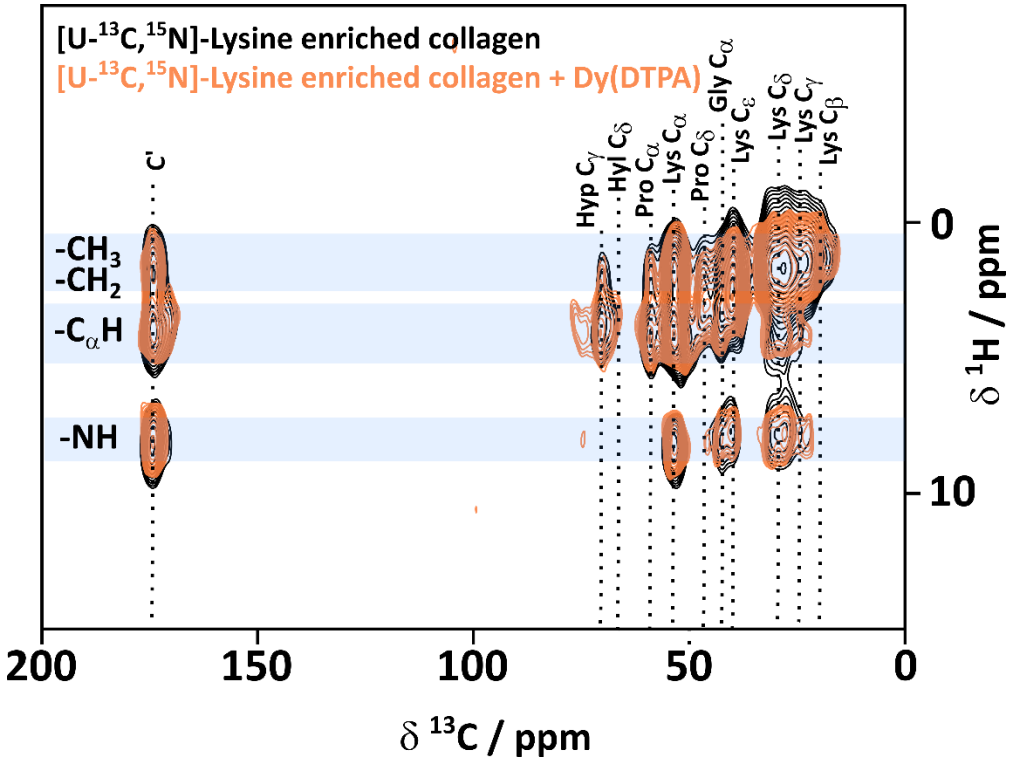

Figure 8.8. 2D ^{1}H-^{13}C heteronuclear correlation NMR spectra of [U-^{13}C,^{15}N]-Lysine enriched collagen extracted from the in vitro cell culture in black compared against the same isotopically enriched sample incubated with 2 mM $Dy(DTPA)^{2-}$ overnight in orange. ^{13}C Lys signals are assigned in the spectra; in light blue main proton environments are shown.

8.3.1. ^{13}C T_1 Relaxation of [U-^{13}C,^{15}N]-Lysine Enriched Collagen with Paramagnetic Doping

The contour plots of the ^{13}C T_1 relaxation data for the Lys and Arg isotopically-enriched ECM collagen samples showed two differently relaxing regions (see Section 7.3.). It was suggested that these regions could correspond to differently relaxing nuclei in the interior and exterior, or gap and overlap zones of the collagen fibrils. We tested a selection of the paramagnetic tags to see if they can be used to differentiate between different regions of the collagen fibrils. In Figure 8.9. contour plots of ^{13}C T_1 relaxation data of U-^{13}C-Lys enriched extracellular collagen incubated with different paramagnetic reagents are presented. All relaxation data were processed using the same parameters for the ILT input to generate comparable contour plots. The TEM images (Figure 8.3.) showed that the paramagnetic species are present in the gap zones of the collagen fibrils and therefore we would expect to see a decrease or loss of the signal intensities in the ILT contour plot for one of two regions if they are indeed a representative of the gap and overlap zones.

The ILT contour plot of the Lys sample incubated with $CuCl_2$ salt shows broadening of the signals along the T_1 dimension compared to that for the reference sample. Bi-modal relaxation behaviour is still observed for the observed nuclei, where both relaxing regions have shorter T_1 values compared to the reference sample. However, both observed relaxing regions in the ILT plots have significant signal intensities. This suggests that the paramagnetic effect felt by the observed nuclei is the same across both zones of the collagen fibrils: whether it is interior and exterior or gap and overlap regions.

The significant decrease in the signal intensity of the fast-relaxing components is almost absent and lost when the collagen fibrils are treated with $Pr(DTPA)^{2-}$ and $Eu(DTPA)^{2-}$ complexes respectively. The T_1 relaxation constants for observed nuclei in both samples are shorter (lower T_1) when compared to the reference sample, and furthermore, the regions with higher T_1 values show similar relaxation constants between the $Pr/Eu(DTPA)^{2-}$ treated samples, suggesting that both of these probes have similar paramagnetic effects on the collagen fibrils. However, one of the paramagnetic complexes is more efficient in removing the fast-relaxing component from the ILT contour plot.

Interesting T_1 results were observed in the sample treated with $Dy(DTPA)^{2-}$ complex, where the two differently relaxing regions are still observed. At first, it appears that the signal intensity of the slow (higher T_1) relaxing region in the ILT plot is lost when compared directly to the

reference sample. However, after examining the T_1 relaxation constants in the ILT plots of the nuclei treated with $Dy(DTPA)^{2-}$ complex, the T_1 values of the slow (higher T_1) relaxing region is almost identical to the reference sample. All of these lanthanide complexes are the same in terms of size and their interactions with the collagen fibrils. Therefore, it is unlikely that the $Dy(DTPA)^{2-}$ complex will interact with the collagen fibrils in a different manner that would result in a change of the observed nuclear relaxation behaviour. The 1D ^{13}C NMR spectra (Figure 8.7.) showed poorer S/N ratio of the carbon signals when they are in the presence of the $Dy(DTPA)^{2-}$ complex, showing that Dy (III) introduces the strongest PRE effect compared to the other paramagnetic complexes. Therefore, the resulting T_1 constants for the observed carbon nuclei should be the shortest (lowest T_1). If we consider these facts, then the fast-relaxing region (low T_1) in the contour plots are corresponding to the nuclei that are close to the paramagnetic centres. The observed weak components with longer T_1 values are likely to be due to natural abundance ^{13}C signals from of Pro, Gly and Hyp signals from their chemical shifts. We speculate that the longer relaxation constants are from hydrophobic regions of the collagen fibrils which do not bind the paramagnetic tags, and so the lack of ^{13}C-^{13}C dipolar interactions the resulting relaxation constants are higher in general.

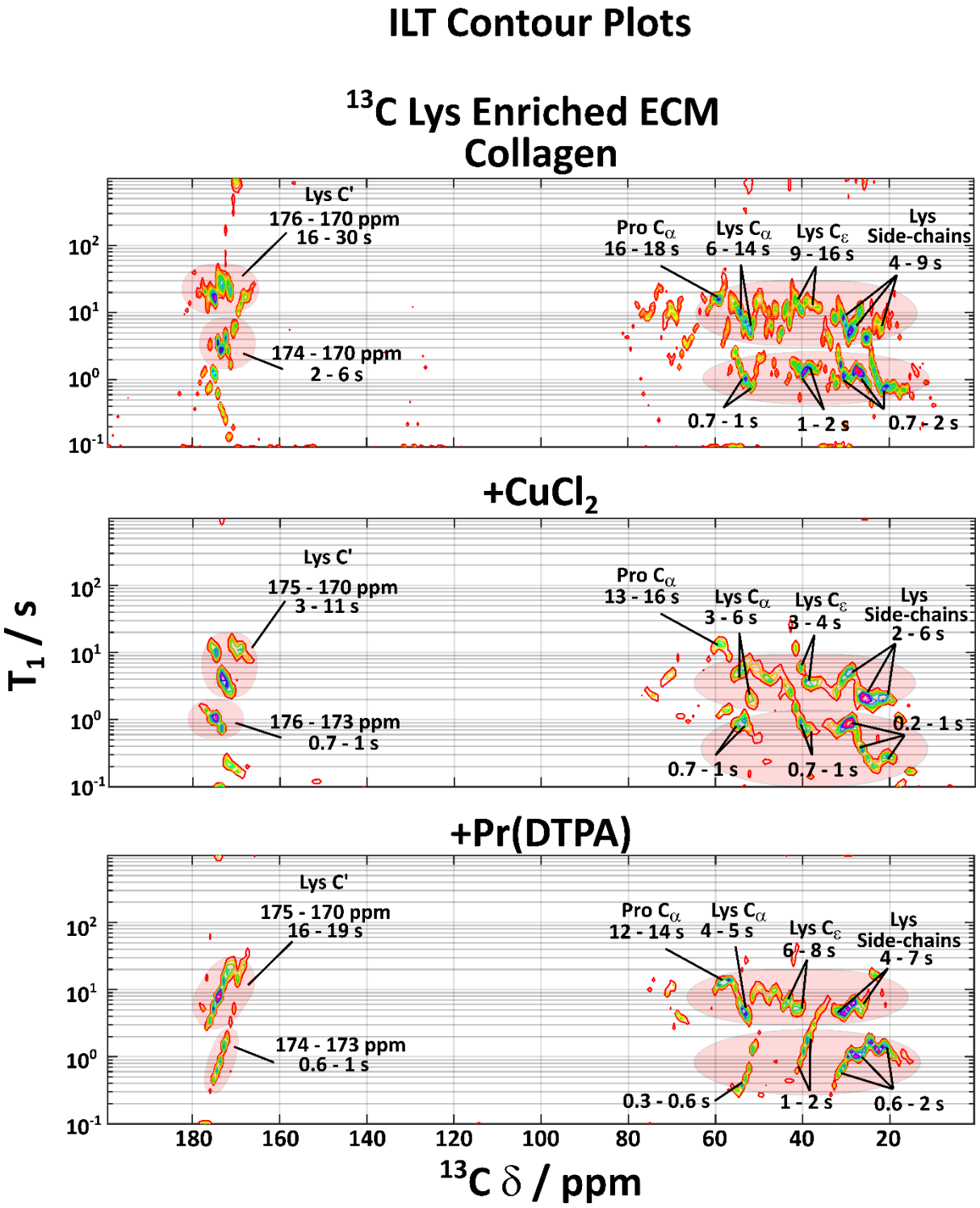

Figure 8.9. ILT contour plots of [U-^{13}C,^{15}N]-Lysine ^{13}C T_1 relaxation results from top to bottom of extracted collagen from in vitro ECM, incubated with $CuCl_2$, $Pr(DTPA)^{2-}$, $Eu(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$. The number of scans per sample was varied to gain the best signal-to-noise ratio possible.

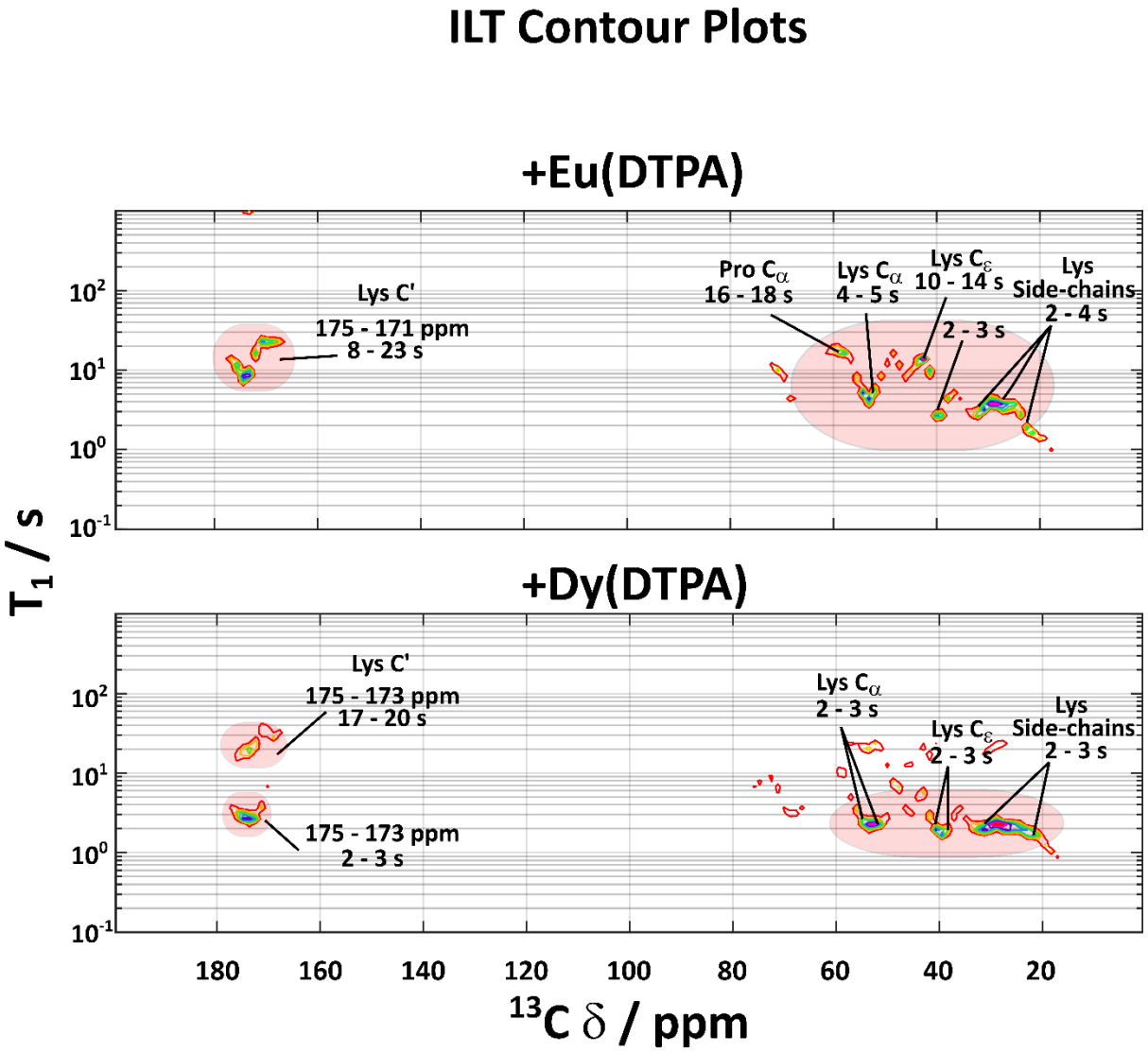

Figure 8.9. (Continued) ILT contour plots of [U-^{13}C,^{15}N]-Lysine ^{13}C T_1 relaxation results from top to bottom of extracted collagen from in vitro ECM, incubated with $CuCl_2$, $Pr(DTPA)^{2-}$, $Eu(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$. The number of scans per sample was varied to gain the best signal-to-noise ratio possible.

8.4. Unexpected Effects of the Paramagnetic Doping of Collagen Fibrils

In this work, two types of different paramagnetic probes were tested – PCSs and PREs. The PCS probes shift signals of nuclei close to the paramagnetic centre.[233–235] Probes tested in this work were: DTPA complexes of Pr (III), Eu (III) and Dy(III). However no significant ^{13}C chemical shift changes were observed in either unlabelled or [U-^{13}C,^{15}N]-Lys isotopically enriched collagen fibrils (See Figures 8.5. and 8.7.). This may be partly due to the chelating DTPA ligand which increase the distance between the amino acid sidechains on the collagen and the paramagnetic centre by ca. 2.5 Å. The atomic radii of the Pr (III), Eu (III) and Dy (III)

are between 0.9 – 1 Å, where the chelating ligand results in the total radii of the probe to be around 3.5 Å. However, in the solution-state, $Dy(DTPA)^{2-}$ can produce detectable shifts for proton nuclei up to a ca. 40 Å distance.[217] However, no observable ^{1}H shifts in ^{13}C-Lys-enriched collagen were observed in the 2D ^{1}H-^{13}C correlation NMR spectra (Figure 8.8.). The lack of the observed shifted ^{1}H NMR signals could be related to the strong dipolar interaction between the nuclei and the paramagnetic ion.

Strong PRE effects were observed for the samples treated with $Gd(DTPA)^{2-}$ and $Dy(DTPA)^{2-}$, where the signal intensity was affected the most for both collagen samples, even more so for the Lys-labelled collagen sample incubated with $Gd(DTPA)^{2-}$ where no ^{13}C NMR signal was observed. The loss of the signal intensities in the ECM collagen samples treated with $Gd(DTPA)^{2-}$ could be related to the smaller size of the collagen fibrils and the more efficient proton spin diffusion across the sample. The Gd (III) has seven unpaired electrons that are isotropically distributed across the HOMO molecular orbitals resulting in efficient enhanced relaxation of surrounding nuclei, where the effect is further transmitted through the sample via proton network. The ECM produced collagen samples could still contain small amounts of the other extracellular proteins or further glycosylations of collagen compared to the reconstituted collagen fibrils that can potentially contribute towards the increased proton network.

Weaker PRE effects on the carbon nuclei were observed for the samples treated with Cu (II) and $Eu/Pr(DTPA)^{2-}$ complexes. Nevertheless, surprising results was observed for the Lys enriched collagen doped with $Eu(DTPA)^{2-}$ (Figure 8.7.), where the S/N ratio was better than the reference sample. This sample was further subjected to 2D NMR – PDSD and POST-C7 (Figure 8.10.). The 2D NMR spectra revealed new chemical shift correlations that were not observed in the reference sample; these are highlighted in Figure 8.10. These new signals correspond to collagen hydroxylysyl post-translational modifications previously detected only with DNP signal enhancement (Figure 8.10. (c) and (d)).[59] However, note that the DNP samples were from different tissue types (skin, and in vitro ECM from vascular smooth muscle cells), where the levels of the post-translational modifications can be lower compared to the foetal sheep osteoblast model used in this work. The $Eu(DTPA)^{2-}$ may prove a useful probe to highlight various amino acid environments and gain more detailed insights into different structural features of the collagen fibrils.

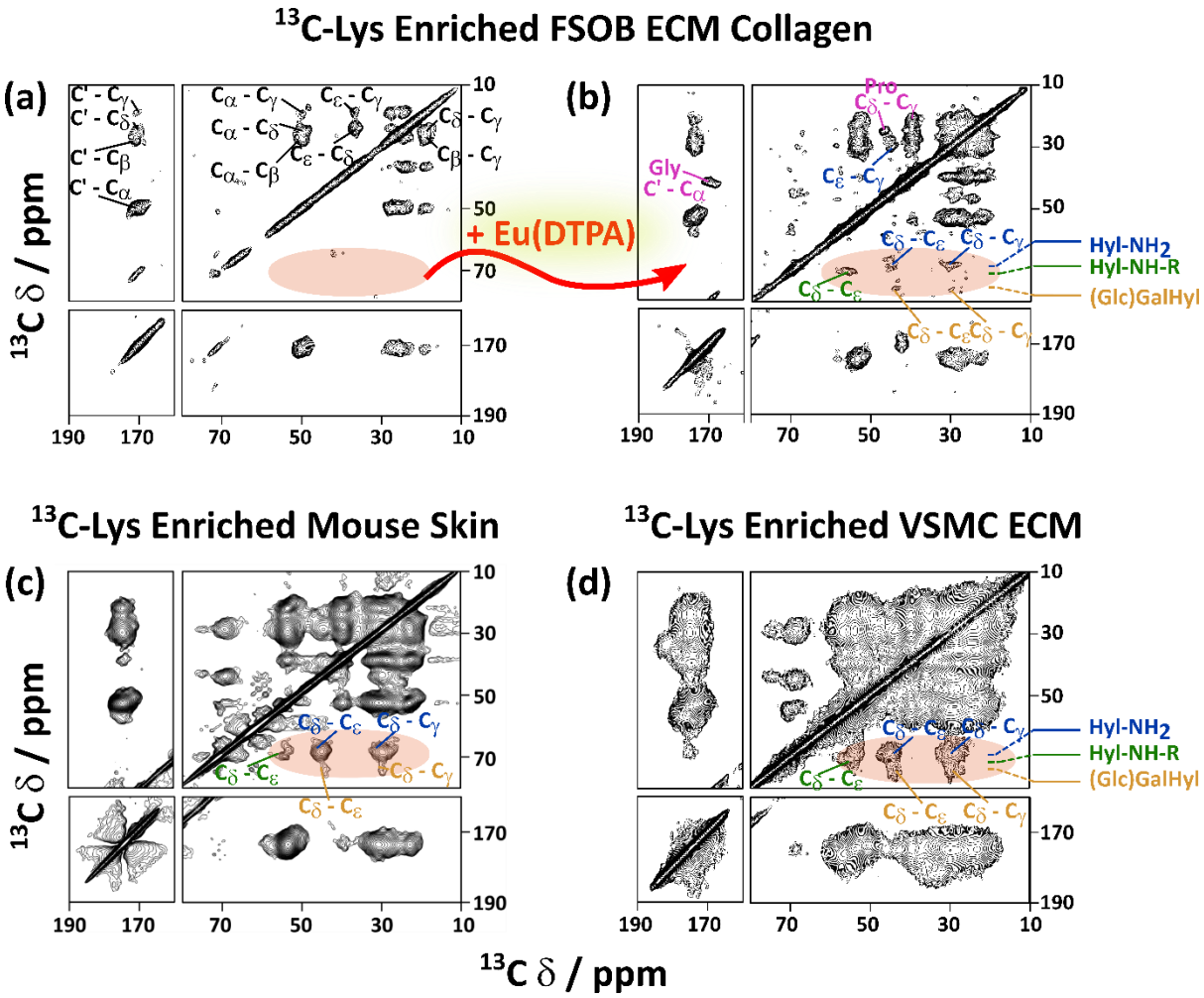

Figure 8.10. 2D ^{13}C-^{13}C PDSD NMR spectra of [U-^{13}C,^{15}N]-Lys enriched in vitro collagen fibrils extracted from FSOB ECM (a) in absence of paramagnetic probes, (b) incubated with 2 mM $Eu(DTPA)^{2-}$ complex in D_2O. 2D ^{13}C-^{13}C DARR DNP NMR spectra of (c) [U-$^{13}C_6$]-Lys enriched mouse skin sample and (d) [U-^{13}C,^{15}N]-Lys enriched bovine vascular smooth muscle cell produced ECM collagen. In (a) cross-peaks of lysine carbons highlighted in black; (b) cross-peaks of glycine and proline carbon atoms highlighted in purple; (b) - (d) cross-peaks of the hydroxylysine (Hyl-NH) highlighted in blue, modified Hyl due to glycation/ cross-links in green, and glycosylated hydroxylysine ((Glc)GalHyl)) in light orange. All NMR parameters used for the (a) - (d) PDSD experiments are summarised in the Materials and Methods section 5.6.4.3. DARR experimental parameters are given in reference [60]. PDSD and DARR experiments were performed with 20 and 5 ms mixing time, respectively. PDSD NMR of ^{13}C-Lys enriched FSOB ECM collagen was recorded at MAS of 9 kHz.

Finally, all paramagnetic probes were tested to see if it is possible to differentiate between the different zones of the collagen fibrils and help to assign the two differently relaxing regions of Lys-enriched ECM collagen. The ^{13}C T_1 relaxation results showed that the paramagnetic effects are influencing both observed relaxing regions. This results in an overall decrease in T_1 constants in both regions due to the spin diffusion. However, the fast-relaxing component in the Lys labelled sample were reduced in intensity by treatment with $Pr(DTPA)^{2-}$, and removed completely by $Eu(DTPA)^{2-}$ and Dy(DTPA) (Figure 8.9.). Considering these relaxation results together with the TEM images of the collagen fibrils, we infer that the fast-relaxing component can be assigned to the gap region of the collagen fibrils. This is an interesting

result, as it shows that the paramagnetic probes can be used to differentiate between the gap and overlap zones in the isotopically enriched collagen samples. However, such separation was not observed in the unlabelled reconstituted collagen treated with different paramagnetic complexes. The lack of the separation could be attributed to the high amounts of the Gly, Pro and Hyp hydrophobic residues present in the collagen that are observed to be dominating the ^{13}C NMR spectrum (Figure 8.5. (a)) and the resulting T_1 relaxation constants can be assigned to most of these residues. Due to the hydrophobic nature of these amino acids, they will be more hidden from the surrounding water, and thus further away from the paramagnetic lanthanide complex. As a result, the interpretation of the differently relaxing regions of these unlabelled samples is much more complex. Nevertheless, it is an interesting result, as it shows that the paramagnetic probes can be used to differentiate between the gap and overlap zones in the isotopically enriched collagen samples.

8.5. Conclusions and Future Perspective

The results showed that the paramagnetic doping of the collagen fibrils could be used to gain more detailed structural information. The Eu(DTPA) complex appeared to enhance signals from low abundance species, like hydroxylysyl post-translational modifications that until now have only been detected in intact collagen in similar samples using DNP NMR spectroscopy. However, at this time we do not fully understand why the $Eu(DTPA)^{2-}$ produces such signal enhancement on the Hyl residues and its modifications. To better understand the effects of $Eu(DTPA)^{2-}$ and other paramagnetic complexes, more systematic titration experiments should be performed. Nevertheless, such paramagnetic probes offer potential for studying collagen in much needed detail at physiologically relevant temperatures rather than the cryogenic temperatures needed for DNP. The results obtained with the Cu (II) complex suggest that no degradation of collagen fibrillar structure is induced from incorporating this potentially residue-binding metal ion, as indicated by the unchanged ^{13}C chemical shifts. These small aquo ion paramagnetic probes appear to penetrate collagen fibrils and access the interior of fibrils as well. This suggests that uncomplexed lanthanide ions could be used in a similar manner. The size of the lanthanide ions are similar to Ca^{2+} ions and could be useful in future to study potentially a calcification of the collagen fibrils. Furthermore, lanthanide ions would be a good probes to study for example collagen interactions with polyanion poly(ADP-ribose), where the pyrophosphate groups of the adenosine nucleotides are observed to bind strongly with Ca^{2+} ions;[236] or polyanionic proteoglycan and glycosaminoglycan interactions with collagen fibrils. TEM imaging could be beneficial in these studies as it would show where the lanthanide species are located on the collagen fibrils as in this work. Other collagen protein

interactions can also be studied, where the collagen ligand binding sites could be used to bind a protein that binds via metal-mediated interactions, such as collagen integrin binding, by changing the diamagnetic metal ion in the binding to paramagnetic ones,, e.g. for integrin $\alpha 2\beta 1$ protein binding to collagen with the Mg^{2+} ion substituted by paramagnetic Mn (II); or the proteins could be modified by attaching a paramagnetic tag. Furthermore, the paramagnetic probes like Mn (II) and Gd (III) could potentially be used as sources of electronic polarization in DNP NMR instead of the commonly used radicals, opening the possibility of studying specific collagen sites with affinity for these ions.

These paramagnetic probes can be useful in studying collagen which has been pathologically modified such as by hyperglycaemic glycation. Cu (II) ions are known to bind to some advanced glycation end products, such as carboxylmethyl lysine (CML), which can chelate copper (II) via its N-donor atom (the ζ-amine) and the nonbonding carboxylate.[237] Preliminary studies (collaboration with Miss Holly Chetwood, Department of Chemistry, University of Cambridge) showed that $CuCl_2$ selectively removed signals of (labelled) glycation products generated by reaction of reconstituted collagen with [U-^{13}C]-ribose (Figure 8.11.).

In conclusion, we have shown that paramagnetic tagging of the collagen can potentially offer new and exciting ways of studying these structurally and biologically vital proteins in more detailed manner, from fibril surfaces to interior and by offering spectral editing to examine very specific binding sites.

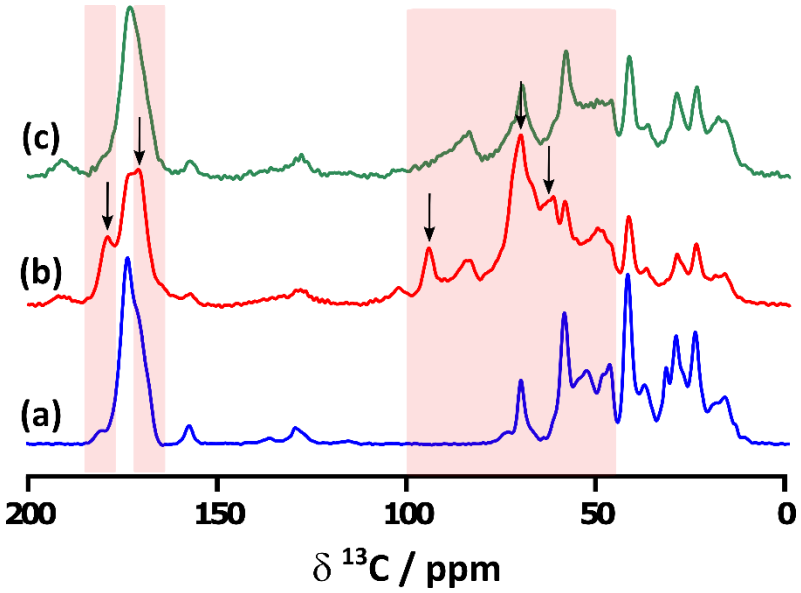

Figure 8.11. 1D ^{13}C CP NMR spectra of reconstituted (a) bovine Achilles tendon collagen, (b) incubated with 50 mM [U-^{13}C]-Ribose for 6 weeks, (c) the sample in (b) incubated with $CuCl_2$ overnight at 4 °C in D_2O and washed twice with D_2O. The areas where glycation products are highlighted with light red rectangles, arrows highlighting regions where most of the signals are lost after treatment with $CuCl_2$. The experiments in producing the samples and NMR spectra were carried out by Miss Holly Chetwood.

9

Conclusions and Future Perspective

In this thesis we demonstrate how solid-state NMR can be used to probe the effects of genetic and chemical modifications on the molecular structure, and dynamics of a range of collagen-based materials. In the first part of the project we used synthetic model peptides to investigate collagen backbone motions via ^{15}N T_1 relaxation. The tailored model peptide sequences represent variations of normal and pathological sequence compositions. ^{15}N T_1 relaxation results showed that small variations in the sequence can have a significant effect on the local conformation and flexibility of the collagen triple helix. Furthermore, pathological mutations of the collagen native sequence showed that the collagen backbone becomes more flexible. This increased local flexibility is very likely to influence the affinity and binding of other extracellular regulatory proteins to the collagen scaffold, as increased mobility comes with greater disordering of the collagen triple helix that can propagate the misalignment of collagen fibrils. These model peptide studies will help us to better understand the influence of sequence on

collagen triple helix flexibility and the role these might play and may enable us to design sequences with specific physical properties.

The chemistry of collagen glycation and subsequent effects on the extracellular matrix, surrounding cells and contributions to multiple diseases have been studied extensively.[15,29,67,204,238,239] However, the underlying chemistry of the resultant products and their effects on the collagen structure are still far from fully understood due to the complex nature and unpredictability of the reaction pathways. Our group has shown that solid-state NMR can be used to identify many glycation products in unprocessed tissue, and to look more closely at resulting molecular structural changes.[16,17,207] The work presented in this thesis focused on assessing molecular dynamic changes of the glycated collagen by using ^{13}C T_1 relaxation measurements. The current understanding that glycation "stiffens" collagen is attributed to the formation of interhelix cross-links; the work in this thesis shows that this is not necessarily the case. We proposed this increase in fibrillar stiffness arises rather from changes in collagen protein backbone and side-chain dynamics, disrupting molecular ordering not only in hydrophilic regions but being transmitted to hydrophobic parts of the collagen fibrils as well. Hydrophilic glycation species attract more hydrating water into the collagen fibrils, while increased compositional and structural heterogeneity will change their mechanical properties. These can have downstream effects on biological properties. Significant Phe structural changes were observed; Phe residues are important components of integrin recognition sites.[209] Such structural changes after glycation help explain the change in ordering of collagen-protein interactions leading to defective cell functions.[204] These results suggest how solid-state NMR structural dynamic studies can elucidate the molecular level contrasts between normal and disease state collagen fibrils.

Besides using only ^{13}C or ^{15}N T_1 relaxation measurements, other types of relaxation experiments like $T_{1\rho}$ or T_2 could be performed to generate insights on the slower molecular motions of the collagen fibrils as well as the peptides. The ^{15}N 1D NMR results (Figure 7.3.) showed that the cross-polarization dynamics differ for the healthy and diabetically affected collagen fibrils showing that the molecular motions at ms time-scales are different for the glycated collagen fibrils. The other type of relaxation experiments could give a better overview of the molecular dynamics at different time-scales that would cover not only fast molecular motions required for biological processes but also slower molecular motions that are relevant for the mechanical properties of the collagen fibrils. Furthermore, ^{2}H isotopic enrichment of the peptides and the collagen fibrils could further complement the structural and dynamic studies.

Besides using solid-state NMR for studying dynamic changes of the collagen fibrils, we investigated whether paramagnetic molecular probes can help towards site-specific information. These studies utilised hydrophilic paramagnetic chelates with preferential affinity (TEM data) towards collagen gap zones, and identified two different Lys (and Arg) relaxation domains. The studies further showed some unexpected sensitivity gains with a $Eu(DTPA)^{2-}$ complex, whereby Hyl and its post-translational modifications were detected, where before this has been done using only DNP NMR. These studies open new ways of studying collagen fibrils in their native state, gaining new site-specific information as well as suggesting new approaches for detection and investigation of less abundant species at biologically relevant (non-cryogenic) temperatures as discussed in the previous Section 8.5.The techniques presented in this thesis and insights gleaned will undoubtedly be of use in future studies directed at elucidating the complex structure and dynamic properties of collagen-based materials.

Appendices

Appendix A

^{13}C INEPT of lyophilised bovine Achilles tendon

In Figure 10.1. the spectra of 1D CP and INEPT ^{13}C NMR spectra of lyophilised bovine Achilles tendon (Sigma) are compared. The INEPT ^{13}C NMR shows only sharp signals corresponding to the unsaturated fatty acid signals at 129.7, 128.2, 30.1, 27.5, 22.9 and 14.1 ppm.

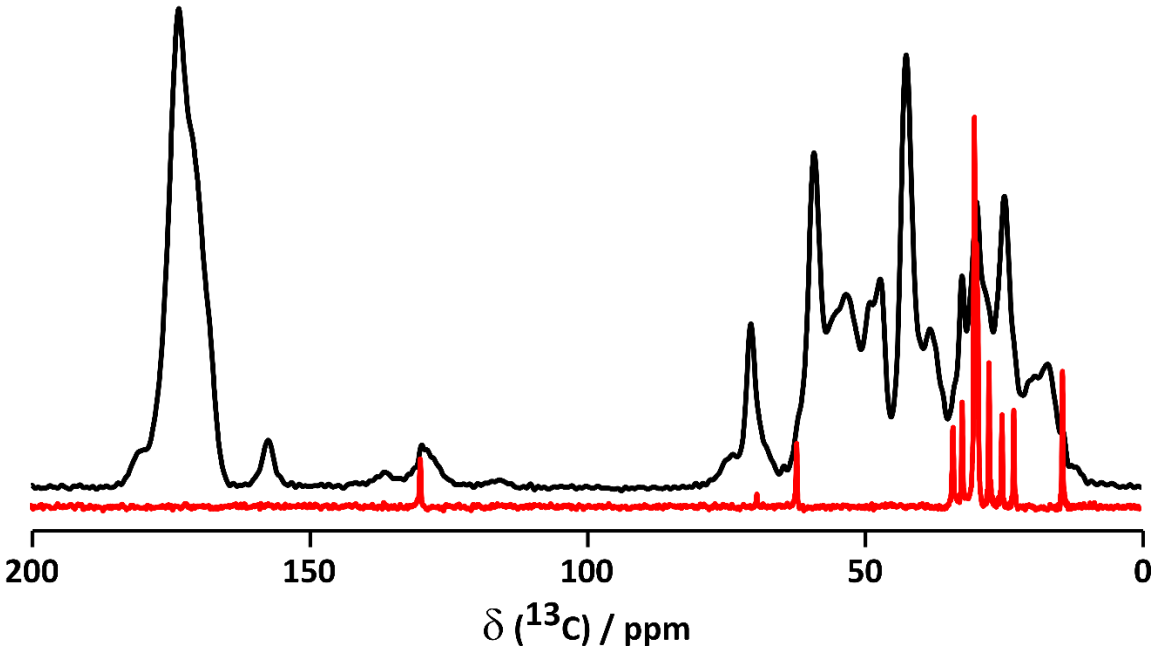

Figure 10.1. ^{13}C 1D CP (black) and INEPT (red) NMR spectra of lyophilised bovine Achilles tendon shown. Sharp signals in the ^{13}C INEPT NMR spectrum corresponds to the unsaturated fatty acid signals.

Appendix B

Imino-Rich Peptides

The figure below illustrates the integrated ^{15}N signal intensities of Gly and Pro for the P-type peptides. The single exponential fit was performed on the relaxation data in order to extract corresponding ^{15}N T_1 constants that are summarised in Table 10.1. These constants were further compared to the ones obtained from the ILT fit, further discussion is presented in Section 6.4.1.

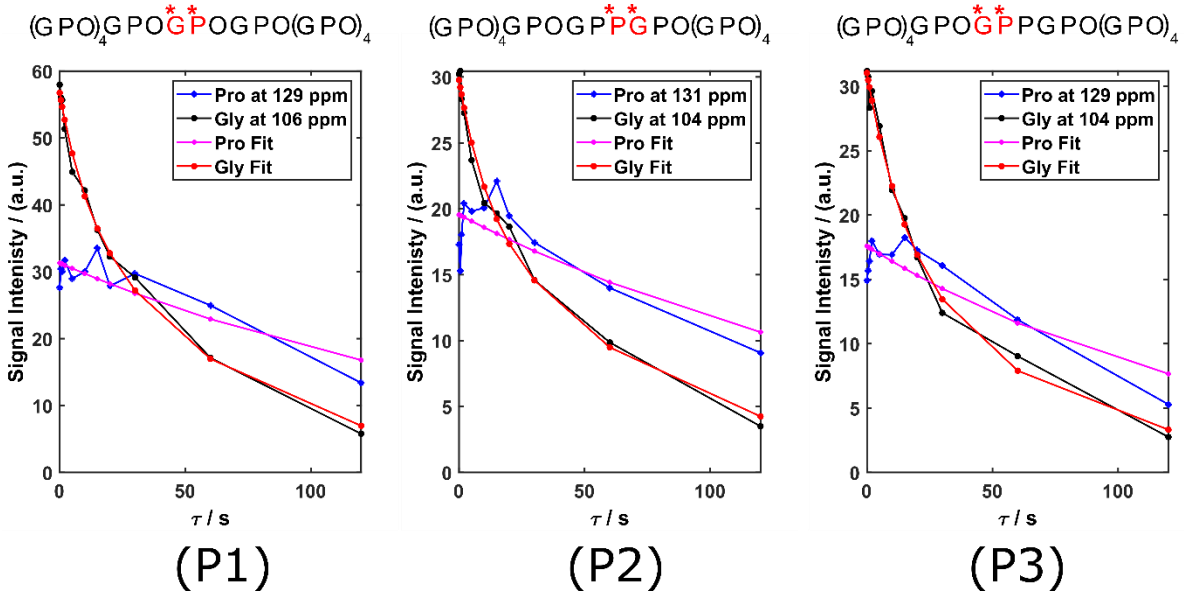

Figure 10.2. Integrated signal intensity of the Pro and Gly signals of peptides (i) P1, (ii) P2 and (iii) P3 and the single exponential fitted by using Equation 4.3 shown.

Table 10.1. The table below shows ^{15}N T_1 relaxation constants calculated for glycine and proline residues in P-type peptides at temperature of 290 K. Constants were calculated using a single exponential fit of Equation (4.2). Fit was performed using a MatLab script.

Peptide and Residue		δ_{iso} / ppm	290 K
P1	**Pro**	129	180.8 ± 74.3
	Gly	106	42.8 ± 8.1
P2	**Pro**	131	146.8 ± 83.8
	Gly	104	44.7 ± 10.6
P3	**Pro**	129	137.3 ± 70.8
	Gly	104	37.7 ± 6.5

Bibliography

1 L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U.S.A.*, 1951, **37**, 272–281.

2 B. P. G. N. Ramachandran and G. Kartha, *Nature*, 1955, **4402**, 593–595.

3 A. Rich and F. H. C. Crick, *Nature*, 1955, **4489**, 915–916.

4 A. Rich and F. H. C. Crick, *J. Mol. Biol.*, 1961, **3**, 483–506.

5 W. Y. Chow, R. Rajan, K. H. Müller, D. G. Reid, J. N. Skepper, W. C. V. Wong, R. A. Brookes, M. Green, D. Bihan, R. W. Farndale, D. A. Slatter, C. M. Shanahan and M. J. Duer, *Science*, 2014, **344**, 742–746.

6 V. W. C. Wong, D. G. Reid, W. Y. Chow, R. Rajan, M. Green, R. A. Brooks and M. J. Duer, *J. Biomol. NMR*, 2015, **63**, 119–123.

7 A. Mohs, T. Silva, T. Yoshida, R. Amin, S. Lukomski, M. Inouye and B. Brodsky, *J. Biol. Chem.*, 2007, **282**, 29757–29765.

8 C. Xu, Z. Yu, M. Inouye, B. Brodsky and O. Mirochnitchenko, *Biomacromolecules*, 2010, **11**, 348–356.

9 Y. Y. Peng, T. Nebl, V. Glattauer and J. A. M. Ramshaw, *Acta Biomater.*, 2018, **80**, 169–175.

10 O. Shoseyov, Y. Posen and F. Grynspan, *Bioengineered*, 2013, **5**, 49–52.

11 G. Chen, N. Kawazoe and T. Tateishi, in *Natural-Based Polymers for Biomedical Applications*, Elsevier, 2008, pp. 396–415.

12 K. M. Pawelec, S. M. Best and R. E. Cameron, *J. Mater. Chem. B*, 2016, **4**, 6484–6496.

13 I. V. Yannas, D. S. Tzeranis, B. A. Harley and P. T. C. So, *Philos. Trans. Royal Soc. A*, 2010, **368**, 2123–2139.

14 M. P. Vitek, K. Bhattacharya, J. M. Glendening, E. Stopat, H. Vlassara, R. Bucala, K. Manogue and A. Cerami, *Proc. Natl. Acad. Sci. U.S.A.*, 1994, **91**, 4766–4770.

15 S. L. Fishman, H. Sonmez, C. Basman, V. Singh and L. Poretsky, *Mol. Med.*, 2018, **24**, 1–12.

16 R. Li, R. Rajan, W. C. V. Wong, D. G. Reid, M. J. Duer, V. J. Somovilla, N. Martinez-Saez, G. J. L. Bernardes, R. Hayward and C. M. Shanahan, *Chem. Commun.*, 2017, **53**, 13275–13380.

17 I. Goldberga, R. Li and M. J. Duer, *Acc. Chem. Res.*, 2018, **51**, 1621–1629.

18 A. J. Heim, W. G. Matthews and T. J. Koob, *Appl. Phys. Lett.*, 2006, **89**, 181902.

19 L. Yang, K. O. van der Werf, C. F. C. Fitié, M. L. Bennink, P. J. Dijkstra and J. Feijen, *Biophys. J.*, 2008, **94**, 2204–2211.

20 A. J. Heim, T. J. Koob and W. G. Matthews, *Biomacromolecules*, 2007, **8**, 3298–3301.

21 M. P. E. Wenger, L. Bozec, M. A. Horton and P. Mesquidaz, *Biophys. J.*, 2007, **93**, 1255–1263.

22 M. Minary-Jolandan and M.-F. Yu, *Biomacromolecules*, 2009, **10**, 2565–2570.

23 A. J. Pell, G. Pintacuda and C. P. Grey, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2019, **111**, 1–271.

24 J. K. Mouw, G. Ou and V. M. Weaver, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 771–785.

25 K. Okuyama, C. Hongo, R. Fukushima, G. Wu, H. Narita, K. Noguchi, Y. Tanaka and N. Nishino, *Biopolymers*, 2004, **76**, 367–377.

26 The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

27 D. J. Prockop and K. I. Kivirikko, *Annu. Rev. Biochem.*, 1995, **64**, 403–434.

28 B. Kern, J. Shen, M. Starbuck and G. Karsenty, *J. Biol. Chem.*, 2001, **276**, 7101–7107.

29 A. J. Bailey, R. G. Paul and L. Knott, *Mech. Ageing Dev.*, 1998, **106**, 1–56.

30 J. Brinckmann, H. Notbohm and P. K. Müller, *Primer in Structure, Processing and Assembly, Collagen*, Springer, 2005.

31 D. E. Hay, *Cell Biology of Extracellular Matrix*, 1991.

32 P. Fratzl, *Collagen: Structure and Mechanics*, Springer, 2008.

33 M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958.

34 H. M. Berman, R. Z. Kramer, J. Bella, P. Mayville and B. Brodsky, *Nat. Struct. Biol.*, 1999, **6**, 454–457.

35 K. Okuyama, X. Xu, M. Iguchi and K. Noguchi, *Biopolymers*, 2006, **84**, 181–191.

36 R. Z. Kramer, M. G. Venugopal, J. Bella, P. Mayville, B. Brodsky and H. M. Berman, *J. Mol. Biol.*, 2000, **301**, 1191–1205.

37 J. Engel, *Annu. Rev. Biophys.*, 1991, **20**, 137–152.

38 J. C. Marini, A. Forlino, W. A. Cabral, A. M. Barnes, J. D. San Antonio, S. Milgrom, C. J. Hyland, J. Körkkö, D. J. Prockop, A. De Paepe, P. Coucke, S. Symoens, F. H. Glorieux, P. J. Roughley, A. M. Lund, K. Kuurila-svahn, H. Hartikka, D. H. Cohn, D. Krakow, M. Mottes, U. Schwarze, D. Chen, K. Yang, C. Kuslich, J. Troendle, R. Dalgleish and P. H. Byers, *Hum. Mutat.*, 2007, **28**, 209–221.

39 H. Kuivaniemi, G. Tromp and D. J. Prockop, *FASEB J.*, 1991, **5**, 2052–2060.

40 R. Bogaert, G. E. Mary, A. Weis, H. E. Gruber, D. L. Rimoin, D. H. Cohn and D. R. Eyre, *J. Biol. Chem.*, 1992, **267**, 22522–22526.

41 D. J. Prockop, *J. Biol. Chem.*, 1990, **265**, 15349–15352.

42 G. W. Cole, *Prog. Nucleic Acid Res. Mol. Biol.*, 1994, **47**, 29–80.

43 J. P. R. O. Orgel, T. C. Irving, A. Miller and T. J. Wess, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 9001–9005.

44 D. E. Birk and R. L. Trelstad, *J. Cell Biol.*, 1986, **103**, 231–240.

45 J. A. Chapman, *Biopolymers*, 1989, **28**, 1367–1382.

46 J. E. Scott, *Biochem. J.*, 1988, **252**, 313–323.

47 A. Veis, J. Anesey, L. Yuan and S. J. Levy, *Proc. Natl. Acad. Sci. U.S.A.*, 1973, **70**, 1464–1467.

48 B. L. Trus and K. A. Piex, *Nature*, 1980, **286**, 5770–5771.

49 J. Zhu, C. L. Hoop, D. A. Case and J. Baum, *Sci. Rep.*, 2018, **8**, 16646.

50 J. Myllyharju, in *Collagen*, 2005, pp. 115–147.

51 S. M. Krane, *Amino Acids*, 2008, **35**, 703–710.

52 T. V Burjanadze, *Biopolymers*, 1979, **18**, 931–938.

53 L. E. Bretscher, M. Taylor and R. T. Raines, *Chem. Biol.*, 1999, **6**, 63–70.

54 C. L. Jenkins and R. T. Raines, *Nat. Prod. Rep.*, 2002, **19**, 49–59.

55 A. V Persikov, J. A. M. Ramshaw and B. Brodsky, *J. Biol. Chem.*, 2005, **280**, 19343–19349.

56 R. G. Spiro, *Glycobiology*, 2002, **12**, 43–56.

57 B. Batge, C. Winter, H. Notbohm, Y. Acil, J. Brinckmann and P. K. Müller, *J. Biochem.*, 1997, **122**, 109–115.

58 E. Song, A. Mayampurath, C. Y. Yu, H. Tang and Y. Mechref, *J. Proteome Res.*, 2014, **13**, 5570–5580.

59 W. Y. Chow, R. Li, I. Goldberga, D. G. Reid, R. Rajan, J. Clark, H. Oschkinat, M. J. Duer, R. Hayward and C. M. Shanahan, *Chem. Commun.*, 2018, **54**, 12570–12573.

60 D. A. Hanson and D. R. Eyre, *J. Biol. Chem.*, 1996, **271**, 26508–26516.

61 R. C. Siegel, *Int. Rev. Connect. Tissue Res.*, 1979, **8**, 73–118.

62 D. R. Eyre, M. A. Weis and J. J. Wu, *Methods*, 2008, **45**, 65–74.

63 N. Light and A. J. Bailey, *Fed. Eur. Biochem. Soc.*, 1985, **182**, 503–508.

64 A. Bierhaus, M. A. Hofmann, R. Ziegler and P. P. Nawroth, *Cardiovasc. Res.*, 1998, **37**, 586–600.

65 C. Han, Y. Lu, Y. Wei, Y. Liu and R. He, *PLoS One*, 2011, **6**, 1–10.

66 R. Sandwick, M. Johanson and E. Breuer, *Ann. N. Y. Acad. Sci.*, 2005, **1043**, 85–96.

67 D. Vashishth, G. Gibson, J. Khoury, M. Schaffler, J. Kimura and D. Fyhrie, *Bone*, 2001, **28**, 195–201.

68 C. Bonnans, J. Chou and Z. Werb, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 786–801.

69 B. N. Mason, A. Starchenko, R. M. Williams, L. J. Bonassar and C. A. Reinhart-King, *Acta Biomater.*, 2013, **9**, 4635–4644.

70 N. Ogawa, T. Yamaguchi, S. Yano, M. Yamauchi, M. Yamamoto and T. Sugimoto, *Horm. Metab. Res.*, 2007, **39**, 871–875.

71 J. P. R. O. Orgel, J. D. San Antonio and O. Antipova, *Connect. Tissue Res.*, 2011, **52**, 2–17.

72 S. W. Chang, S. J. Shefelbine and M. J. Buehler, *Biophys. J.*, 2012, **102**, 640–648.

73 T. Ushiki and C. Ide, *Cell Tissue Res.*, 1990, **260**, 175–184.

74 M. Raspanti, M. Protasoni, A. Manelli, S. Guizzardi, V. Mantovani and A. Sala, *Micron*, 2006, **37**, 81–86.

75 M. Tzaphlidou, C. Cummings and J. A. Chapman, *Micron Microsc. Acta*, 1992, **23**, 25–35.

76 K. E. Kadler, D. F. Holmes, J. A. Trotter and J. A. Chapman, *Biochem. J.*, 1996, **11**, 1–11.

77 T. Starborg, N. S. Kalson, Y. Lu, A. Mironov, T. F. Cootes, D. F. Holmes and K. E. Kadler, *Nat. Protoc.*, 2013, **8**, 1433–1448.

78 J. A. Chapman, M. Tzaphlidou, K. M. Meek and K. E. Kadler, *Electron Microsc. Rev.*, 1990, **3**, 143–182.

79 F. Variola, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2950–2959.

80 D. R. Baselt, J. P. Revel and J. D. Baldeschwieler, *Biophys. J.*, 1993, **65**, 2644–2655.

81 M. Janko, A. Zink, A. M. Gigler, W. M. Heckl and R. W. Stark, *Proc. R. Soc. B*, 2010, **277**, 2301–2309.

82 P. Dutov, O. Antipova, S. Varma, J. P. R. O. Orgel and J. D. Schieber, *PLoS One*, 2016, **11**, 1–13.

83 T. Gutsmann, G. E. Fantner, J. H. Kindt, M. Venturoni, S. Danielsen and P. K. Hansma, *Biophys. J.*, 2004, **86**, 3186–3193.

84 D. A. Cisneros, C. Hung, C. M. Franz and D. J. Müller, *J. Struct. Biol.*, 2006, **154**, 232–245.

85 D. F. Holmes, C. J. Gilpin, C. Baldock, U. Ziese, A. J. Koster and K. E. Kadler, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 7307–7312.

86 J. Bella, *Biochem. J.*, 2016, **473**, 1001–1025.

87 G. C. Knight, L. F. Morton, D. S. Tuckwell, R. W. Farndale and M. J. Barnes, *J. Biol. Chem.*, 1998, **275**, 35–40.

88 Y. Xu, S. Gurusiddappa, R. L. Rich, R. T. Owens, D. R. Keene, R. Mayne, A. Höök and M. Höök, *J. Biol. Chem.*, 2000, **275**, 38981–38989.

89 E. Hohenester, T. Sasaki, C. Giudici, R. W. Farndale and H. P. Bachinger, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 18273–18277.

90 N. Pugh, A. M. C. Simpson, P. A. Smethurst, P. G. de Groot, N. Raynal, R. W. Farndale, P. G. De Groot, N. Raynal and R. W. Farndale, *Blood*, 2010, **115**, 5069–5080.

91 R. W. Farndale, *Essays Biochem.*, 2019, **63**, 337–348.

92 W. Y. Chow, C. J. Forman, D. Bihan, A. M. Puszkarska, R. Rajan, D. G. Reid, D. A. Slatter, L. J. Colwell, D. J. Wales, R. W. Farndale and M. J. Duer, *Sci. Rep.*, 2018, **8**, 13809.

93 W. Ying Chow, D. Bihan, C. J. Forman, D. A. Slatter, D. G. Reid, D. J. Wales, R. W. Farndale and M. J. Duer, *Sci. Rep.*, 2015, **5**, 12556.

94 J. A. Hodges and R. T. Raines, *J. Am. Chem. Soc.*, 2003, **125**, 9262–9263.

95 M. D. Shoulders, J. A. Hodges and R. T. Raines, *J. Am. Chem. Soc.*, 2008, **128**, 8112–8113.

96 K. Okuyama, K. Miyama, K. Mizuno and H. P. Ba, *Biopolymers*, 2012, **97**, 607–616.

97 E. Kilchherr, H. Hofmann, W. Steigemann and J. Engel, *J. Mol. Biol.*, 1985, **186**, 403–415.

98 C. G. Long, E. Braswell, D. Zhu, J. Apigo, J. Baum and B. Brodsky, *Biochemistry*, 1993, **32**, 11688–11695.

99 K. Beck, V. C. Chan, N. Shenoy, A. Kirkpatrick, J. A. M. Ramshaw and B. Brodsky, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 4273–4278.

100 Y. Li, B. Brodsky and J. Baum, *J. Biol. Chem.*, 2009, **284**, 20660–20667.

101 J. Bella, M. Eaton, B. Brodsky and H. M. Berman, *Science*, 1994, **266**, 75–81.

102 N. Yamaguchis, P. D. Benyag, M. van der Rest and Y. Ninomiyas, *J. Biochem.*, 1989, **264**, 16022–16029.

103 D. Brazel, I. Oberbaumer, H. Dieringer, W. Babel, R. W. Glanville, R. Deutzmann and K. Kuhn, *Eur. J. Biochem*, 1987, **168**, 529–536.

104 U. Schwarz, D. Schuppan, I. Oberbaumer, R. W. Glanville, R. Deutzmann, R. Timpl and K. Kuhn, *Eur. J. Biochem.*, 1986, **157**, 49–56.

105 P. D. Yurchenco and G. C. Ruben, *Am. J. Pathol.*, 1988, **132**, 278–291.

106 B. G. Hudson, K. Tryggvason, M. Sundaramoorthy and E. G. Neilson, *N. Engl. J. Med.*, 2003, **348**, 2543–2556.

107 J. Zhous, J. M. Hertz, A. Leinonens and K. Tryggvasons, *J. Biol. Chem.*, 1992, **267**, 12475–12481.

108 J. M. Pace, M. Atkinson, M. C. Willing, G. Wallis and P. H. Byers, *Hum. Mutat.*, 2001, **18**, 319–326.

109 J. Bella, J. Liu, R. Kramer, B. Brodsky and H. M. Berman, *J. Mol. Biol.*, 2006, **362**, 298–311.

110 J. Bella, *J. Struct. Biol.*, 2014, **186**, 438–450.

111 C. G. Knight, L. F. Morton, D. J. Onley, A. R. Peachey, A. J. Messent, P. A. Smethurst, D. S. Tuckwell, R. W. Farndale and M. J. Barnes, *J. Biol. Chem.*, 1998, **50**, 33287–33294.

112 J. Emsley, C. G. Knight, R. W. Farndale and M. J. Barnes, *J. Mol. Biol.*, 2004, **335**, 1019–1028.

113 J. Emsley, C. G. Knight, R. W. Farndale, M. J. Barnes and R. C. Liddington, *Cell*, 2000, **101**, 47–56.

114 J. K. Rainey and M. C. Goh, *Protein Sci.*, 2009, **11**, 2748–2754.

115 D. L. Bodian, R. J. Radmer, S. Holbert and T. E. Klein, *Pac. Symp. Biocomput.*, 2011, 193–204.

116 A. Gautieri, S. Vesentini, A. Redaelli and M. J. Buehler, *Nano Lett.*, 2011, **11**, 757–766.

117 C. M. Stultz, *Protein Sci.*, 2006, **15**, 2166–2177.

118 R. Salsas-Escat, P. S. Nerenberg and C. M. Stultz, *Biochemistry*, 2010, **49**, 4147–4158.

119 S. W. Chang and M. J. Buehler, *Mater. Today*, 2014, **17**, 70–76.

120 D. A. Torchia and K. A. Piez, *J. Mol. Biol.*, 1973, **76**, 419–424.

121 D. A. Torchia, J. R. Lyerla and A. J. Quattrone, *Biochemistry*, 1975, **14**, 887–900.

122 D. A. Torchia and D. L. Vanderhart, *J. Mol. Biol.*, 1976, **104**, 315–321.

123 S. K. Sarkar, C. E. Sullivan and D. A. Torchia, *J. Biol. Chem.*, 1983, **258**, 9762–9767.

124 L. W. Jelinski and D. A. Torchia, *J. Mol. Biol.*, 1980, **138**, 255–272.

125 D. A. Torchia, Y. Hiyama, S. K. Sarkar, C. E. Sullivan and P. E. Young, *Biopolymers*, 1985, **24**, 65–75.

126 S. K. Sarkar, *Proc. Natl. Acad. Sci. U.S.A.*, 1984, **81**, 4800–4803.

127 L. W. Jelinski, C. E. Sullivan and D. A. Torchia, *Nature*, 1980, **284**, 531–534.

128 L. W. Jelinski, C. E. Sullivan, L. S. Batchelder and D. A. Torchia, *Biophys. J.*, 1980, **32**, 515–529.

129 S. K. Sarkar, P. E. Young and D. A. Torchia, *J. Am. Chem. Soc.*, 1986, **108**, 6459–6464.

130 J. Schaefer, E. O. Stejskal, C. F. Brewer, H. D. Keiser and H. Sternlicht, *Arch. Biochem. Biophys.*, 1978, **190**, 657–661.

131 D. Huster, *Annu. Reports NMR Spectrosc.*, 2008, **64**, 127–159.

132 W. Y. Chow, in *Collagen: Methods and Protocols*, 2019, pp. 57–77.

133 H. Saitô and M. Yokoi, *J. Biochem.*, 1992, **111**, 376–382.

134 A. E. Aliev, *Biopolymers*, 2005, **77**, 230–245.

135 D. Reichert, O. Pascui, R. deAzevedo, Eduardo, T. J. Bonagamba, K. Arnold and D. Huster, *Magn. Reson. Chem.*, 2004, **42**, 276–284.

136 F. Gobeaux, G. Mosser, A. Anglo, P. Panine, P. Davidson, M. M. Giraud-Guille and E. Belamie, *J. Mol. Biol.*, 2008, **376**, 1509–1522.

137 P. De Sa Peixoto, G. Laurent, T. Azaïs and G. Mosser, *J. Biol. Chem.*, 2013, **288**, 7528–7535.

138 C. Singh and N. Sinha, *J. Phys. Chem. C*, 2016, **120**, 9393–9398.

139 C. Singh, R. K. Rai, F. Aussenac and N. Sinha, *J. Phys. Chem. Lett.*, 2014, **5**, 4044–4048.

140 K. Kar, S. Ibrar, V. Nanda, T. M. Getz, S. P. Kunapuli and B. Brodsky, *Biochemistry*, 2009, **48**, 7959–7968.

141 M. Levitt, *Spin Dynamics*, John Wiley & Sons, Ltd, Chichester, 2008.

142 M. J. Duer, *Introduction to Solid-State NMR Spectroscopy*, Wiley-Blackwell, 2005.

143 R. K. Harris, R. E. Wasylishen and M. J. Duer, *NMR Crystallography*, Wiley, 2009.

144 P. J. Hore, *Nuclear Magnetic Resonance*, Oxford University Press, 1995.

145 D. C. Apperley, R. K. Harris and P. Hodgkinson, *Solid-State NMR: Basic Principles and Practice*, Momentum Press, 2012.

146 J. Keeler, *Understanding NMR Spectroscopy*, Wiley, 2005.

147 M. Bak, J. T. Rasmussen and N. C. Nielsen, *J. Magn. Reson.*, 2000, **147**, 296–330.

148 J.-P. Amoureux, C. Fernandez and L. Frydman, *Chem. Phys. Lett.*, 1996, **259**, 347–355.

149 J. K. M. Sanders and B. K. Hunter, *Modern NMR Spectroscopy: A Guide for Chemists*, Oxford University Press, 1993.

150 L. R. Sarles and R. M. Cotts, *Phys. Rev. J.*, 1958, **111**, 853–859.

151 D. Vanderhart, W. L. Earl and A. Garroway, *J. Magn. Reson.*, 1981, **44**, 361–401.

152 A. E. Bennett, C. M. Rienstra, R. G. Griffin, K. V. Lakshimi, M. Auger, K. V. Lakshmi and R. G. Griffin, *J. Chem. Phys.*, 1995, **103**, 6951–6958.

153 B. M. Fung, A. K. Khitrin and K. Ermolaev, *J. Magn. Reson.*, 2000, **101**, 97–101.

154 A. Bielecki, A. C. Kolbert and M. H. Levitt, *Chem. Phys. Lett.*, 1989, **155**, 341–346.

155 S. R. Hartmann and E. L. Hahn, *Phys. Rev.*, 1962, **128**, 2042–2053.

156 N. M. Szeverenyi, M. J. Sullivan and G. E. Maciel, *J. Magn. Reson.*, 1982, **47**, 462–475.

157 M. Hohwy, H. J. Jakobsen, M. Edén, M. H. Levitt and N. C. Nielsen, *J. Chem. Phys.*, 1998, **108**, 2686–2694.

158 Y. S. Liao, M. Lee, T. Wang, V. I. Sergeyev and M. Hong, *J. Biomol. NMR*, 2017, **176**, 139–148.

159 C. Sauvée, M. Rosay, G. Casano, F. Aussenac, R. T. Weber, O. Ouari and P. Tordo, *Angew. Chemie Int. Ed.*, 2013, **52**, 10858–10861.

160 R. K. Wangsnesst and F. Bloch, *Phys. Rev.*, 1953, **89**, 728–739.

161 F. Bloch, *Phys. Rev.*, 1956, **102**, 104–135.

162 A. G. Redfield, *IBM J. Res. Dev.*, 1957, **1**, 3937–3940.

163 D. M. Dias and A. Ciulli, *Prog. Biophys. Mol. Biol.*, 2014, **116**, 101–112.

164 D. A. Torchia, *J. Magn. Reson.*, 1978, **30**, 613–616.

165 B. J. Wylie, W. Trent Franks and C. M. Rienstra, *J. Phys. Chem. B*, 2006, **110**, 10926–10936.

166 Y. Wei, D. K. Le and A. Ramamoorthy, *J. Am. Chem. Soc.*, 2001, **123**, 6118–6126.

167 I. G. Marini, rilt.m code, https://uk.mathworks.com/matlabcentral/fileexchange/6523-rilt, (accessed 25 July 2019).

168 S. W. Provencher, *Comput. Phys. Commun.*, 1982, **27**, 229–242.

169 S. W. Provencher, CONTIN, http://s-provencher.com/contin.shtml, (accessed 25 July 2019).

170 A. Bogaychuk, N. Sinyavsky and G. Kupriyanova, *Appl. Magn. Reson.*, 2016, **47**, 1409–1417.

171 A. Lupulescu, M. Kotecha and L. Frydman, *J. Am. Chem. Soc.*, 2003, **125**, 3376–3383.

172 T. A. Early, *J. Magn. Reson.*, 1987, **74**, 337–343.

173 J. M. Gibbins and M. P. Mahaut-Smith, *Methods in Molecular Biology: Platlets and Megakaryocytes*, Mumana Press Inc., Totowa, New Jersey, 2004, vol. 2.

174 P. D. Chan, W.C.; White, *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*, Oxford University Press, 2000.

175 J. D. Kretlow, Y. Q. Jin, W. Liu, W. J. Zhang, T. H. Hong, G. Zhou, L. S. Baggett, A. G. Mikos and Y. Cao, *BMC Cell Biol.*, 2008, **9**, 1–13.

176 M. M. Burrell, *Enzymes of Molecular Biology*, 1993.

177 R. F. Doolittle, *Methods Enzymol.*, 1970, **19**, 555–569.

178 D. Jain, A. Y. Stark, P. H. Niewiarowski, T. Miyoshi and A. Dhinojwala, *Sci. Rep.*, 2015, **5**, 1–8.

179 G. Platzer, M. Okon and L. P. McIntosh, *J. Biomol. NMR*, 2014, **60**, 109–129.

180 J. Einbinder and M. Schubert, *J. Biol. Chem.*, 1951, **188**, 335–341.

181 J. M. Menter, G. D. Williamson, K. Carlyle, C. L. Moore and I. Willis, *Photochem. Photobiol.*, 1995, **62**, 402–408.

182 Y. Katio, T. Nishikawa and S. Kawakisji, *Photochem. Photobiol.*, 1995, **61**, 367–372.

183 S. S. Lehrer and G. D. Fasman, *Biochemistry*, 1967, **6**, 757–767.

184 A. J. Gross and I. W. Sizer, *J. Biol. Chem.*, 1959, **234**, 1611–1614.

185 S. Sakura and D. Fujimoto, *Photochem. Photobiol.*, 1984, **40**, 731–734.

186 V. M. Monnier, R. R. Kohnt and A. Cerami, *Proc. Nati. Acad. Sci. U.S.A.*, 1984, **81**, 583–587.

187 Y. Shen, D. Zhu, W. Lu, B. Liu, Y. Li and S. Cao, *Appl. Sci.*, 2018, **8**, 1–10.

188 J. L. Zitnay, Y. Li, Z. Qin, B. H. San, B. Depalle, S. P. Reese, M. J. Buehler, S. M. Yu and J. A. Weiss, *Nat. Commun.*, 2017, **8**, 14913.

189 K. Okuyama, K. Okuyama, S. Arnott, M. Takayanagi and M. Kakudo, *J. Mol. Biol.*, 1981, **152**, 427–443.

190 M. Bhate, X. Wang, J. Baum and B. Brodsky, *Biochemistry*, 2002, **41**, 6539–6547.

191 A. Buevich, J. Baum, S. Lindquist, J. W. Kelly, P. H. Byers, R. Seckler and A. Helenius, *Philos. Trans. R. Soc. B Biol. Sci.*, 2001, **356**, 159–168.

192 J. Baum and B. Brodsky, *Fold. Des.*, 1997, **2**, 53–60.

193 A. V Buevich, Q. Dai, X. Liu, B. Brodsky and J. Baum, *Biochemistry*, 2000, **39**, 4299–4308.

194 J. A. M. Ramshaw, *J. Struct. Biol.*, 1998, **91**, 86–91.

195 R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari and H. M. Berman, *J. Mol. Biol.*, 1998, **280**, 623–638.

196 E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao and J. L. Markley, *Nucleic Acids Res.*, 2008, **36**, 402–408.

197 N. Giraud, M. Blackledge, A. Böckmann, L. Emsley, A. Bo and L. Emsley, *J. Magn. Reson.*, 2007, **184**, 51–61.

198 V. Chevelkov, A. Diehl and B. Reif, *J. Chem. Phys.*, 2008, **128**, 1–5.

199 A. Krushelnitsky, T. Bra, T. Bräuniger and D. Reichert, *J. Magn. Reson.*, 2006, **182**, 339–342.

200 G. Gemmecker, W. Jahnke and H. Kessler, *J. Am. Chem. Soc*, 1993, **115**, 447–465.

201 P. Fan, M. Li, B. Brodsky and J. B. J, *Biochemistry*, 1993, **23**, 13299–13309.

202 M. A. Bag and L. M. Valenzuela, *Int. J. Mol. Sci.*, 2017, **18**, 1–31.

203 R. Singh, A. Barden, T. Mori and L. Beilin, *Diabetologia*, 2001, **44**, 129–146.

204 K. L. Reigle, G. Di Lullo, K. R. Turner, J. A. Last, I. Chervoneva, D. E. Birk, J. L. Funderburgh, E. Elrod, M. W. Germann, R. D. Sanderson and J. D. S. Antonio, *J. Cell Biochem.*, 2009, **104**, 1684–1698.

205 M. Camici, M. Grazia Tozzi and P. Luigi Ipata, *J. Biochem. Biophys. Methods*, 2006, **68**, 145–154.

206 M. G. Tozzi, M. Camici, L. Mascia, F. Sgarrella and P. L. Ipata, *FEBS J.*, 2006, **273**, 1089–1101.

207 P. T. B. Bullock, D. G. Reid, W. Ying Chow, W. P. W. Lau and M. J. Duer, *Biosci. Rep.*, 2014, **34**, 83–92.

208 R. D. B. Fraser, T. P. MacRae and A. Miller, *J. Mol. Biol.*, 1987, **193**, 115–125.

209 K. Kühn and J. Eble, *Trends Cell Biol.*, 1994, **4**, 256–261.

210 M. Yamauchi and M. Sricholpech, *Essays Biochem.*, 2012, **52**, 113–133.

211 K. Uzawa, H. N. Yeowell, K. Yamamoto, Y. Mochida, H. Tanzawa and M. Yamauchi, *Biochem. Biophys. Res. Commun.*, 2003, **305**, 484–487.

212 D. J. S. Hulmes, A. Miller, D. A. D. Parry, K. A. Piez and J. Woodhead-Galloway, *J. Mol. Biol.*, 1973, **79**, 137–148.

213 S. Varma, J. P. R. O. Orgel and J. D. Schieber, *Biophys. J.*, 2016, **111**, 50–56.

214 D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, *J. Biomol. NMR*, 1995, **5**, 67–81.

215 I. Bertini, C. Luchinat and G. Parigi, in *NMR of Biomolecules: Towards Mechanistic Systems Biology*, Wiley-VCH, 2012, pp. 154–171.

216 V. Kocman, G. M. Di Mauro, G. Veglia and A. Ramamoorthy, *Solid State Nucl. Magn. Reson.*, 2019, **102**, 36–46.

217 G. Pintacuda, M. John, X.-C. Su and G. Otting, *Acc. Chem. Res.*, 2007, **40**, 206–212.

218 I. Bertini, C. Luchinat, G. Parigi and R. Pierattelli, *ChemBioChem*, 2005, **6**, 1536–1549.

219 L. Benda, J. Mareš, E. Ravera, G. Parigi, C. Luchinat, M. Kaupp and J. Vaara, *Angew. Chemie Int. Ed.*, 2016, **55**, 14713–14717.

220 C. P. Jaroniec, *J. Magn. Reson.*, 2015, **253**, 50–59.

221 M. Ahn, C. A. Waudby, A. Bernardo-Gancedo, E. De Genst, A. Dhulesia, X. Salvatella, J. Christodoulou, C. M. Dobson and J. R. Kumita, *Sci. Rep.*, 2017, **7**, 1–12.

222 M. Tang, D. A. Berthold and C. M. Rienstra, *J. Phys. Chem. Lett.*, 2011, **2**, 1836–1841.

223 K. H. Mroue, N. MacKinnon, J. Xu, P. Zhu, E. McNerny, D. H. Kohn, M. D. Morris and A. Ramamoorthy, *J. Phys. Chem. B*, 2012, **116**, 11656–11661.

224 M. Kaushik, T. Bahrenberg, T. V. Can, M. A. Caporini, R. Silvers, J. Heiliger, A. A. Smith, H. Schwalbe, R. G. Griffin and B. Corzilius, *Phys. Chem. Chem. Phys.*, 2016, **18**, 27205–27218.

225 H. J. Weinmann, R. C. Brasch and G. E. Wesbey, *Am. J. Roentgenol.*, 1983, **142**, 619–624.

226 G. M. Clore and J. Iwahara, *Chem. Rev.*, 2009, **109**, 4108–4139.

227 I. Bertini, C. Luchinat and G. Parigi, *Solution NMR of Paramagnetic Molecules*, Elsevier Science, 2001.

228 I. Solomon, *Phys. Rev.*, 1955, **99**, 559–565.

229 A. J. Vega and D. Fiat, *Mol. Phys.*, 1976, **31**, 347–355.

230 M. Gueron, *J. Magn. Reson.*, 1975, **19**, 58–66.

231 L. Lee and B. D. Sykes, *Biophys. J.*, 1980, **32**, 193–210.

232 B. Bleaney, *J. Magn. Reson.*, 1972, **8**, 91–100.

233 Y. Kobashigawa, T. Saio, M. Ushio, M. Sekiguchi, M. Yokochi, K. Ogura and F. Inagaki, *J. Biomol. NMR*, 2012, **53**, 53–63.

234 R. R. Biekofsky, F. W. Muskett, J. M. Schmidt, S. R. Martin, J. P. Browne, P. M. Bayley and J. Feeney, *FEBS Lett.*, 1999, **460**, 519–526.

235 H. Yagi, K. B. Pilla, A. Maleckis, B. Graham, T. Huber and G. Otting, *Structure*, 2013, **21**, 883–890.

236 G. V. Fazakerley and D. G. Reid, *Eur. J. Biochem.*, 1979, **93**, 535–543.

237 M. Kamalov, P. W. R. Harris, C. G. Hartinger, G. M. Miskelly, G. J. S. Cooper and M. A. Brimble, *Org. Biomol. Chem.*, 2015, **13**, 3058–3063.

238 R. G. Paul and A. J. Bailey, *Int. J. Biochem. Cell Biol.*, 1996, **28**, 1297–1310.

239 N. Verzijl, J. DeGroot, S. R. Thorpe, R. A. Bank, J. N. Shaw, T. J. Lyons, J. W. J. Bijlsma, F. P. J. G. Lafeber, J. W. Baynes and J. M. TeKoppele, *J. Biol. Chem.*, 2000, **275**, 39027–39031.