MASSEY
UNIVERSITY

Unravelling the Molecular Contributions to Collagen Higher Order Structure

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science
In
Biochemistry

At Massey University, Manawatu
New Zealand

Danielle Renee Visser

2019

For Mum and Dad

Abstract

Abnormal levels of cross-linking in fibrillar collagen strands have been shown to cause a number of human and animal diseases. Cross-linking is a vital step in fibrillogenesis and contributes greatly to the structural integrity of collagenous tissues. Conversely, defects in cross-link formation can significantly alter fibrillar organisation and lead to pathogenesis. Because collagen cross-links form on collagen-specific hydroxylated lysine residues, an understanding of the link between hydroxylysine and cross-link concentrations is needed to determine whether the level of hydroxylysine, the stereochemistry of these hydroxylysine residues, or other post-translational modifications such as glycosylation affect the level of cross-linking in tissue. While some research has been done to elucidate the connection between the two in different tissue types from the same animal, little has been undertaken to relate hydroxylation and glycosylation of lysine and hydroxylysine to the concentration and types of cross-links in different species. Furthermore, no research has been done to compare the relative distribution of diastereomers of hydroxylysine even within the same species. In order to make a valid comparison, collagen needs to be purified from skin to a high degree and separated into different collagen types and sub-structures as much as possible. To achieve this, the extraction and purification of collagen from the skins of four different mammalian species displaying different skin tensile strengths has been optimised. Different extraction methods were used to prevent the loss of specific features of the collagens that were characterised that may otherwise be lost. Amino acid analysis revealed that while the ratios of the two hydroxylysine diastereomers differed between different animals and extraction methods, the differences were not significant. Mass spectral analysis of cross-links showed that goat skin differed from the other three animals in its cross-link profile. Amino acid analysis combined with mass spectral analysis revealed that on average 70% of proline residues were hydroxylated, a figure much higher than previously thought. Mass spectral analysis also revealed that there are some differences between the glycosylation pattern of different animals, and the ratios of the different types of collagen which are extracted from each animal. While these findings need to be confirmed, they challenge some long held beliefs about the collagen molecule and provide a firm foundation for future work.

Acknowledgements

Firstly, thank you so much to Associate Professor Gill Norris for your support, patience and guidance. I knew during undergrad that I wanted you to be my supervisor, and I am so glad that I made that decision - you have been the best supervisor I could have possibly asked for. Go proteins!

Thank you to Dr. Meekyung Ahn, for your wisdom and enthusiasm for collagen and science in general, and for the many collagen-based chit chats around the tea-room table (sorry to the rest of X-lab for this)! Thanks also to Catherine Maidment, for your tips and tricks of the trade along the way, and for always being able to whip out a gel or reagent recipe whenever I needed one.

Thank you to Dr. Rafea Naffa for your many tried and tested methods and invaluable knowledge of protein chemistry and purification, and for your infectious enthusiasm. I'm proud to be able to add my own work to follow on from your own superb research.

Thanks also go to Trever Loo, whose technical support and expertise in the use of most of the equipment utilised in this project have been instrumental.

To the rest of X-lab and SFS, thank you from the bottom of my heart for making every day a blast, no matter the challenges you always provided a sympathetic ear, an alternative method or theory, and a good debate around the lunch table! Thanks #lunchclub #labladies

Finally, thank you to my family, especially to my Mum and Dad for your unfailing love and support. It was your belief in me that encouraged me and supported me through this big change in career and made me believe that I could undertake this challenge successfully!

List of Abbreviations

3D	Three-dimensional
aa	Amino acid
Å	Ångström (0.1 nm)
AcOH	Acetic acid
C_3H_7OH	Propan-2-ol (isopropanol)
C_4H_9OH	Butan-1-ol (butanol)
cm	Centimetre
da	Dalton
°C	Degrees Celcius
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
g	Gram
GAG	Glucosaminoglycan
GlcNAc	N-acetylglucosamine
HCl	Hydrochloric acid
Hyl	Hydroxylysine
Hyp	Hydroxyproline
kDa	Kilodalton
L	Litre
M	Molar (moles per litre)
MeCN	Acetonitrile
MeOH	Methanol
mg	Milligram
µg	Microgram
mL	Millilitre
µL	Microlitre
mM	Millimolar (10^{-3} molar)
µM	Micromolar (10^{-6} molar)
µS	Micro siemens

MW	Molecular weight
$NaBH_4$	Sodium Borohydride
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
$NH_4CH_3CO_2$	Ammonium Acetate
nM	Nanomolar (10^{-9} molar)
PAGE	Polyacrylamide gel electrophoresis
PG	Proteoglycan
pM	Picomolar (10^{-12} molar)
PTM	Post-translational modification
SDS	Sodium dodecyl sulfate
TEMED	*N,N,N',N'*-tetramethylethylenediamine
T	Temperature (°C)
V	Volts
v/v	Volume/volume
w/v	Weight/volume
x g	Multiple of earth's gravitational force

Amino Acid Abbreviations

Full Name	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Hydroxylysine	Hyl	
Hydroxyproline	Hyp	
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

Table of Contents

Abstract i
Acknowledgements iii
List of abbreviations iv
Amino acid abbreviations vi
Table of contents vii
List of figures xi
List of tables xvi
1. Introduction 1
1.1 Skin structure 4
1.2 A short history of collagen 5
1.3 Collagen family and structure 6
1.4 Collagen biosynthesis 11
1.5 Post-translational modifications 15
1.5.1 Prolyl hydroxylation 15
1.5.2 Lysyl hydroxylation 16
1.5.3 Hydroxylysine diastereomers 18
1.5.4 Glycosylation 19
1.6 Cross-links 20
1.6.1 Immature cross-links 21
1.6.2 Mature cross-links 23
1.6.3 Histidine-containing cross-links 23
1.7 Fibril orientation 24
1.8 Aims of the study 25
2. Materials and methods 29
2.1 General materials and methods 29
2.1.1 Sterilisation of buffers and equipment 29
2.1.2 Protein assays 29
2.1.2.1 Bradford assay 29

2.1.2.2 Sircol collagen assay (SCA) 29
2.1.2.3 Bicinchoninic acid assay (BCA) 30
2.1.3 SDS-PAGE 30
3. Collagen extraction and purification 33
3.1 Introduction 33
3.2 Methods 33
3.2.1 Skin preparation 33
3.2.2 Reduction of cross-links 33
3.2.3 Skin dicing and pulping 34
3.2.4 Extraction of collagen from skin 34
3.2.5 Separation from skin 34
3.2.6 NaCl precipitation 35
3.2.7 Dialysis and lyophilisation 35
3.2.8 Collagen re-solubilisation 35
3.2.9 Initial determination of protein concentration 35
3.3 Results and discussion 36
3.3.1 Enzymatic extraction 36
3.3.2 Pepsin inactivation 39
3.3.3 Boiling *vs* non-boiling of samples prior to SDS-PAGE 44
3.3.4 Pre-treatment *vs* straight to pepsin 46
3.3.5 Acidic and neutral extractions to retain telopeptide regions 47
3.3.6 Final extraction method 48
3.3.7 Putative collagen yield and purity 51
3.4 Conclusion 55
4. Amino acid analysis 57
4.1 Introduction 57
4.2 Methods 58
4.2.1 Hydrolysis of collagen 58
4.2.2 AQC derivatisation 59
4.2.3 High performance liquid chromatography (HPLC) and amino acid analysis (AAA) 60

4.3 Results and discussion 61
4.3.1 Optimisation of the amino acid elution profile 61
4.3.2 Improving the resolution of the hydroxylysine diastereomers 67
4.3.3 Calculating the response factors of each amino acid 69
4.3.4 Hydroxyproline correction 72
4.3.5 Determination of the amino acids in collagen for each species 72
4.3.6 Hydroxylysine and its diastereomers 84
4.3.6.1 In an acidic environment 84
4.3.6.2 In a neutral environment 86
4.4 Conclusion 87
5. Cross-links and glycosylation 91
5.1 Introduction 91
5.2 Methods 92
5.2.1 Hydrolysis of collagen 92
5.2.2 Cross-link enrichment by CF-11 column 93
5.2.3 Cross-link separation and quantitation 93
5.2.4 Cross-link separation 93
5.2.5 Software and instrumentation 94
5.2.5.1 Cross-link analysis 94
5.2.5.2 Collagen PTM analysis 94
5.2.6 In-solution tryptic digestion of collagen 94
5.3 Results and discussion 95
5.3.1 Mass spectral identification of collagen type I cross-links 95
5.3.2 Glycosylation of collagen type I 102
5.3.3 Glycosylation differences between species 106
5.3.4 Prolyl hydroxylation 106
5.3.5 Ratio of collagen types and α-chains 107
5.4 Conclusion 109
6. General summary and conclusions 113
6.1 Overall differences in collagen modifications between species and extraction types 113

6.2 Future directions 116
7. References 117
8. Appendices 135

List of Figures

Figure 1. Cross-section diagram of skin structure. Original artwork..... 4
Figure 2. Hydroxylysine and its diastereomer allohydroxylysine..... 19
Figure 3. An overview of collagen cross-link substrates and products. Reprinted from Methods, 45, Eyre et al., Advances in collagen cross-link analysis, Pg. 68, Copyright (2008), with permission from Elsevier..... 22
Figure 4. 5 % SDS-PAGE gels after the first pepsin digestion at 25° C (P), followed by 3 M NaCl precipitation at 4° C (N) of collagen from sheep and cow skins. White arrows indicate the positions of the α1 and α2 chains of collagen type I, yellow arrows represent β form of collagen (two chain dimer), and the green arrows denote the probable γ form of collagen (three chain or more association). All samples are a 10-fold dilution of the protein extract. Gel was stained with colloidal Coomassie stain overnight..... 37
Figure 5. 5 % SDS-PAGE gels after second pepsin digestion (P), followed by NaCl precipitation (N) of collagen from sheep and cow skins. P1N1 is one pepsin extraction followed by one NaCl precipitation. P2N1 is the second pepsin treatment followed by one NaCl precipitation. P1 is the first pepsin extraction without precipitation. Fractions above lanes denote dilutions of sample prior to loading. Black arrows denote β-collagen strand positions. 39
Figure 6. Pepsin extraction of collagen at 25° C followed by pH inactivation of pepsin at pH 9.0, with a subsequent decrease to pH 7.0. Precipitation in 3 M NaCl was carried out in 10 μM sodium phosphate buffer, pH 7.4. The supernatant was dialysed against MilliQ water to remove excess salt before analysis by 5 % SDS-PAGE. P is after pepsin extraction, PNF is after precipitation, dialysis and lyophilisation. All samples were diluted 10-fold in AcOH prior to loading. White arrow shows putative β-collagen position. 40
Figure 7. 5 % SDS-PAGE gels of collagen from four different species skins after pepsin digestion and 3 M NaCl precipitation, dialysis and lyophilisation. S is sheep skin, C is cow skin, D is deer skin, and G is goat skin. Arrows denote proposed α, β, and γ forms of collagen as indicated. Samples were diluted 10-fold in AcOH prior to loading. 41
Figure 8. 5 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and 4° C (25C and 4C, respectively). 25C and 4C Samples were diluted 10-fold in acetic acid

prior to loading. 25P and 4P denotes acetic acid containing pepsin only and no skin samples, to test pepsin viability after 24 h incubation at the respective temperatures. 42
Figure 9. Comparison between 25°C and 4°C pepsin incubation temperatures, and 0.8 M NaCl and 3 M NaCl collagen precipitation on (a) 5 % and (b) 15 % SDS-PAGE gels. P is pepsin only lane. 43
Figure 10. 15 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and 4° C and precipitation with 0.8 and 3 M NaCl. Samples were diluted 10-fold in 0.5 M acetic acid prior to loading. U is unboiled samples and B is boiling for 5 mins prior to loading. 44
Figure 11. 5 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and precipitation with 0.8 and 3 M NaCl. Samples were diluted 10-fold in 0.5 M acetic acid prior to loading. U is unboiled samples. 2 and 4 are with boiling for 2 and 4 mins respectively prior to loading. 1 h samples were heated to 80 °C in a heating block for 1 h prior to loading. 45
Figure 12. 15 % SDS-PAGE gels comparing "Straight to pepsin extraction" (P-COW) *vs* 0.1 M NaOH treatment of skin followed by defatting with 10 % Butyl Alcohol before pepsin extraction (N-COW). U is unboiled prior to loading, compared to heating for 2 mins at 85 °C. Fractions above lanes denote dilution of extractions in 0.5 M acetic acid. 46
Figure 13. 5 % SDS-PAGE gels showing pepsin, acid and neutral extractions for both sheep and cow skin. In pepsin and acid extractions, B is before NaCl precipitation, A is after precipitation and before lyophilisation. For neutral extractions, A is after concentration and F is the flow through. α-, β-, and γ- collagens are labelled on collagen I standard lane (Coll). 47
Figure 14. Extraction and purification of collagen from raw skin using pepsin, acidic, and neutral extraction methods, after optimisation of each stage of extraction and purification. 49
Figure 15. Method for preparation of pepsin, acidic, and neutral extractions of collagen for analysis on SDS-PAGE gels and subsequent future molecular analysis. 50
Figure 16. Yields of collagen from dry skin weight across extraction types and species. Pink represents pepsin extraction, blue is acid extraction, and green is neutral (salt) extraction.

Pale versions of each colour represent samples reduced with $NaBH_4$ prior to extraction. 54
Figure 17. Amino acids and amines react with AQC to produce highly stable, fluorescent derivatives. The excess reagent reacts with water to form a free amine. 60
Figure 18. Effect of temperature change on elution profile of amino acids from C18 column. Mobile phase consisted of a gradient of 5 mM ammonium acetate pH 5.05, and 60% acetonitrile. 61
Figure 19. HPLC chromatogram of the separation of amino acid standards derivitised with AQC dissolved in DMSO. Buffer A was 5 mM ammonium acetate pH 5.05 and Buffer B was 60% acetonitrile and the total run time was 120 mins. All amino acids were eluted from the column after 70 mins. 63
Figure 20. Comparison of the chromatographic separation of amino acid standards with 5 mM, 20 mM and 50 mM buffer A solutions. 64
Figure 21. Example of improvement to chromatographic separation over several equilibration runs. Shown here are run number 3 (blue line) and run number 6 (black line). 65
Figure 22. Improved resolution of hydrophobic amino acids in HPLC of amino acid standard eluted over a 130 minute run time. 66
Figure 23. Comparison of different column oven temperatures on amino acid elution profile. 67
Figure 24. Standard curve of amino acid standards as eluted from a C18 column with RP-HPLC. 69
Figure 25. RP-HPLC of the amino acids extracted from four different species skins using pepsin as eluted from a C18 column. 70
Figure 26. RP-HPLC of the amino acids extracted from four different species skins using acetic acid as eluted from a C18 column. 71
Figure 27. Amino acid molar % across all four species, as determined by RP-HPLC amino acid analysis. A is the amino acid profile of pepsin extracted collagen; B is from acid extracted collagen. 76
Figure 28. Combined proline and hydroxyproline % molarity measured for each animal versus the theoretical combined values determined by the FASTA sequence. 79

Figure 29. Comparison of number of amino acid residues found experimentally and in the literature for either bovine skin or bovine collagen.....82
Figure 30. Ratio of Hyl1 to Hyl2. Positive bars rising above the 0.00 line indicates a higher level of Hyl2 present, while bars falling below the 0.00 line indicate higher levels of Hyl1. 85
Figure 31. HPLC chromatogram of salt-extracted collagen type I enzymatically hydrolysed sequentially with actinidin, proteinase K and carboxypeptidase y. The black line denotes the chromatogram of the enzymatically hydrolysed sample (no spike), and the blue line is this same sample with the addition of a racemic Hyl standard containing both diastereomers, added to the sample prior to hydrolysis (spiked sample). Arrows denote potential Hyl diastereomer peaks. 86
Figure 32. The four classifications of collagen subjected to cross-link analysis.....92
Figure 33. Mass spectral identification the standards of seven cross-link types found in collagen type I; DHLNL (dihydroxylysinonorleucine), HLNL (hydroxylysinonorleucine), HHL (histidinolysinonorleucine), HHMD (histidinohydroxymerdesmosine), LNL (lysinonorleucine), DPYR (deoxypyridinoline), and PYR (pyridinoline). 96
Figure 34. Comparison of the molar percent of each type of intermediate, reducible cross link across species and extraction type. Pink boxes denote pepsin-extracted collagen, blue boxes are acid-extracted collagen, and green boxes are salt-extracted collagen. Data labels are as follows: the first letter denotes extraction type, the second letter is species (C: Cow, S: Sheep, D: Deer, G: Goat); the third letter denotes whether the skins were reduced with $NaBH_4$ (R) or non-reduced (N). 98
Figure 35. Comparison of the molar percent of each type of mature cross link across species and extraction type. Pink boxes denote pepsin-extracted collagen, blue boxes are acid-extracted collagen, and green boxes are salt-extracted collagen. Data labels are as follows: the first letter denotes extraction type, the second letter is species (C: Cow, S: Sheep, D: Deer, G: Goat); the third letter denotes whether the skins were reduced with $NaBH_4$ (R) or non-reduced (N). 99
Figure 36. Ratio of mature cross-links (HLNL+DHLNL) to immature cross-links (HHMD+HHL) across all four species in reduced skins and non-reduced skins. 101

Figure 37. Percent glycosylation of collagen α chains across four different species. Blue bars denote galactosylation (Gal), red bars denote glucosylgalactosylation (Glcgal). 103
Figure 38. Map showing glycosylation of hydroxylysine residues in collagen α1(I) and α2(I).The left axis shows the total number of technical replicates (n=3) that the modification was found on. The first letter of each bar label relates to the animal (c, d, g, s are cow, deer, goat and sheep, respectively), the second letter/s relate to the sugar attachment (-G is galactosyl, -GG is glucosylgalactosyl). Blue brackets represent the borders of the telopeptide containing collagen molecule, areas outside of the brackets belong to the propeptide and signal peptide regions, which are cleaved during/after secretion from the cell. 105
Figure 39. Percent of proline residues in collagen type I that are detected by mass spectrometry as modified by a hydroxy group (hydroxyproline). Blue columns denote proline residues originating from the collagen α1(I) chain and red columns denote prolines from the collagen α2(I) chain. 107
Figure 40. Percent of collagen types (chains) across species as determined by mass spectroscopy. All proteins were detected with high confidence (low false detection rate (FDR)). 108

List of Tables

Table 1. Collagen groups based on general structural similarities.....8
Table 2. Overview of collagen types, their chains, molecular and supramolecular structure and tissue distribution.....10
Table 3. Glycine SDS-PAGE gel components.....30
Table 4. Overview of methods used for collagen extraction from animal skins.....34
Table 5. Yields of collagen following first and second pepsin digestions.....38
Table 6. Yields of collagen extracted from biological replicates of cow, sheep, deer, and goatskins.....52
Table 7. Binary gradient for elution of AQC derivatised amino acids using RP-HPLC...68
Table 8. Amino acid profile of pepsin extracted collagen expressed as molar percentages. 73
Table 9. Amino acid profile of acid extracted collagen expressed as molar percentages..74
Table 10. Theoretical molar percentages of amino acids based on FASTA data for the COL1A gene in Bos taurus (bovine). 75
Table 11. Comparison of amino acid residues/1000 in bovine collagen or skin as reported by previous groups, in this experimental study, and theoretical (FASTA sequence). A: All amino acids. B: Hyp and Pro combined. 1. Cohen & Seifter [171], 2. Eastoe [169], 3. Gallop et al [170], 4. Deb Choudhury & Norris [168], 5. Naffa [144].....81
Table 12. Comparison of coefficient of variation (CV) between technical replicates for each species and extraction type. P or A before the species denotes pepsin and acid extractions, respectively.....83
Table 13. Molar percentages of hydroxylysine diastereomers as determined by RP-HPLC. 84
Table 14. μmol cross-links per μmol collagen from cow, sheep, deer, and goat skins. NT- before the species name denotes non-telopeptide collagen extracted with pepsin. T- denotes telopeptide containing collagen extracted using either salt (reduced cross-links) or acid (non-reduced cross-links). Cross-links names in the first column that are preceded by R- have been reduced with sodium borohydride. μmol/ μmol collagen was calculated using

the values for the molarity of collagen in each sample calculated in section 4 determined by the hydroxyproline content..... 97
Table 15. Comparison of the FASTA sequence for Col1A (bovine) against the number of hydroxylysine residues found in chapter 4 and the number of glycosylated residues detected by mass spectrometry. Calculated based on glycosylated sites that were detected in ≥2 out of 3 technical replicates..... 103

1. Introduction

The ability to determine stability and structure from primary amino acid sequence is a vital step in the understanding of protein function and the consequences arising from pathological mutations. Essential structural proteins such as collagen, with its hierarchical structure and ubiquity in nearly all tissues are no exception to this rule. Collagen is the major structural protein in the connective tissues in animal bodies, making it the single most abundant protein in vertebrates, constituting from 25% to 35% of the entire body protein content [1]. Collagen makes up one to two percent of muscle tissue, and contributes approximately 6% of the weight of strong, tendinous muscles [2]. It has a highly complex supramolecular structure which is highly diverse across different tissues, allowing it to perform a wide range of biological functions, from cushioning as an extracellular matrix component, to providing rigidity to bone, strength to tendons, and flexibility to skin. The vast universality of collagen across the entire animal kingdom is a clear indication of its significance in biology. Collagen has been identified in nearly all known vertebrates and invertebrates, from hydra to human. The fibrous molecule forming a thick layer around the ejaculatory duct of adult male locusts [3] was found to be very similar to that of mammalian type I collagen [4]. A silk-like protein excreted from the larvae of a sawfly species was shown to originate from three small collagen proteins with all the characteristics of mammalian proteins except for a noticeable lack of hydroxyproline [5]. Collagen is found in the model organism *Drosophila melanogaster* [6], and even in the simplest of all multicellular organisms, the marine sponge [7-9]. But collagen's ubiquity in nature does not begin at the multicellular level. Collagen-like proteins have also been identified in some bacteria that although lacking in hydroxyproline and glycosylated hydroxylysine, still have the Gly-X-Y repeating sequence, the triple helical structure, and the high thermal stability of mammalian collagens [10]. Rasmussen, Jacobsson [11] found not only a number of bacteria but also bacteriophages contained genes which encoded several collagen-like proteins, showing that the collagen sequence motif is found even in the smallest known organism.

Given the ubiquity of collagen across tissues and species, it is clear that this protein is vitally important and warrants extensive study, particularly given its more recent use as a biomaterial. Collagen is present in such a large proportion of human tissues because of its varied and unique properties, such as its ability to form 3D scaffolds in a multitude of different matrices, resulting in mechanical properties which provide not only strength but also flexibility. It is therefore an attractive choice as a raw material for many bio-engineering applications and can be used to produce artificial prostheses that mimic mammalian organs, especially those comprised of collagen type I or II. Because it is a naturally abundant component of the ECM it is already endogenous to the body and is less likely to be viewed as foreign due to its weak antigenicity [12]. It can be extracted in large quantities from a number of different sources such as skin, bone, tendons and the pericardium and is completely biodegradable. Medical grade collagen has been used for decades as wound dressings, surgical sponges, intravaginal contraceptive diaphragms, artificial heart valves and as a matrix for cartilage and bone repair [13]. Recent advances in technology such as 3D printing have seen collagen used to print scaffolds for cell growth in virtually limitless shapes and designs [14-17]. In 2016, the regenerative medicine company CollPlant received funding from Israel's Ministry of Economy for a project focused on the exploration of new uses for their human collagen-like material, including the development of a bioink material for use in the 3D printing of human tissues and organs [18]. The obvious benefits of such applications to biomedical engineering and consequently human and animal health cannot be underestimated. However, the ultimate success of such ventures requires in-depth knowledge of not only collagen's molecular structure and function, but also the many variations between collagen in different tissues and different animals, and the contributions of collagen's many *in-natura* post translational modifications and molecular interactions to its higher order structure.

Collagen molecules interact with one another and with their immediate environment to produce higher order structures with complex hierarchical layers of association related to specific functions. Because it is a major structural component of the extracellular matrix (ECM), collagen plays a vital role in maintaining the structural and biological integrity of cells and tissues. Collagen defects affecting fibrillogenesis have

been implicated in a number of human diseases, including scoliosis and degenerative disc disease, rheumatoid arthritis, oesteogenesis imperfecta (OI) and Ehlers-Danlos Syndrome (EDS). They have also been implicated in the pathogenesis of arterial and myocardial stiffening associated with aging and diabetes [19-27]. In EDS type VI, lysyl hydroxylase enzyme deficiency results in a low hydroxylysine concentration which causes abnormalities in the level of reducible cross-links in bone, cartilage and skin collagens [24]. For example, when the tissues from two sisters diagnosed with the disease were analysed, it was found that the hydroxylysine content of the skin was only 5 % of normal levels [24]. The disease Epidermolysis Bullosa Simplex (EBS) has been shown to be correlated to a deficiency in the enzyme, galactosylhydroxylysyl glucosyltransferase, which catalyses the addition of a second sugar moiety onto an already glycosylated residue of collagen hydroxylysine [28]. Defects in the synthesis and post-translational processing of collagen in cattle has been shown to cause extreme skin fragility. Analysis of their skins showed that impairment to the parallel packing of collagen fibres resulted in loose, disorganised collagen fibres [29, 30]. Because both glycosylation and crosslinking only occur on residues of hydroxylysine in collagen, the number of hydroxylated lysines in the procollagen molecule available for modification will affect the resultant skin structure. When the relative ratio of hydroxylysine diastereomers, which affects the angle of weave of collagen, is added into this complexity, it is clear that a greater understanding of the relationship between collagen cross-links, glycosylation, stereochemistry, and levels of hydroxylysine is required.

Skin organisation, collagen and its many types, sub types, and modifications is, understandably, a vast topic to cover. While this review cannot possibly cover everything there is to know about collagen, it will describe in detail the topics which are central to the context of the hypothesis guiding this study, by providing the background necessary to understand the aspects of collagen involved in this research. In order to keep this plethora of information as logical as possible, a "top down" approach will be taken. It will begin by describing the structure of skin, the arrangement of collagen fibrils and their contribution to skin structure. After describing the structures and functions of the different members of the collagen family the focus will turn to collagen type I, the 'hero' of this research, its functions and its unique structure.

The biosynthesis of type I collagen will then be described, detailing the journey from pro-procollagen to its final form. A comprehensive account of a select few of the post-translational modifications and their roles in collagen and skin structure will then be described. Finally, a hypothesis will be proposed positing a possible reason that explains why the skins of four common New Zealand animals have differences in both strength and collagen fibril orientation.

1.1 Skin structure

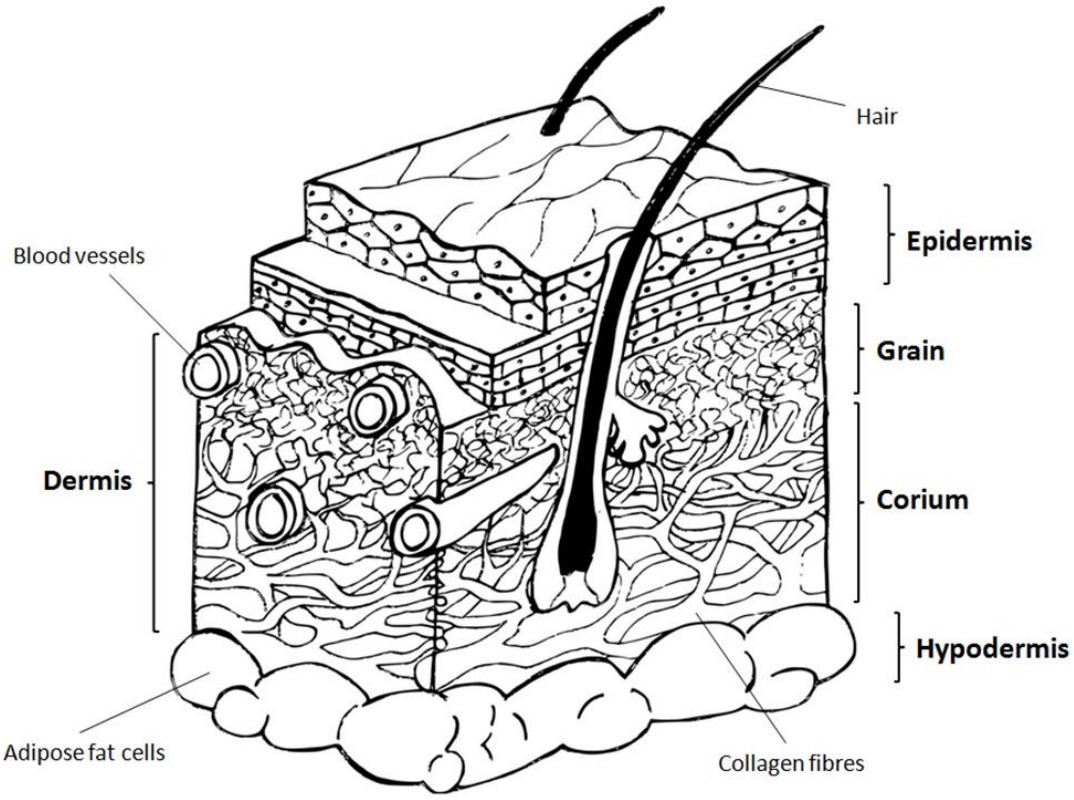


Figure 1. Cross-section diagram of skin structure. Original artwork.

The skin or, integument, of vertebrates is an incredible organ. Not only is it the largest organ in the body, but it is also one of the most complex, with many different cell types contributing to its local structure, along with a multitude of transient cells and molecules from the immune and circulatory systems. It vastly outweighs any other organ of the body in terms of the number of functions it performs, providing not only protection from chemical, physical, pathogenic and irradiation sources to name only a

few, but also thermoregulation, endocrine functions such as vitamin D synthesis, and sensory functions that include non-verbal communication. Skin's complex structure can be divided into three well-defined layers; the epidermis, dermis and hypodermis (Figure 1). The epidermis is the uppermost layer of the skin and is made up mainly of keratinocytes (95%), the remainder being melanocytes, Langerhans cells, and Merkel cells. The hypodermis, also known as the flesh layer, is composed of mainly fat cells. In between these two layers lies the dermis, which makes up 90% of total skin weight and is made up of collagen, elastin, and glycosaminoglycans, collectively termed the extracellular matrix (ECM), as well as fibroblasts which synthesise and secrete them. The dermis is highly vascular, and is interspersed with nerve endings, sweat glands, mast cells, dermal adipose cells, and transient leucocytes. The dermis is often further divided into two layers when discussed from the point of view of leathermakers, into the grain and the corium. The grain is composed of loose fibrous collagen, mainly type III, whereas the corium contains more compact fibres comprised mostly of collagen type I, which run parallel to the surface of the skin [31]. These collagen fibres together account for around 70% of skin protein and are responsible for the tensile strength of skin structure. Alongside, and often interacting with collagen, are other fibrous proteins such as keratin and elastin along with proteoglycans such as dermatan sulphate, and glycosaminoglycans such as hyaluronic acid [32].

1.2 A Short history of collagen

The term 'collagen' was first coined in the 19th century, a derivatisation of the Greek words *kola*, meaning 'glue', and *gennao,* meaning 'I produce', most likely due to the practice at the time of boiling animal skin, tendons and ligaments to produce a crude glue which is now known as gelatin. This adhesive, commonly referred to in historical texts as "animal glue" has been used extensively throughout human history. Hide, or animal "glue" has been discovered in many diverse places such as staffs found at the foot of shaman's tombs at Xiaohe Cemetery, Xinjiang, China, dating to approximately 1500 BCE [33], tile adhesive used in the tombs of the pyramids of Third dynasty Egypt, 2500 BCE [34], and on the surface of bows discovered in Zurich, Switzerland dating to the 4th Millennium BCE [35]. Most recently, one of the oldest samples of animal

glue was discovered on baskets found in a Nahal Hemar cave in Israel, dating back to approximately 8000 BCE [36]. Theophilus Presbyter, a German Benedictine monk and craftsman, wrote about animal glue in his extensive three-volume work about the techniques of all known contemporary crafts; *De Diversis Artibus, ca* 1140. These included the manufacture and use of adhesives made from animal hides and horns, and fish bladders [37]. Theophilus provided a recipe of sorts, suggesting, when making animal glue:

> *"You test it in this way; moisten your fingers in the water (in which the hides and horns have been boiled), and if, when they are cool, they stick together, the glue is good; but if not, heat it until they do stick together."*

Leather made from various animal hides has also been used by the human species for an extended period of time. Pinhasi, Gasparian [38] recently reported the discovery of the oldest leather shoe, dating to approximately 3500 BCE, while the "Tyrolean Iceman", a 5300 year old mummy discovered in the Otztal Alps in 1991 had leather clothing made from a number of different animals, including cow, sheep, goat, deer and bear [39].

These discoveries show just how long and how widely humans have identified if not understood the strength and diversity of collagen, the major extracellular matrix (ECM) protein that provides support for cell growth and is responsible for the mechanical properties of connective tissues by way of its spontaneous formation into self-striated fibrils.

1.3 Collagen family and structure

The term "collagen" captures the entire family of glycoproteins that are categorised by several unique features. All members of the collagen family of proteins contain the distinctive, repeating sequence of amino acids in the primary structure each of the three alpha chains, which frequently follows the pattern Gly-Pro-X or Gly-X-Hyp, where Hyp is hydroxyproline (a post-translationally modified form of proline), and X may be almost any other amino acid residue. The X and Y positions are generally occupied

by proline or its post-translationally modified form; hydroxyproline. Proline and hydroxyproline together make up around 1/6 of the total amino acids, and glycine around a 1/3, meaning that approximately half of the collagen sequence is glycine, proline or hydroxyproline. The collagen molecule also consists of a unique quaternary structure formed by the individual three left-handed α-helical polypeptide chains of identical length coming together to create a right-handed triple helix known commonly as tropocollagen. These characteristic features, particularly the Gly-X-Y repeating motif, have been used to identify homologous gene sequences and ultimately the classification of 40 vertebrate collagen genes that produce 29 genetically distinct collagen types to date. While collagen types differ markedly across tissues and species, types I, II and III constitute the majority of total mammalian collagen, making up around 80-90% [40].

Roman numerals are used to identify collagen type, with Greek letters denoting the different chains (α), (β), and the higher molecular weight structure (γ). The three-chained super-helical nature of collagen allows for varying homo and heterotopic combinations of the α-chains. The specific combination of these α-chains comprise isoforms of each collagen type, for example, collagen type I generally exists as heterotrimer of two α1 chains and one α2 chain but can also be found as a homotrimer consisting of three α1 chains. Many collagen types, such as types II, III, and VII, exist exclusively as homotrimers, however, collagen IV, which is encoded by six different α-chain genes can be found in many isoforms which form tissue-specific basement membranes.

While some of the currently identified collagen types exhibit highly unique features, characteristic of specific locations and tissues most of them appear to be structurally conserved. Collagen classification is based on the primary structure, triple helical domain length and interruptions to the Gly-X-Y motif, molecular weight, charge profile along the helix, size and shape, and cleavage or retention of the telopeptides (terminal domains), and variations in the post-translation modifications. As a result, four overarching collagen groups have been identified (Table 1).

Although collagens can be categorised into these groups based on their general primary, secondary, and some tertiary structures, the types within these groups vary greatly not only in both location and function, but also, in some cases, their quaternary structure. For example, although similar in primary structure, the short chain type VIII and type X collagens form hexagonal lattices, while type VI collagen forms beaded microfilaments. To get a snapshot of the enormous variety in function, location and structure of the different collagen types, they have been summarised in Table 2. Many

Table 1. Collagen groups based on general structural similarities

Group	*Type*	*Description*
Fibrillar	Type I, II, III, V, XI, (XXIV & XXVI)	Fibril-forming collagens
Basement membranes	Type IV, VII, (XXVIII)	Forms basal lamina
Short chain	Type VI, VIII, X	Named for their triple helical regions which extend 100-150 nm.
Collagens with helical interruptions	Multiplexins: Type XV, XVIII FACITs: Type IX, XII, XIV, XVI, XIX, XX, XXI MACITs: Type XIII, XVII	Multiplexins are Multiple Triple Helix domains with Interruptions, FACITs are Fibril Associated Collagens with Interrupted Triple Helices, and MACITs are Membrane Associated Collagens with Interrupted Triple Helices. Also include some transmembrane collagens.

other proteins related to collagen have also been identified, such as lung surfactant proteins [41], complement 1q (C1q), collagen-like lectins (collectins) [42], the tail of the asymmetric form of acetylcholinesterase [43], and macrophage scavenger receptors [44]. These proteins all contain triple helical domains composed of Gly-X-Y repeats.

Fibrillar collagens are the largest group by far, because they are major components of skin, bone and tendon. Type I collagen is found in skin, bone, tendon, and vasculature; type II collagen is the main component of cartilage; type III, is the main component of reticular fibers and a minor component of skin; type IV, forms the secreted epithelium layer of the basement membrane, the basal lamina; type V collagen is found

in hair, on cell surfaces, and in the placenta as well as being a minor component of skin. All of these except Type IV are fibrillar collagens. The fibrillar collagens also include another less common collagen, type IX, which is found in fetal cartilage where it forms thin fibrils and acts as a regulator of collagen types I and II fibrillogenesis [45, 46].

Fibrillar collagen molecules are defined by an uninterrupted central triple helical region and a nonhelical region at both the N and C termini, known as telopeptides. The major fibrillar collagen, and the focus of this research, type I collagen contains no tryptophan or cysteine and is very low in tyrosine and histidine [47]. The high glycine content contributes to the stabilisation of collagen, allowing the very tight packing that permits the formation of the right-handed triple helix as the small glycine residue is the only amino acid that can fit into its core. The three α-chain helices wind around each other forming a super helix that results in a linear structure of ≈300 000 g mole^{-1} with a diameter of ~1.4 nm in diameter and a length of 280 nm. The triple helix is stabilised by hydrogen bonds between the amines of the glycyl residues and the carboxyl groups of the second position residue in adjacent chains. They are further stabilised by hydrogen bonds between the hydroxyl group of a hydroxyproline (Hyp) with a Hyp in an adjacent chain *via* bridging water molecules resulting in an average of two hydrogen bonds per triplet [48, 49]. These properties of the individual α-chains result in elongated left-handed polyproline II-like helices [50]. Three left-handed helices are then drawn together by weak interactions around a central axis in an arrangement that results in each helix being twisted by approximately 30° at every turn and staggered by one residue relative to each other. The resulting right-handed superhelix has a conformation in which every third amino acid lies at the centre of the superhelix, with only glycine able to occupy this position due to steric hinderance. The alicyclic structures of proline and hydroxyproline in the adjacent positions stiffen the structure of the α-chain, helping to stabilise the superhelix by preventing rotation of the C-N bond.

Collagen also contains another uncommon post-translationally modified amino acid; hydroxylysine (Hyl). Hyl acts as a substrate for enzymes which transfer galactose or

Table 2. Overview of collagen types, their chains, molecular and supramolecular structure and tissue distribution.

Collagen type	Chains	Molecular assembly	Supramolecular structure	Mw [kDa]/α-chain	Tissue distribution
I (Heterotrimer)	$[\alpha1(I)]_2\,\alpha2(I)]$	Monomers staggered by 67 nm	Large diameter, 67 nm banded fibrils	95	Skin, tendon, ligament, cornea, organ capsules, dura mater of brain and spinal cord, the main organic component of bone
I (Homotrimer)	$[\alpha1(I)]_3$	Monomers staggered by 67 nm	67 nm banded fibrils		Tumours, dermis, bone 95 95
II	$[\alpha1(II)]_3$	Monomers staggered by 67 nm	67 nm banded fibrils	95	Cartilage, vitreous, cartilagenous zones of tendon, intervertebral disc
III	$[\alpha1(III)]_3$	Monomers staggered by 67 nm	Small diameter, 67 nm banded fibrils	95	Dermis, aorta, uterus, admixture in tendon, intestine, blood vessels, in the reticular connective tissue of liver, spleen and surrounding internal organs
IV	$[\alpha1(IV)_2\,\alpha2(IV)]$, α3(IV), α4(IV), α5(IV), α 6(IV)	Association of 4N- and 2C-termini	Nonfibrillar meshwork	170–180	Basement membranes
V	$[\alpha1(V)]_2\,\alpha2(V)$, [α1(V) α2(V) α3(V)],$[\alpha1(V)]_3$	Monomers staggered by 67 nm	9 nm diameter banded fibrils	120–145	Placental/embryonic tissue, dermis, bone, cornea, cell surfaces
VI	[α1(VI) α2(VI) α3(VI)]	Association into tetramers that aggregate end to end	5–10 nm diameter, beaded microfibrils, 100 nm periodicity	αI(VI) 140, α2(VI) 140, α(VI) 340	Uterus, dermis, cartilage, muscle
VII	XVI [α1(VII)]3	Lateral aggregation of antiparallel dimers	Anchoring fibrils	170	Skin, amniotic membrane, cornea, mucosal epithelium
VIII	$[\alpha1(VIII)]_2\,\alpha2(VIII)$	Interrupted helical structure	Nonfibrillar, hexagonal lattice	61	Descemet's membrane, endothelial cells
IX	[α1(IX) α2(IX) α3(IX)]	Covalently crosslinked to surface of collagen II fibrils	FACIT; nonfibrillar	68–115	Cartilage, vitreous, admixture in tendon, co-distributes with collagen II
X	$[\alpha1(X)]_3$	Assemble a mat-like structure	Nonfibrillar, hexagonal lattice	59	Calcifying cartilage (including parts of tendons)
XI	[α1(XI) α2(XI) α3(XI)]	Monomers staggered by 67 nm	Fine fibrils similar to those of collagen V	110–145	Cartilage, intervertebral disc
XII	$[\alpha1(XII)]_3$	Associates with surface of collagen fibrils	FACIT; nonfibrillar	220, 340	Dermis, tendon, cartilage
XIII	$[\alpha1(XIII)]_3$	150 nm rod with two flexible hinges	Transmembrane	62–67	Endothelial cells, epidermis
XIV	$[\alpha1(XIV)]_3$	Disulfide-linked cross-shape	FACIT; nonfibrillar	220	Dermis, tendon, cartilage
XV	$[\alpha1(XV)]_3$	Figure eight knot configuration	MULTIPLEXIN; nonfibrillar	125	Placenta, kidney, heart, ovary, testis
XVI	$[\alpha1(XVI)]_3$	Associates with dermal fibrillin; associates with banded collagen in cartilage	FACIT; nonfibrillar	150-160	Heart, kidney, muscle
XVII	[α1 (XVII)]3	Shed from cell surface into shorter soluble form	Membrane-intercalated	180	Hemidesmosomes (skin), specialized epithelia
XVIII	[α1(XVIII)]3		MULTIPLEXIN; nonfibrillar	200	Kidney, liver
XIX	[α1(XIX)]3	Sharply kinked and higher order complexes	FACIT; nonfibrillar	165	Transitory embryonic expression, interneurons and formation of hippocampal synapses, basement membranes, muscle cell, rhabdomyosarcoma
XX	[α1(XX)]3	Binds to collagen fibrils with amino terminal domains away from fibrillar surface	FACIT	185, 170, and 135	Corneal epithelium, embryonic skin, sternal cartilage, tendon
XXI	[α1(XXI)]3		FACIT		Blood vessel walls, secreted by smooth muscle cells
XXII	[α1(XXII)]3	Associates with cartilage microfibrils	FACIT	200	Tissue junctions
XXIII	[α1(XXIII)]3		Transmembrane		Tumors (prostate)
XXIV	[α1(XXIV)]3	Associates with vertebrate fibrillar	Fibrillar, fibril associated		Regulation of collagen I fibrillogenesis, osteoblast differentiation marker
XXV	[α1(XXV)]3	Binds to fibrillized Aβ	Transmembrane	50/100	Interaction with β amyloid plaques in Alzheimer's disease
XXVI	[α1(XXVI)]3		FACIT	≈80	Ovary and testis
XXVII	[α1(XXVII)]3	10 nm network organization	Thin nonstriated fibrils		Hypertrophic cartilage
XXVIII	[α1(XXVIII)]3	Associates with nonmyelinated regions	Beaded filament forming	≈50	Basement membrane of Schwann cells, peripheral nervous system

glucose molecules to the γ or δ-hydroxyl groups of the Hyl sidechain. These same hydroxyl groups are modified by the enzyme lysyl oxidase, a prerequisite for the formation of cross-links. Studies have shown that once the triple-helical collagen structure is formed, Hyl can no longer be glycosylated [51]. Hydroxylysine glycosylation thus controls which lysines are involved in forming crosslinks, which in turn, influences fibril diameter [52]. Thus, the biosynthesis of collagen needs to be highly ordered and tightly controlled to ensure that the many modifications which are required to form fully functioning fibrils are achieved to a high level of precision.

1.4 Collagen Biosynthesis

With collagen structure so hierarchically ordered in tissues it is not surprising that its biosynthesis is a tightly ordered hierarchical process, involving many steps and multiple chaperones. The initial step of intracellular biosynthesis of the collagen molecule involves the transcription of mRNA molecules specific to the type and isoform of collagen being encoded by various three-chain combinations of α-chain *COL* genes. The initial form of collagen is called pre-procollagen, which is then converted into procollagen upon removal of the signal peptide. Two of the major post-translational modifications (PTMs) of collagen, hydroxylation of some proline and lysine residues, and glycosylation, also occur within the ER lumen on the nascent polypeptide. Super-helical formation begins at the C-terminus, however, in order for the signal peptides to be removed and co-translational modifications to occur the pro-α-chains need to remain untangled for the length of time required to complete the modifications and finalise translation, after which the correct three pro-α-chains must align precisely at the C-termini to allow triple helix formation to begin [53]. This process is mediated by several chaperone proteins which prevent the α-chains from getting tangled including prolyl 4-hydroxylase (P4-H), protein disulfide isomerase (PDI), and the molecular chaperone specific to procollagen, heat shock protein 47 (hsp47), among others [54]. More than 20 hsp47 molecules per triple helix are required to stabilise the procollagen helix at body temperature [55]. It has been recently suggested that the intracellular SPARC (Secreted Protein Acidic and Rich in Cysteine) protein may also be a collagen-

specific chaperone, as it has been found to bind to the triple-helical domain of procollagens, and its downregulation leads to defects in collagen deposition [56].

Because different isomers of collagen require a specific combination of α-chains, it is essential that the α-chains associate in a controlled manner to eliminate the possibility of incorrect pairing. For example, skin fibroblasts synthesise three fibril-forming collagens; types I, III and V. These are comprised of six homologous but different α-chains; α1(I), α2(I), α1(III), α1(V), α2(V) and α3(V), which require assembly in a specific manner for each collagen type. Unlike the nucleic acids in the base pair-formation of a DNA duplex, the Gly-X-Y sequences in collagen are not capable of complementary recognition in a residue-to-residue manner. Instead, the C-propeptide domains of the procollagen molecule provide a system to ensure collagen-type chain assembly specificity, eliminating the generation of hybrid collagens composed of incorrect α-chain combinations. Observations of artificial deletions and naturally occurring mutations to the C-propeptide domains in fibril-forming procollagen molecules have shown that this domain is essential for ensuring that each α–chain is correctly aligned allowing procollagen folding to begin [57, 58]. Intermolecular disulfide bridges in the C-terminal -propeptide domains (are non-collagenous domains approximately 250 amino acids in length) ensure that the chains associate correctly with one another [59-62]. These domains are highly homologous [63], each containing eight conserved cysteine residues which are oxidised into four cystines by the action of the enzyme protein disulfide isomerase (PDI) during the trimerization and folding step. The four cysteines situated at the N-terminus of the C-propeptide have been shown to be essential for intermolecular stabilisation, while the four cysteines at the C-terminus form intramolecular disulfide bridges [64-67]. A peptide sequence of 23 amino acid residues was found to be responsible for type-specific pro-*α-chain* association, and within this sequence, 15 discontinuous amino acid residues were shown to direct α-chain selectivity [68].

Following the recognition of correct type and chain associations, and formation of the stabilising disulfide bridges, the triple helical regions of the procollagens, which contain approximately 1000 amino acids begin to coil around one another to form a right-

handed super coil, proceeding in the direction of the N-termini. The individual α-chains are not stabilised by hydrogen bonds between the peptide backbone carbonyl and amide groups as in a standard α helix; instead, NH-CO hydrogen bonds are formed between the different α-chains. This requires concomitant formation of both the polyproline helices of individual chains and the triple-helix. Hydroxylation of at least two specific prolines in the C-terminus of the helical region of each α-chain has been shown to ensure that trimerization begins at the correct location between α-chains, as misalignment would result in collagen molecules that had only partial or interrupted triple helical regions [69]. Intermolecular hydrogen bonding and triple-helix formation begins at these residues and is propagated towards the N-terminus. The formation of the triple helix is however a slow process compared to the formation of globular proteins because the prolines and hydroxyprolines need to be in a *trans*-conformation rather than their typical *cis* configuration prior to the formation of the triple helix [70, 71]. This is achieved by the action of peptidyl prolyl *cis-trans* isomerases (PPIases) which also act as molecular chaperones. It is at this point that other PTMs occur, such as hydroxylation of lysine residues by the enzyme lysyl oxidase, and subsequent glycosylation of some of the resulting hydroxylysines to form galactosyl-Hyl and glucosyl-galactosyl-Hyl residues.

After the triple helix has been successfully formed, the procollagen molecule is transported to the Golgi apparatus. Generally, secretory proteins are packaged into vesicles for transport from the ER to the Golgi. As these vesicles are approximately 60-80 nm in diameter, they are far too small for the 300 nm long procollagen molecule. Stephens and Pepperkok [72] showed that procollagen transport utilises a distinctly different pathway from other secretory protein-containing vesicles using recombinant procollagen fused with green fluorescent protein (GFP). Although electron microscopy showed that the morphology of the procollagen-containing ER-to-Golgi transport complexes are tubular, sack-like structures with greater than 300 nm in length, and therefore large enough to contain the entire procollagen molecule, the transport system is still not well understood [73, 74].

Upon entry to the cisternae of the Golgi apparatus the procollagen molecules self-associate into aggregates that are approximately 320 nm long and 170 nm thick [73-75]. It is within the *cis*-Golgi that procollagen molecules are liberated from Hsp47 for the first time. This dissociation of Hsp47 is generally considered to be the likely cause of aggregation of procollagen, and although the functional advantage of forming aggregates is yet to be determined, it is speculated that it may be a way to protect the molecules from attack by proteases in the Golgi apparatus [54]. Aggregate formation may also help to provide thermal stability to the procollagen molecules, which are thermally unstable at body temperature [76]. These procollagen aggregates then move across the Golgi stacks without leaving the lumen of the Golgi cisternae, and then incorporated into secretory vesicles where the aggregates increase in length [77]. Procollagen molecules are then secreted into the extracellular space, most likely in their aggregated form. During or following their secretion into the ECM, the N- and C-propeptide domains are cleaved by specialised N- and C-proteinases, to produce collagen molecules with short, non-triple helical regions of approximately 9–26 amino acids in length at each termini called telopeptides, which perform a vital role in the crosslinking of collagen α-chains. The removal of the propeptide regions is a necessary prerequisite to the self-assembly of collagen molecules into the typical quarter-staggered arrangement in a lateral and head-to-tail fashion resulting in the formation of cross-striated fibrils [78].

Studies have shown that some regulation of collagen biosynthesis is achieved through the action of a negative feedback loop, in which the free propeptides are able to re-enter the cell to pretranslationally regulate collagen production [79]. The presence of the propeptides of intact procollagen also work to prevent the premature supramolecular assembly of fibrils within the cell but are removed from the ECM to allow collagen assembly or fibrillogenesis. While this process is not fully understood, the most likely model has it happening within deep invaginations of the plasma membranes at the surface of fibroblasts. Although enveloped partially by the plasma membrane, the interior of these invaginations is part of the extracellular space. It is within this space that removal of the propeptides occurs, followed by aggregation of tropocollagen, to form fibrillar intermediates that grow out of their invaginated pockets.

1.5 Post-translational modifications

Two major post-translational modifications (PTMs) of collagen, hydroxylation of both proline and lysine residues, and glycosylation of hydroxylysine, occur in the endoplasmic reticulum, while another major modification, cross-linking, occurs after secretion to the extracellular space. These modifications contribute to the mechanical and thermal stability of the collagen triple helix and higher assembly forms. Many studies have reported pathological conditions as a direct result of interference with these PTMs, either by nutritional deficiencies or genetic alterations, demonstrating the importance of these modifications to the structural integrity both at the molecular and fibrillar level.

1.5.1 Prolyl hydroxylation

By matter of necessity, the first post-translational modification of procollagen is the conversion of some Pro residues in the triple helical region into Hyp. This must occur soon after translation and prior to the formation of the triple helix, as Hyp is required for coordination of the network of water molecules residing in the interior of the triple helix of collagen and directs the formation of hydrogen bonds between collagen chains [80, 81]. Hydroxylation appears to be favoured on Pro at the Y position of the Gly-X-Y triplet, as this serves to direct correct alignment of the water bridges necessary for the most stable packing of the collagen triple helix [82, 83]. Prolines residing in the telopeptide regions are not converted to Hyp, as these regions lack the Gly-X-Y motif and are not directly involved in triple-helical formation. Hydroxylation of the helical prolines is catalysed by two enzymes residing in the ER lumen; prolyl-4-hydroxylase (P4Hs), which catalyses proline 4-hydroxylation (4Hyp) at the Y position, and prolyl-3-hydroxylase (P3Hs), which catalyses proline 3-hydroxylation (3Hyp) at the X position. [78] Prolyl 4-hydroxylation is the single most prevalent post-translational modification in humans, which is not surprising, given that collagen is the most abundant protein and a quarter of mammalian fibrillar collagen residues are prolines, approximately 50% of which are 4-hydroxylated. Prolyl-3-hydroxylation however is very rare, with 3Hyp occurring on average at only one to two residues per α-chain in collagen type I. It has been established that Prolyl-3-hydroxylation occurs

predominantly within the sequence Gly-Pro-4Hyp (5,9), resulting in a Gly-3Hyp-4-Hyp sequence. The role of prolyl-3-hydroxylation is not clear but may be related to modifications of the C-termini of collagen type I and III α-chains. These regions are rich in stretches of $[GPP]_5$ and $[GPP]_7$ which seem to be preferentially hydroxylated by P3H and appear to play an important role in the increase of local thermal stability around the C-terminus [84]. There is also some evidence to suggest that P3H may act to prevent premature aggregation of collagen during its synthesis within the cell, by operating as molecular chaperone [85]. Whatever P3H's function, it is clearly an important enzyme for collagen synthesis as its loss has been implicated in autosomal recessive forms of osteogenesis imperfecta [86].

The importance of P4H hydroxylation of helical prolines is clearly demonstrated by the pathological results of long-term vitamin C deficiency. Ascorbic acid, commonly known as vitamin C is an essential co-factor of P4H as deficiency results in procollagen being unable to leave the ER due to improper folding, leading to a lack of synthesis of new collagen fibrils. Humans (as well as primates and guinea pigs) are unable to synthesise vitamin C in the liver and need to obtain it exogenously through their diet. A lack of vitamin C results in scurvy; a disease which is characterized by inadequate renewal of connective tissue [87].

For mammals that maintain a body temperature of around 37 °C, a minimum of 100 Pro residues per α-chain need to be converted to Hyp to achieve thermal stability. This level of prolyl hydroxylation in mammals equates to approximately 13.4% of total amino acid content, and this figure has been used for the last seven decades as a yardstick for determining total collagen content from total protein content in hydrolysates [88]. Less hydroxylation has been observed in cold blooded species such as arctic cod [89].

1.5.2 Lysyl hydroxylation

Human Type I collagen has 38 Lys residues in the α1 chain (one in each of the C and N telopeptide regions and 36 in the helical domain) and 31 Lys residues in the α2 chain (one in the N telopeptide region but not in the C telopeptide, and 30 in the

helical domain). The positions of Lys residues in the α1 chain are highly conserved between different species, while there are differences in the quantity and distribution of Lys in the α2 chain. Mouse and rat type I α2 collagens have one extra Lys residue in their helical domains giving a total of 32 rather than 31, and some of the sites vary (Yamauchi & Sricholpech, 2012). Lys residues involved in covalent cross-links in both the α1 and α2 chains are highly conserved between species. Analysis has shown that lysine hydroxylation is highly variable in relation to the hydroxylation of proline. Approximately 50 % of proline residues are hydroxylated in the collagens of different tissues, in contrast to lysine hydroxylation, which can vary between 15-90 % with collagen type. Significant variation is seen between type I collagens alone and depends on tissue type as well as the physiological and pathological conditions of the host cells. Hydroxylation of lysine residues in the α-chains of procollagen to 5-hydroxylysine (Hyl) is achieved through the action of lysyl hydroxylase (LH) which is located in the rough ER. Hyl residues are found in both the telopeptide and the triple helical regions of collagen, and in both of these locations specific Hyl residues are vital to the formation of covalent cross links. Hyl residues also serve as an attachment site for certain monosaccharides (more on this in the following section). Studies of mice, rats, and humans have shown there are three lysyl hydroxylase isoenzymes; LH1, LH2, and LH3, and that the expression of these LH genes is tissue dependent [90, 91], and developmentally regulated. LH1 is shown to be continuously expressed throughout mice embryonic development, while LH2 and LH3 are only highly expressed until day 7 of development, with LH3 expression re-emerging at day 15 [92].

Lysyl hydroxylation within the helical region of collagen occurs exclusively on lysines located within the sequence Gly-X-Lys; if a Lys is present in the X position it is not hydroxylated. Hydroxylated telopeptide lysines are not usually found within this sequence, leading to speculation that two classes of LH must exist; one which targets lysines at the specific locations within the triple-helical region, and another which targets Lys resides in the telopeptides.

Collagen-related pathogenesis has so far provided the most direct evidence that both a telopeptide and helical lysyl hydroxylase must exist. Cross-link studies in Ehlers–

Danlos syndrome type VIA and Bruck syndrome have opposing characteristics of pathogenesis. EDS-VIA patients have decreased levels of Hyl in the triple helix yet have normal levels of pyridinoline cross-links [93, 94]. Pyridinoline cross-links are formed with telopeptide hydroxyallysines, and therefore, despite lowered levels of Hyl in the helical region in EDS-VIA, a normal amount of Hyl is present in the telopeptides, indicating that it is the helical lysyl hydroxylase gene which is mutated in this disease. Conversely, in Bruck syndrome, patients exhibit normal Hyl levels in the helical region, yet pyridinoline cross-links are virtually absent [95]. Despite normal levels of helical Hyl, telopeptide Hyl is deficient, indicating that the telopeptide-specific lysyl hydroxylase is either absent or inactive. In EDS-VIA the PLOD1 gene [96, 97], which expresses the LH1 isotype was found to be mutated, indicating that LH1 is a helical-specific lysyl hydroxylase. In Bruck syndrome the PLOD2 gene was shown to be mutated [98-101], showing that LH2 is a telopeptide-specific lysyl hydroxylase. Several other studies show that increased levels of LH2 expression correlate with an increase in hydroxylation of the telopeptides [100, 102-105] supporting the importance of these genes. More recently it has been shown that mutations to a gene which encodes the prolyl *cis–trans* isomerase, FKBP65, also contributes to under-hydroxylation of lysine residues in some Bruck Syndrome and Oesteogenesis Imperfecta patients, suggesting that incorrect orientation of neighbouring proline residues may affect binding of LH2 to its substrate sequence [106].

1.5.3 Hydroxylysine diastereomers

With hydroxylation of lysine residues being such a vital step in the creation of cross-links which stabilize the collagen structure, it stands to reason that the orientation of these residues and modifications will greatly impact the direction of the resultant cross-links. In 1954, Piez discovered a possible diastereomer of hydroxylysine (Figure 2) and reported that approximately 20 % of Hyl in acid hydrolysates was in fact the diastereomer, allohydroxylysine [107]. Hamilton and Anderson in 1955 [108] also reported this, but suggested that the allohydroxylysine was formed as a result of the acid conditions used in amino acid hydrolysis rather than occurring naturally. Several other researchers [109-111] subsequently confirmed the existence of hydroxylysine diastereomers after basic hydrolysis. Witkop [112] showed that acid treatment of

hydroxylysine resulted in the formation of a lactone whereas acid treatment of allohydroxylysine did not, suggesting that the latter is a more stable compound which does not support Hamilton and Anderson's 1955 conclusion. There have been no further reports on the legitimacy of hydroxylysine diastereomers in skin. Thus, whether they are truly native to skin structure, or whether they are purely a product of some chemical reaction involved in the isolation of collagen from skin remains to be established. If they do exist *in vivo*, they would provide yet another possible influence on collagen fibril conformation in different skin types.

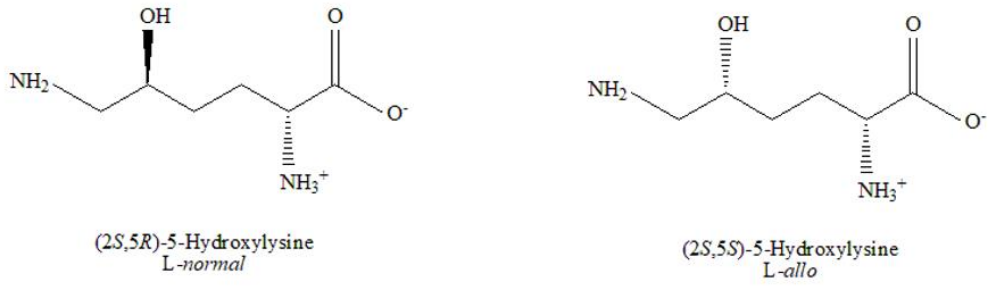


Figure 2. Hydroxylysine and its diastereomer allohydroxylysine.

1.5.4 Glycosylation

Glycosylation is a very common post-translational modification that occurs when enzymes such as glycosyl and oligosaccharyl transferases covalently attach a monosaccharide or oligosaccharide to a polypeptide chain. Linkage can either be *via* the hydroxyl oxygen of serine threonine, or hydroxylysine side chains (O-linked glycosylation) or *via* the amido nitrogen of an asparagine sidechain. For a glycoprotein, type I collagen has a comparatively low carbohydrate content at less than 1 %. Specific hydroxylysine residues are O-glycosylated at the 5-hydroxyl group with a monosaccharide β-D-galactopyranose, (Gal) or a disaccharide α-D-glucopyranosyl-(1-2)-β-D-galactopyranose, (GlcGal). This results in two glycosylated forms of Hyl: galactosylhydroxylysine (GHyl) and glucosylgalactosylhydroxylysine (GGHyl) [113].

While glycosylation is necessary for the correct functioning of collagen, the extent of glycosylation of hydroxylysine varies greatly between tissues (Spiro *et al*, 1971). Studies with collagen produced by chick embryo fibroblasts showed that some hydroxylated

lysines were preferentially glycosylated while the nascent collagen polypeptide chain was still bound to the ribosome, and that these could be identified by amino acid analysis following alkaline hydrolysis, but not after acid hydrolysis (Brownell & Veis, 1975). Interestingly, extracellular collagen contained more free hydroxylysine than glycosylated hydroxylysine suggesting that lysine residues destined to become glycosylated in mature collagen are preferentially hydroxylated while the nascent polypeptide chain is still ribosome bound.

Human lysyl hydroxylase isoform 3 (LH3) performs both lysyl hydroxylase and glucosyl transferase (GGT) reactions [114]. Studies of LH3 transfection in osteoblast cultures showed it was responsible for the production of five glycosylated Hyl residues in type I collagen, one of which is a major helical cross-linking site. Further experiments showed that manipulation of the glycosylation of these sites altered the cross-linking and fibrillogenesis of the tropocollagen molecules [115]. This supports the long-standing view that the attachment of sugar components to Hyl residues is important for modulation of fibrillogenesis. In addition, analyses of fibrillar collagens have shown that increased glycosylation of hydroxylysine residues is associated with decreased collagen fibril diameter [116, 117].

1.6 Cross-links

Collagen's strength, enzymatic resilience and stability is further enhanced by the formation of strong intermolecular crosslinks. In collagen type I a total of only four cross-links have been identified: two in the telopeptide region of the collagen molecule and two in the triple-helix. The enzyme responsible for the initiation of formation of these cross-links is lysyl oxidase (LO), an extracellular, copper-dependent enzyme that catalyses the formation of aldehydes from lysine and hydroxylysine residues, by converting the ε-NH_2 group of lysine to an aldehyde [87]. These newly formed side chain aldehydes are extremely volatile and will spontaneously condense with the ε-amino groups of other lysine or hydroxylysine sidechains to form Schiff bases (aldimines), or with other aldehydes to form aldol crosslinks [118-120]. The covalent cross-links thus formed are essential for the formation and stabilisation of collagen fibrils, and ultimately the integrity of the skin structure. Cross-links can also form

between two different mature collagen molecules resulting in a head-to-tail bonds called aldimide bridges that form along the fibrils, [121]. The formation of different forms of cross-links, either immature or mature depends largely on their location in the collagen molecule, whether in the helical region or either telopeptide. Figure 3 gives an overview of the current understanding of cross-link formation. Because hydroxylysine residues are essential for cross-link formation, and these are usually in specific locations in the collagen sequence, the number and locations of Hyl residues will direct the position and type of cross-link formed which may be different between species. Given that skin collagen molecules are known to be glycosylated, and that these modifications are to Hyl residues it stands to reason that Hyl residues which are glycosylated will not be able to form cross-links, making glycosylation another controller of cross-link formation.

1.6.1 Immature cross-links

Initial cross-linking begins with deamidation of lysine and hydroxylysine residues to allysine and hydroxyallysine respectively (Figure 3). The lysine pathway leads to aldol condensation and the formation of crosslinks between adjacent α-chains within the triple helix (intramolecular), which can be observed in SDS-PAGE gels as β-bands. The hydroxylysine pathways create ketoimine and aldimine cross-links, which connect adjacent collagen molecules (intermolecular). These "immature" intermolecular cross-links are known as divalent cross-links. Telopeptide hydroxyallysines react with helical hydroxylysines or lysines to form ketoamine cross-links called hydroxylysine-keto-norleucine (HLKNL) and lysine-keto-norleucine (LKNL), respectively. These cross-links are formed through an amadori rearrangement, and mainly occur in bone and cartilage where the telopeptide lysines are highly hydroxylated. In contrast, in tissues such as skin, where telopeptide hydroxylation is low, telopeptide allysines react with helical hydroxylysines or lysines to form the aldimine crosslinks dehydro-hydroxylysinonorleucine (Δ-HLNL) and dehydro-lysinonorleucine (Δ-LNL), respectively [122]. These divalent aldimine cross-links are reducible by treatment with sodium borohydride ($NaBH_4$) [123]. In most tissues, the number of borohydride-reducible cross-links decreases with age due to further maturation into more stable, non-

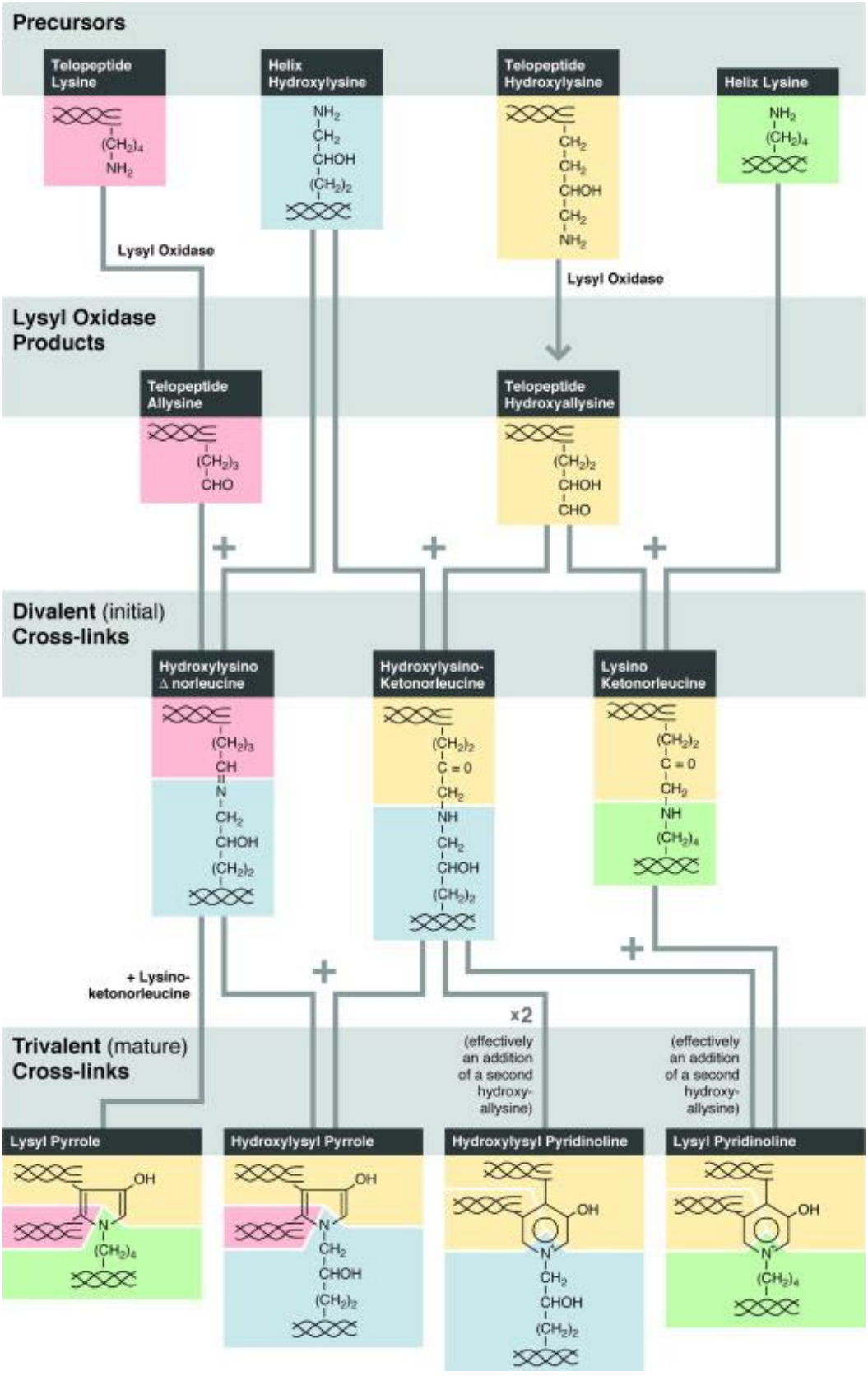


Figure 3. An overview of collagen cross-link substrates and products. Reprinted from Methods, 45, Eyre et al., Advances in collagen cross-link analysis, Pg. 68, Copyright (2008), with permission from Elsevier.

reducible trivalent cross-links.

1.6.2 Mature cross-links

The free aldehyde or aldol of an immature cross-link can further react with the ε-amino group of a hydroxylysine or histidine from a third collagen molecule to form mature trivalent or tetravalent cross-links. These mature cross-links covalently bridge three or four collagen molecules together resulting in the head to tail staggered arrangement which stabilises the collagen fibril. The trivalent cross-links are known as pyrroles and pyridinolines, and their names and structures are shown in Figure 3. The pyridinoline crosslinks are eventually released into the bloodstream after collagen tissue remodelling and subsequently excreted in the urine, where they can be quantified to assess collagen turnover in the body using HPLC or ELISA [124, 125]. An approximate 1:1 ratio of pyridinoline to pyrrole is characteristic of mature bone collagen; conversely, hydroxylysyl pyridinoline is the predominant mature cross-link found in cartilage. The ratio is ultimately determined by the extent or lack of hydroxylation of lysine residues and whether their locations in the collagen molecule is in the telopeptide or helical region of the molecule.

1.6.3 Histidine-containing cross-links

There are two further cross-links not shown in Figure 3, which include a histidine residue as part of their structure. Histidinohydroxylysinonorleucine (HHL) a trivalent cross-link found mainly in skin [118, 126-128], is formed from a telopeptide allysine, a helical hydroxyallysine from an adjacent collagen, (forming the divalent cross-link HLNL), and a histidine in the helical domain of another tropocollagen molecule [6]. Synthesis of this cross-link begins at birth, and increases rapidly initially, then gradually increases with age [97]. A decrease in HLNL cross-links and concomitant increase in HHL cross-links has been linked to the stiffness of mature skin [127-130]. Histidinohydroxymerodesmosine (HHMD) is a tetravalent cross-link formed from histidine, hydroxylysine and an aldol condensation product, or by the reaction of HHL with DHLNL [126, 127, 131]. Its existence *in vivo* is debateable; although it has been isolated and identified in several studies [126, 132-134], it is thought that it may be

formed as an artefact of base-catalysed Michael addition during sodium borohydride reduction [126, 135-141].

1.7 Fibril Orientation

After the formation of the cross-links, tropocollagen monomers self-assemble into ordered fibrillar structures. Leikina *et al* [76] found that monomers of type I collagen are actually unstable at body temperature, and that the conformation that is most preferred is that of a random coil, leading to the conclusion that the higher order structure must stabilise the triple helices. Collagen fibrillogenesis occurs through the assembly of fibril segments, called microfibrils which are in turn composed of tropocollagen molecules. These molecules of tropocollagen come together to spontaneously form microfibrils, which are ordered in an axial and parallel manner to form fibres. The fibrillar structures form a staggered arrangement consisting of five collagen molecules, overlapping to produce a gap region resulting in a characteristic D-periodicity of 67nm (the axial staggering of tropocollagen as measured by a distance that is the sum of the gap and overlap regions) [119]. The factors that direct the arrangement of individual tropocollagen monomers within the structure of the microfibril, and the composition of these microfibrils within a collagen fibril are not yet understood [142]. While there are many reports on the contribution of covalent cross-links to fibrillar conformation, there is very little relating to the contribution of other post-translational modifications such as hydroxylation, glycosylation and the presence and concentration of lysine diastereomers to the fibrillar architecture. While proteoglycans are known to play an important role in fibril organisation and size, differentiation of the individual contributions of specific collagen-associated proteoglycans and GAGs to collagen fibre architecture remains to be explored. Buehler [143], questioned the contribution of molecular properties to overall fibril structure, although their studies again mostly looked at the length of the tropocollagen molecules and the number of cross-links, with no investigation into these other molecular contributions.

Differences in fibril conformation and orientation is clearly important in determining the tensile strength of fibrils, as the skin from animals with different compositions of

corium to grain have different tensile strengths [31]. However, little is known about the molecular contributions at the collagen level that cause these differences in higher order fibril structures. Hydroxylation and glycosylation of hydroxylysine residues of collagen, and proteoglycan and glycosaminoglycan content in the extra cellular matrix (ECM) of tissues has been extensively studied. However, very little could be found in terms of how these post-translational modifications and other small collagen associated molecules might control cross-linking and thus contribute to fibril conformation in different tissues. There is also the outstanding question of whether a hydroxylysine diastereomer exists *in vivo*. If they do, exist, there is the possibility they may contribute to changes in fibril orientation through crosslinking.

1.8 Aims of the study

Animal skin structure demonstrates a complex architecture of mainly collagen fibrils, the conformation of which contributes differing properties between skin types. The collagen monomer structure has been comprehensively studied and its higher order structure has been extensively unravelled in recent years in the context of fibrillogenesis. Furthermore, many groups have focused on the contributions of covalent cross-links to fibril conformation. However, there is very little data on the influence of hydroxylation, glycosylation, proteoglycan and glycosaminoglycan interaction and possible diastereomers of hydroxylysine on this quaternary structure. It has been suggested in much earlier literature that Hyl residues that are to be glycosylated may be preferentially hydroxylated while the nascent collagen peptide is still bound to the ribosome, however no more recent studies could be found that attempt to provide a mechanism for this reaction. Questions over the possibility of a diastereomer of hydroxylysine were raised in the 1950s, yet since these initial conclusions no further research has been undertaken to confirm or negate its existence using modern methods. The answers to these currently unknown aspects of molecular interaction with collagen would be a valuable asset in our understanding of skin biology, not only in how this information could be used to help with specific pathogeneses as previously mentioned, but for numerous other applications as well, including human and animal aging, wound healing, and biomedical product development. As technology takes mankind to even greater heights collagen may find potential in the

development of more realistic prosthetics, or even robotics. These applications would all require a thorough understanding of the molecular associations which underpin collagen structure and strength.

In order to elucidate any link between lysine hydroxylation, and subsequent cross-linking and glycosylation of these hydroxylysines to cross-linking in skin, this study will analyse these modifications in the skins of four animals: cow, sheep, deer and goat. These animals have been chosen as they are common in New Zealand agricultural practice and their skins are also used to produce leather. As such, their physical properties have been extensively studied, particularly in previous work undertaken by this laboratory. In particular, a study which correlates strength and resistance in the leather produced by these animals with the collagen fibre widths and orientations has recently been completed and will be the foundation upon which this study is based. A summary of these results is found in Naffa (2017, Table 26, p. 274) [144].

This current study will examine four aspects of collagen, in the form of the following objectives:

1. To extract and purify collagens from sheep, cow, deer and goatskins, and to determine the types and relative concentrations.
2. To determine the concentration of amino acids in collagen type I (the primary collagen in skin), and the relative ratio of hydroxylysine diastereomers from each animal.
3. To quantify the types of collagen cross-links in collagen type I from each animal.
4. To determine the types and relative concentrations of glycosylated hydroxylysine in collagen type I from each animal.

These objectives will provide a comprehensive analysis of the major post-translational modifications present in the collagen type I from each of these animals, and compare them firstly to one another, and ultimately to the higher order structures of their collagens elucidated from previous studies. A comparison of all the major modifications in four different mammals to our knowledge has not been undertaken

previously, making this a novel study that should provide a significant contribution to our current understanding of collagen structure and function as a whole.

2. Materials and Methods

2.1 General materials and methods

Sheep, goat, deer and cow skin samples were obtained from the New Zealand Leather and Shoe Association Inc. (LASRA) at the raw stage, or from kind donation by David Allan (Massey). All samples for analysis were taken from the OSP (Official Sampling Position) region of the skins for consistency.

2.1.1 Sterilisation of buffers and equipment

Pipette tips and bottles were sterilised in an autoclave at 120 °C under pressure for 1 hour. Schott bottles for HPLC buffers, and glass vials for hydrolysis of collagen were cleaned with 50% nitric acid to remove the potential for protein contamination.

2.1.2 Protein assays

2.1.2.1. Bradford Assay

The modified Bradford method by Ernst & Zor [145] (originally by Bradford [146]) was followed. Briefly, samples were prepared at a concentration of 100 µg mL, then diluted 10-fold, 50-fold, and 100-fold. One hundred µL of each sample and concentration were added in triplicate to a 96-well plate. One hundred µL of 2.5-fold diluted Bradford reagent (Bio-Rad) was added to each well, and the plate was incubated at room temperature for 20 min. Absorbance was measured at 590 nm and 450 nm. A calibration curve was also prepared by reading the absorbance of collagen type I standard dilutions in the range of 5–100 µg per 0.1 ml. The concentration of the samples was then calculated from the linear equation of the collagen standard calibration curve.

2.1.2.2. Sircol Collagen Assay (SCA)

Sircol red (Sircol™ Collagen Assay, Biocolor Ltd., Northern Ireland) was used to attempt collagen-specific protein determination, and the manufacturer's protocol was followed. Briefly, 100 µL of sample was added to 1 mL of the colorimetric reagent (the dye SR in picric acid) and agitated for 45 min followed by centrifugation at 10,000 x g for 10 min. The SR dye was released from the pellet with alkali reagent (1 M NaOH) and spectrophotometric readings were taken at 540 nm on a Fluostar Optima microplate reader

(BMG Labtech, Offenburg, Germany). Absolute values were attained with a standard graph composed from collagen type I standard supplied with the kit in the range 5–100 µg per 0.1 mL.

2.1.2.3. Bicinchoninic acid assay (BCA)

Pierce™ BCA Protein Assay Kit (Pierce™, Rockford, IL, USA) was used. BCA Reagent A: sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1 M sodium hydroxide. BCA Reagent B: 4% cupric sulfate. Collagen I standard (Sigma) was prepared in a range of concentrations from 20–2000 µg mL^{-1}, samples were prepared in 10–40 fold dilutions. Working reagent was prepared by mixing 50 parts BCA Reagent A with 1 part BCA Reagent B. 25 µL duplicates of each standard and sample were pipetted into a 96-well microplate (Pierce™, Rockford, IL, USA). 200 µL of the working reagent was then added to each well and mixed thoroughly on a plate shaker for 30 seconds, then covered with foil and incubated at 37 °C for 30 minutes. The plate was then cooled to RT, and absorbance was measured at 562 nm on a plate reader.

2.1.3 SDS-PAGE

Table 3. Glycine SDS-PAGE gel components

Component	*4 %*	*5 %*	*10 %*	*15 %*
MQ Water	5.8 mL	5.3 mL		3.75 mL
40 % Polyacrylamide	1 mL	1.5 mL		3.75 mL
1.5 M Tris-HCl buffer, pH 8.8		2.5 mL	2.5 mL	2.5 mL
0.5 M Tris-HCl buffer, pH 6.8	2.5 mL			
10 % SDS	0.1 mL	0.1 mL	0.1 mL	0.1 mL
10% APS	0.1 mL	0.1 mL	0.1 mL	0.1 mL
TEMED	10 µL	5 µL	5 µL	5 µL

Glycine sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the method of Laemmli (Laemmli, 1970) with a 15, 10, or 5 % (w/v) separating gel and a 4 % (w/v) stacking gel (Table 3), and a Mini Protean III cell (Bio-Rad,

U.S.A.). SDS loading buffer (0.1 M Tris-HCl, pH 6.8, 20 % (v/v) glycerol (Vickers Laboratories Ltd), 2 % (w/v) SDS (Affymetrix), 0.02 % (w/v) bromophenol blue (Sigma) and 0.1 M dithiothreitol (DTT) (Sigma)) was added at a 1:5 buffer to sample ratio. High range Precision Plus Protein™ Dual Xtra Prestained Protein Standards (Bio-Rad, U.S.A) were used as a molecular marker ladder.

3. Collagen extraction and purification

3.1 Introduction

The usual pattern seen on SDS-PAGE gels in the literature and from comparison with a collagen type I standard is three distinctive bands; one at ~250 kDa, and two bands between 100–150 kDa. These lower weight bands typically follow a pattern where the higher weight band is ~2 x the concentration of the lower band and is consistent with a structure of two α1 chains and one α2 chain making up tropocollagen I. The higher weight band at MW ~250 kDa is consistent with a dimer of two chains held together by a crosslink and is known as β-collagen.

Objective: Optimise the extraction of collagen type I from skins in one of these three, forms or a or higher molecular weight form known as γ–collagen (three chains held together) and remove any lower molecular weight proteins such as collagen fragments from the sample.

3.2 Methods

3.2.1 Skin preparation

Skins were received raw with little to no flesh attached. All skins were taken from the OSP (official sampling position) region from each animal, as defined in Covington (1997) [31]. They were first rinsed with cold water, then the hair/wool was clipped with scissors until short (< 5 mm) before being shaved with a Wahl electric trimmer until very short (< 2 mm).

3.2.2 Reduction of cross-links

Skins were cut into 5 g samples and collagen cross-links were reduced prior to any further processing of the sample. Sodium borohydride was dissolved in 1 mM sodium hydroxide, and reduction was carried out by incubating each skin sample in the solution at a ratio of 1:30 $NaBH_4$ to skin (w/w) overnight (14 h) at 37 °C. After removal of the $NaBH_4$ solution by decanting, the skins were washed in MilliQ water until the pH was neutralised.

3.2.3 Skin dicing and pulping

Skins were diced into ~5 mm^3 pieces using a sterile scalpel blade, then homogenised using a hand-held kitchen stick blender (Kenwood Tri blade HDP306WH). In short, the ~5 g skin pieces were immersed in 50 mL of either 0.5 M acetic acid (pepsin and acid extractions) or 6 M guanidine HCl and homogenised until the size of the largest skin particles had reduced to < 2 mm^3.

3.2.4 Extraction of collagen from skin

Finely diced and pulped skins (~5 g each) were treated with 20–60 unit/mg pepsin (from porcine gastric mucosa, Sigma-Aldrich) in 0.5 M acetic acid for 12-36 hours (h) according to the methods of both Yamauchi and Shiiba, and Wang *et al* [147, 148] for the pepsin extraction. The acid extraction method was identical to this procedure except that pepsin was excluded. The salt extraction was performed with 6 M guanidine HCl in 0.05 M Tris-HCl at pH 7.0 according to the method of Yamauchi and Shiiba [147]. See Table 4 for a summary of these three methods.

Table 4. Overview of methods used for collagen extraction from animal skins.

Extraction method	*Reagent*	*Solvent*	*Concentration*	*pH*
Enzymatic	Pepsin	0.5 M Acetic acid	20-60 units/mg	~2.5
Acidic	Acetic acid	MilliQ Water	0.5 M	~2.5
Salt	Guanidine HCl	MilliQ Water	6 M	~7.0

3.2.5 Separation from skin

After extraction was complete the sample was centrifuged at 4696 x g for 1 hour at 4 °C or until the skins formed a pellet firm enough to decant any remaining liquid. Because the extract was very thick and gelatinous, several washes with additional acetic acid were sometimes required to remove as much supernatant as possible from the skin. In some cases, particularly for deer and cow skins, a pellet did not form. In such cases, the proteinaceous 'jelly' formed during extraction was removed by pressing the skins into a sieve, and washing them with small volumes of acetic acid, to push the 'jellied' supernatant through the mesh of the sieve. In some instances this washing step had to be repeated

several times over a three day course of extraction. When the extract became too thick or filled the 50 mL Nunc™ tube the supernatant was removed and set aside in a fresh tube, while the treatment was repeated on the remaining skin for the rest of the incubation. In other cases, the extract never filled the 50 mL Nunc™ tube, and so the separation from skin was only done once at the end of the 72 hours extraction.

3.2.6 NaCl precipitation

Sodium chloride (NaCl, LabServ) was made at double the final precipitant concentration so that it could be added in a 1:1 ratio of protein extract (v/v) to give the final desired concentration.

The pepsin, acid, or salt-soluble collagen was then removed by centrifugation at 4696 x g, followed by overnight precipitation at 4 °C with stirring to give a final NaCl concentration of either 0.7 M or 3 M NaCl. Precipitated material was pelleted by centrifugation (4696 x g) and once the supernatant had been removed it was washed three times with MilliQ water, before being dissolved in the minimum volume of 0.5 M acetic acid.

3.2.7 Dialysis and lyophilisation

Samples were then dialysed with Visking tubing 32 mm cellulose membrane (The Scientific Instrument Centre, London, UK) against MilliQ water for 36 hours changing water every 6 hour until the, conductivity (measured before each change) reached <10 μS. Deionised collagen extracts were then snap frozen with liquid nitrogen then lyophilised for 48 h using a VirTis Freezemobile 35EL Sentry 2.0 (SP Scientific, Stone Ridge, NY, USA)

3.2.8 Collagen re-solubilisation

Lyophilised collagen was re-solubilised in 0.5 M acetic acid for pepsin and acid extracts, and Milli-Q water for neutral extracts. Samples were then kept at -80 °C until needed.

3.2.9 Initial determination of protein concentration

The Bradford method was initially used to determine the concentration of the protein extracted from skins. However, an acceptable standard curve for collagen I standard (Sigma-Aldrich, NZ) could not be obtained even though numerous attempts were made to troubleshoot the lack of linearity. This is most likely due to the low aromatic amino acid

content of collagen, which are the amino acids with which the Bradford reagent preferentially reacts. Likewise, when a Sircol collagen assay was used, to determine the collagen content in a mixture of proteins an acceptable standard curve could not be achieved using the Sircol dye reagent either. This may be due to the fact that the Sircol kit available was for soluble collagen only, and not for use with insoluble collagen. Because the collagen for this study is extracted from mature tissue a vast quantity is expected to be covalently cross-linked, insoluble collagen, which will not react with the dye reagent in the Sircol assay. The bicinchoninic acid assay (BCA) was found to give an acceptable standard curve with an R-squared value of 0.9842, and subsequently this assay was used to determine protein concentration.

3.3 Results and discussion

3.3.1 Enzymatic extraction

Collagen can be extracted enzymatically from skin, and in some cases leather, for further purification or structural or quantitative analysis [130, 147-153]. Enzymatic extraction is often necessary to produce high yields of collagen, as the intermolecular cross-links located in the telopeptide regions result in low acid-solubility of the intact collagen molecule [154]. Pepsin, trypsin or various other proteases are commonly used in these extractions. Pepsin preferentially cleaves the polypeptide chain between hydrophobic and aromatic residues. As collagen type I contains no aromatic residues in the helical region, pepsin is expected to cleave collagen only in the telopeptide regions of the molecule, producing collagen triple helices without telopeptides, and without damage to the triple-helical region of the collagen [150]. Because pepsin is only active at low pH, pepsin extraction will produce a mixture of acid-soluble and pepsin-soluble collagen.

Trypsin cleaves polypeptides on the C-terminal side of lysine and arginine residues which are present within the helical region of the collagen molecule. It is commonly used if samples are to be analysed by mass spectrometry, as it imparts a favourable charge to the resulting peptides. As the first objective was to purify the whole collagen molecule in various forms, trypsin was not used at this stage.

Initial extractions were carried out by incubating shaved, diced sheep and cow skins at both 4 °C and 25 °C, over 24 hours, using the method of Yamauchi and Shiiba [147]. Although the optimum temperature for pepsin activity is 37-42° C, higher temperatures were not used because of concern about microbial spoilage occurring at 37° C or higher. While the use of protease inhibitors has been shown to prevent microbial growth they could not be used due to the risk of inhibiting the pepsin itself. Following collagen extraction with pepsin the liquid containing the samples was raised to pH 7.0 and NaCl added to a final concentration of 3.0 M and incubated overnight. The precipitated protein was pelleted by centrifugation, and the supernatant containing the soluble collagen desalted and analysed by SDS polyacrylamide gel electrophoresis (SDS PAGE). The results can be seen in Figure 4 and Figure 5.

This extraction produced a clear, distinct three-band collagen pattern as seen in Figure 4, and subsequent 3 M NaCl precipitation (while clearly reducing the concentration of these bands) removed most other presumably non-collagenous proteins. Determination of the

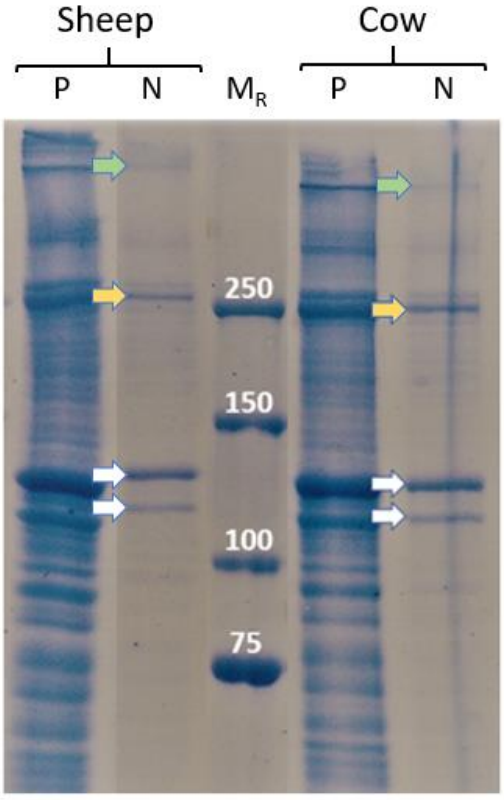


Figure 4. 5 % SDS-PAGE gels after the first pepsin digestion at 25° C (P), followed by 3 M NaCl precipitation at 4° C (N) of collagen from sheep and cow skins. White arrows indicate the positions of the α1 and α2 chains of collagen type I, yellow arrows represent β form of collagen (two chain dimer), and the green arrows denote the probable γ form of collagen (three chain or more association). All samples are a 10-fold dilution of the protein extract. Gel was stained with colloidal Coomassie stain overnight.

concentration of protein using the Bradford method was unsuccessful at this stage for reasons alluded to above.

A second round of pepsin extraction was carried out on the pellet to determine if any collagen remained in the skin. A comparison of the first and second pepsin extractions is shown in Figure 5. It is immediately obvious that the 10-fold dilutions from the first and second pepsin treatments look very similar. in contrast to the 10-fold diluted samples after NaCl precipitation (Figure 4 N lanes), which show bands with intensities similar to the 100-fold dilution lanes in Figure 5. To check that there are indeed differences in concentration, an attempt was made to use the Sircol (collagen-specific) assay using a collagen I standard. Because an acceptable standard curve could not be obtained a comparison of the weight of lyophilised product after purification compared with the wet skin weight it came from was used to give an indication of collagen yield. Table 5 shows a breakdown of the yields of each species after each pepsin extraction and purification.

Table 5. Yields of collagen following first and second pepsin digestions.

Skin	*Wet weight.*	*P1 dry weight.*	*% Dry/wet*	*P2 dry weight.*	*% Dry/wet*
Sheep	1.05 g	80.8 mg	7.69 %	38.7 mg	3.68 %
Cow	1.03 g	91.5 mg	8.88 %	46.3 mg	4.49 %

NB. P1 is first pepsin digestion and P2 is second digestion. Wet wt. is weight of raw skin prior to pepsin treatment, dry wt. is weight of sample following further purification and lyophilisation.

The gel in Figure 5 also shows differences in the patterns for the pepsin digest seen at higher molecular weight, specifically those representing the putative collagen β-chains at around 250 kDa. Compared to the distinct pattern seen clearly in Figure 4 in which the doublet in the P lanes has a more concentrated lower band, the 10-fold concentrated sample in Figure 5 has a lower concentration of these β-chain bands. with the top band now having greater density than the bottom band. While this may be due to greater loss of the bottom band during the process, either during NaCl precipitation or through dialysis it seems unlikely. Because the molecular cut-off of the dialysis tubing is not known, there is a possibility that some tropocollagen long helical chains may be able to pass through the tubing in an end-

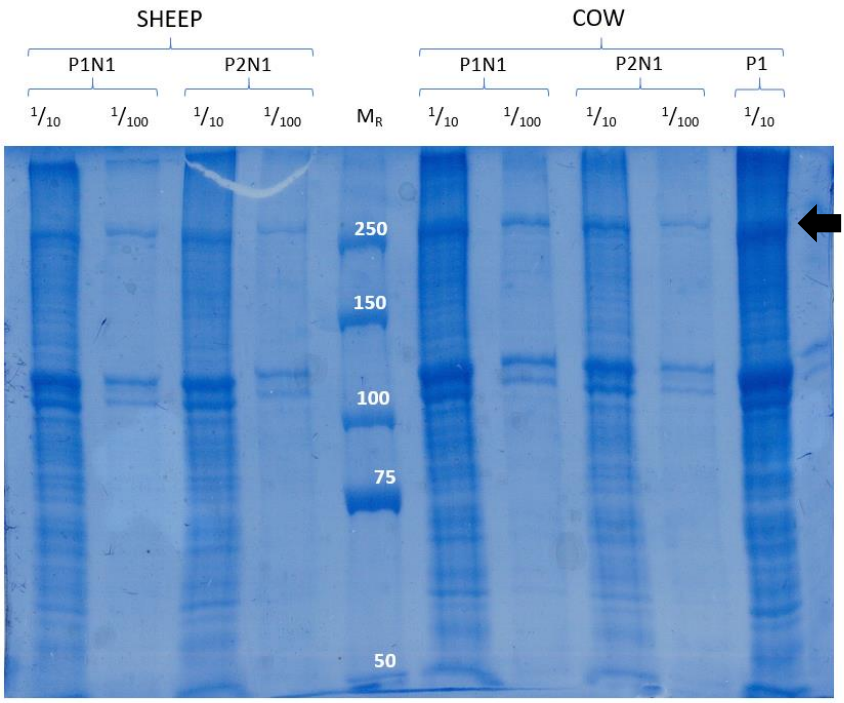


Figure 5. 5 % SDS-PAGE gels after second pepsin digestion (P), followed by NaCl precipitation (N) of collagen from sheep and cow skins. P1N1 is one pepsin extraction followed by one NaCl precipitation. P2N1 is the second pepsin treatment followed by one NaCl precipitation. P1 is the first pepsin extraction without precipitation. Fractions above lanes denote dilutions of sample prior to loading. Black arrows denote β-collagen strand positions.

on orientation. Presumably this would be more likely to be an α-chain, as two chains bound together by a cross-link (β-chain form collagen) should provide enough girth to prevent elution. This could be the reason why the lower MW β-chain band is attenuated following purification. On the other hand, there may be real differences between the two β-collagen forms. For example, one form may be held together by an intramolecular cross-link between an α1 chain and an α2 chain, and the other between two α1 chains, or there may be differences in the quaternary association between the chains that is affected differently during the purification procedure. If the cross-links holding these two different structures together are between different intramolecular regions of their respective peptides, the quaternary structures could vary rendering the molecules less or more susceptible to SDS unfolding. Furthermore, some cross-links, such as aldimine cross-links that are formed between telopeptide allysines and Lys/Hyl residues in helical regions are less stable than other types of cross-links in the acidic conditions [133] used in these preparations.

3.3.2 Pepsin inactivation

The observation that higher weight bands appeared to be attenuated over time prompted curiosity about whether the pepsin was still actively degrading the collagen molecules during precipitation and dialysis. There is no mention for the need to specifically inactivate pepsin

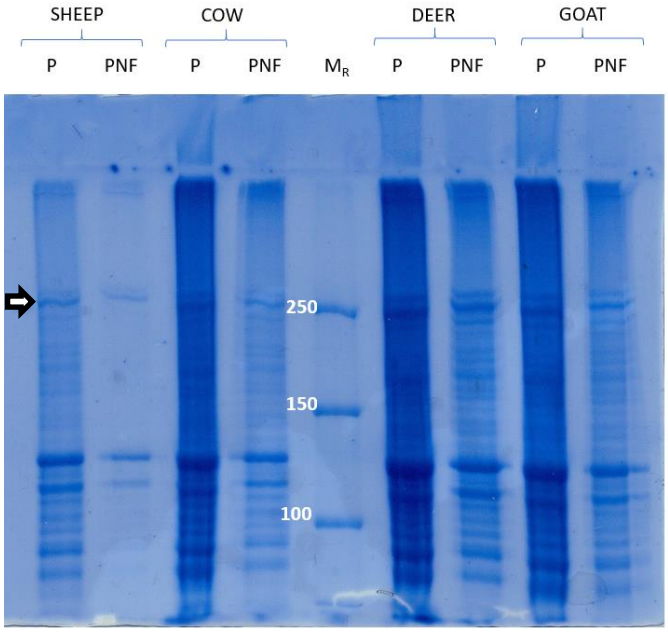


Figure 6. Pepsin extraction of collagen at 25° C followed by pH inactivation of pepsin at pH 9.0, with a subsequent decrease to pH 7.0. Precipitation in 3 M NaCl was carried out in 10 µM sodium phosphate buffer, pH 7.4. The supernatant was dialysed against MilliQ water to remove excess salt before analysis by 5 % SDS-PAGE. P is after pepsin extraction, PNF is after precipitation, dialysis and lyophilisation. All samples were diluted 10-fold in AcOH prior to loading. White arrow shows putative β-collagen position.

prior to salt precipitation in the literature. However, Piper and Fenton's report on pepsin [155] shows that although the enzyme is most active from pH 2.0–4.0, and inactive at neutral pH, it is not irreversibly inactivated until pH 8.0–8.5. Therefore, if a sample at neutral pH is acidified again, any pepsin remaining in solution will most likely be reactivated.

The pepsin used in these experiments was from porcine gastric mucosa (Sigma Aldrich, NZ) in lyophilized powder form, with an activity of ≥2,500 units/mg. At a 1:100 enzyme to skin ratio (w/w), each gram of skin is treated with 25 x 10^3 units of enzyme, or 25 units/mg skin, as used by Zeugolis *et al* [156]. This relatively low level of enzyme was used to restrict cleavage to the telopeptide regions in order to free the collagen helical chains.

In the next experiment pepsin was inactivated by increasing the pH following collagen extraction from fresh sheep, cow, deer, and goat skin samples. A second change involved NaCl precipitation. This time sodium phosphate buffer, pH 7.4 was used, as monitoring of pH during the previous two extractions showed that the pH had dropped to ~4.05.0 during the precipitation steps. This may have contributed to re-activation of pepsin and subsequent further degradation to the collagen bands, particularly the β- and γ-forms. Furthermore, tests of the tap distilled water used for dialysis showed that this water had a pH of 4.0–5.0

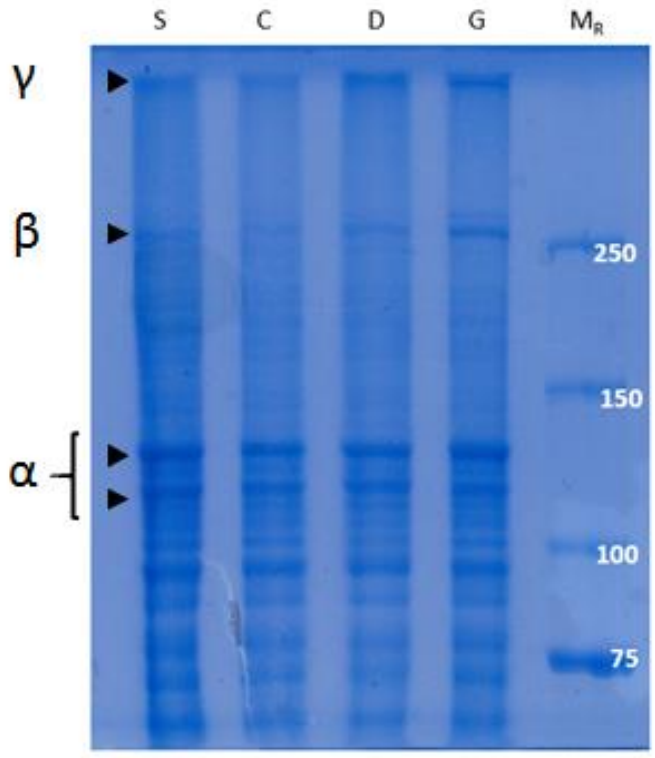


Figure 7. 5 % SDS-PAGE gels of collagen from four different species skins after pepsin digestion and 3 M NaCl precipitation, dialysis and lyophilisation. S is sheep skin, C is cow skin, D is deer skin, and G is goat skin. Arrows denote proposed α, β, and γ forms of collagen as indicated. Samples were diluted 10-fold in AcOH prior to loading.

rather than 6.5–7.0. Additionally, pepsin was inactivated by increasing the pH to 9.0 for 30 mins before bringing it back to neutral with 3 M HCl. The results of the pH-controlled extraction are shown in Figure 6. While the collagen bands extracted from cow, deer, and goat skins looked similar, those extracted from sheep skin appeared to be present in much lower concentration, particularly the β-collagen. In addition, the yield of lyophilised product compared to the wet skin weight used were very different. The calculated yield for all samples was between 35 %, similar to the results obtained for the second pepsin extraction of experiment 1, rather than the ~10 % yields of collagen achieved without pH adjustment. This low yield led to a return to the original method.

The next experiment was carried out using an increased quantity of starting material (~2 g) of sheep, goat, cow, and deer skins and the original conditions with milliQ water. The resultant gel (Figure 7) shows that below 150 kDa, goat, cow, and deer skins have bands at similar positions with similar intensities. The lane containing the sheep skin extraction appears to have a very slightly higher concentration of these bands, when compared to the intensity of the strong band common to all samples at just below 100 kDa that has a similar

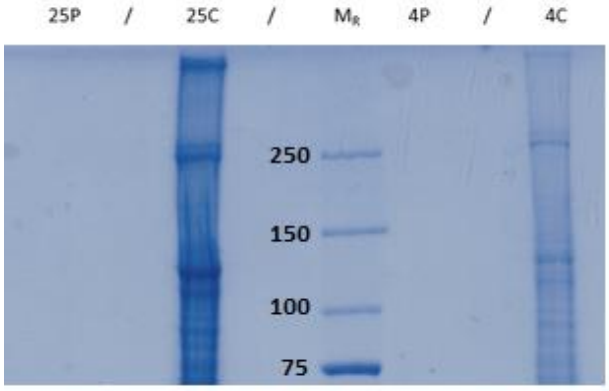


Figure 8. 5 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and 4° C (25C and 4C, respectively). 25C and 4C Samples were diluted 10-fold in acetic acid prior to loading. 25P and 4P denotes acetic acid containing pepsin only and no skin samples, to test pepsin viability after 24 h incubation at the respective temperatures.

intensity in all samples. There are, however, some clear differences in the banding patterns between the different animals at higher MW; the β-collagen band just above 250 kDa is more intense in collagen extracted from goat skin than in the other skins implying that it is present in higher concentrations. Collagen extracted from deer skin has a markedly less intense band at this position while in collagen extracted from sheep and cow skins, the band is barely visible. Collagen from sheep skin has a band approximately halfway between the 150 and 250 kDa markers, which is not apparent in any of the other samples, and deer and goat skins also seem to have a greater amount of γ-collagen sitting at the top of the resolving gel. These differences may be due to differences in the chemical nature of the cross-links as well as their position in the molecule which may account for why some β- and γ-collagens from different species are less likely to break down when digested with pepsin. If cross-links are positioned between two or more helical regions, rather than only between telopeptides, then pepsin will not be able to remove the α-strand from the molecule if the cross-link remains intact, resulting in very stable β and γ forms.

Next, the effects of the pepsin hydrolysis temperature and the NaCl concentration were optimised. The temperature at which the extraction of pepsin-soluble collagen is done is consistently recorded as 4° C in the literature. This is presumably to prevent the possibility of the skin beginning to rot due to the accumulation of microbes, as the optimum temperature for pepsin activity is ~37–42° C. Hydrolysis was repeated at 4 °C and the results compared to those obtained at 25° C (Figure 8). Controls containing pepsin only, in 0.5 M acetic acid were included at each temperature to determine whether pepsin autolysed after 24 h at either temperature (Figure 9B).

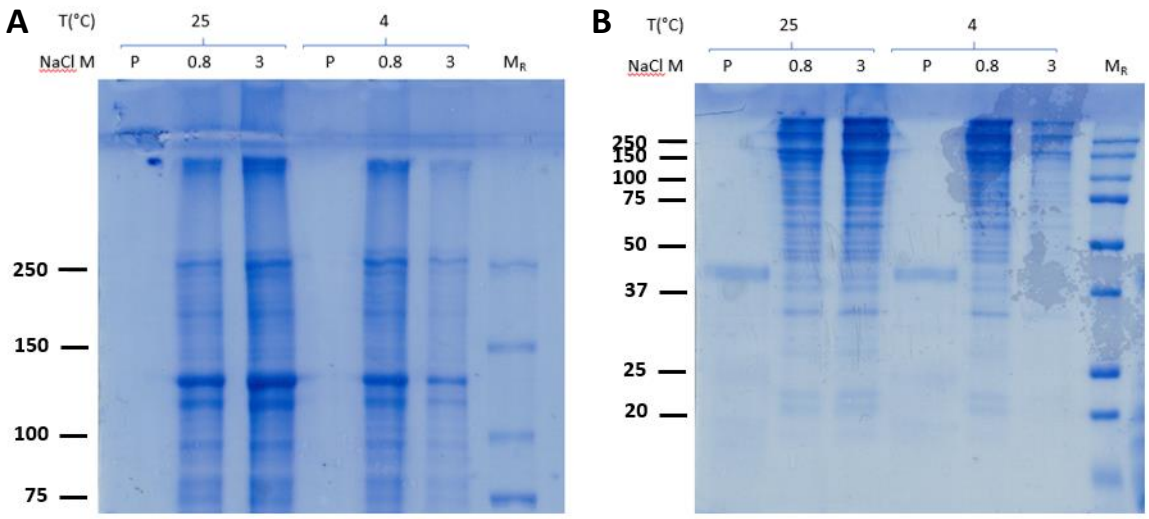


Figure 9. Comparison between 25°C and 4°C pepsin incubation temperatures, and 0.8 M NaCl and 3 M NaCl collagen precipitation on (a) 5 % and (b) 15 % SDS-PAGE gels. P is pepsin only lane.

Figure 8 shows that the amount of collagen extracted at 25 °C over 24 h, is significantly more than the amount extracted at 4 °C and there was no sign of spoilage. These results confirmed that 25° C was the optimum temperature to extract soluble collagen from skin over a 24 period. Following pepsin extraction, both samples were precipitated using two different concentrations of NaCl, 0.8 M and 3 M (Figure 9).

Interestingly, each temperature seems to produce more collagen at different NaCl concentrations. The 25 °C extraction produced a higher concentration of collagen at 3 M NaCl, while the 4 °C extraction produced a greater concentration at 0.8 M NaCl (Figure 9). This may be due to differences in the type of collagen extracted at the different temperatures, as collagen I is known to precipitate at ~0.8 M NaCl, whereas types III and V precipitate at higher concentrations [157]. However, there is little difference between the two 0.8 M treatments at either extraction temperature. The extra bands which can be seen at ~250 kDa in the 4 °C and 0.8 M NaCl extractions, but not in the 25 °C and 0.8 M NaCl extractions, are also present in the 3 M NaCl precipitation of material extracted at 25 °C. Inspection of the gels in Figure 9 indicate that using 3 M NaCl to precipitate collagen extracted at 25 °C results in the highest concentration of all collagens and consequently collagen I for downstream analyses.

3.3.3 Boiling *vs* Non-boiling of samples prior to SDS-PAGE

Of concern following all of the previous gels, was the continued presence of what appeared to be either breakdown products of the collagen itself, or non-collagenous proteins which had not been removed during purification. These bands were not seen in gels produced by other groups using the same purification protocols. All previous samples had been boiled for ~5 minutes in SDS-reducing loading buffer prior to gel analysis to ensure the proteins were completely denatured. The possibility that the collagen was somehow being broken down in the boiling process was considered. The gels in Figure 10 show the result of running the same samples from Figure 9, this time with both boiled and non-boiled samples. The difference between these samples is immediately apparent. While boiled samples showed the typical breakdown of bands below the α-chain positions (between 150-100 kDa), the unboiled samples showed virtually no contaminating proteins below the collagen chains suggesting that the apparent 'contaminating bands' are either breakdown

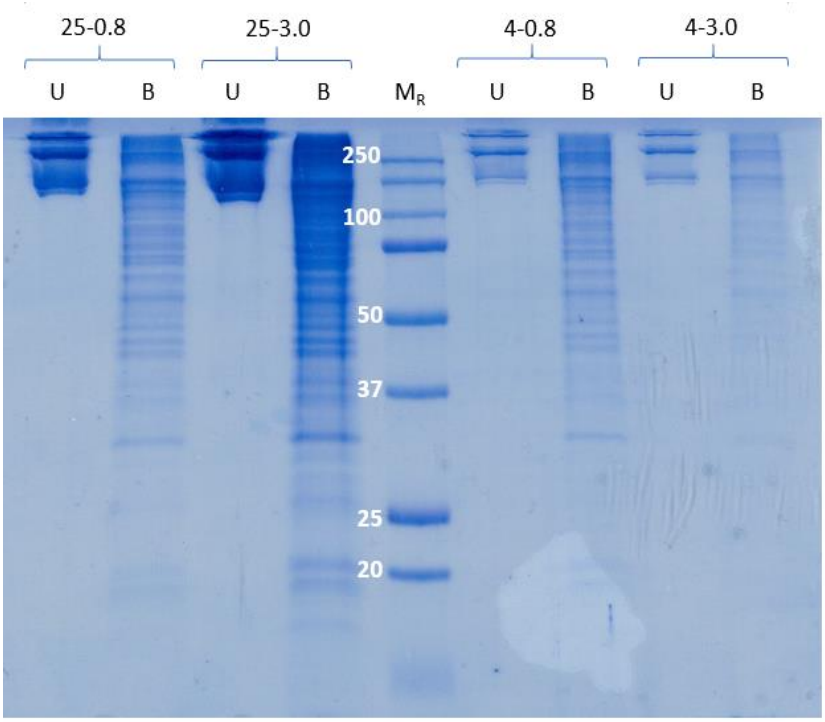


Figure 10. 15 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and 4° C and precipitation with 0.8 and 3 M NaCl. Samples were diluted 10-fold in 0.5 M acetic acid prior to loading. U is unboiled samples and B is boiling for 5 mins prior to loading.

products of the collagen itself, or smaller proteins which are tightly associated with the collagens until removal by boiling.

Further trials to test the effects of temperature on hydrolysis to determine the extent that different heating times and temperatures affected the samples were carried out. Figure 11 shows the results of a comparison of unboiled samples with boiling at 2 or 4 mins, or heating to 80 °C in a temperature-controlled heating block for 1 hour. Compared to the unboiled samples, those that had been boiled for 2, and 4 minutes showed an increase in lower molecular weight bands with time. While the increase in intensity of the lower bands is concomitant with a loss of intensity in the β- and to a lesser extent the γ-bands there doesn't appear to be a great deal of difference between the α-chain bands. These results suggests that the β and γ bands are degraded by boiling although this seems to be rather counterintuitive.

These gels were retained to attempt excision of the bands which appear upon boiling for identification by mass spectrometry using proteomic methods (see future work). Due to the

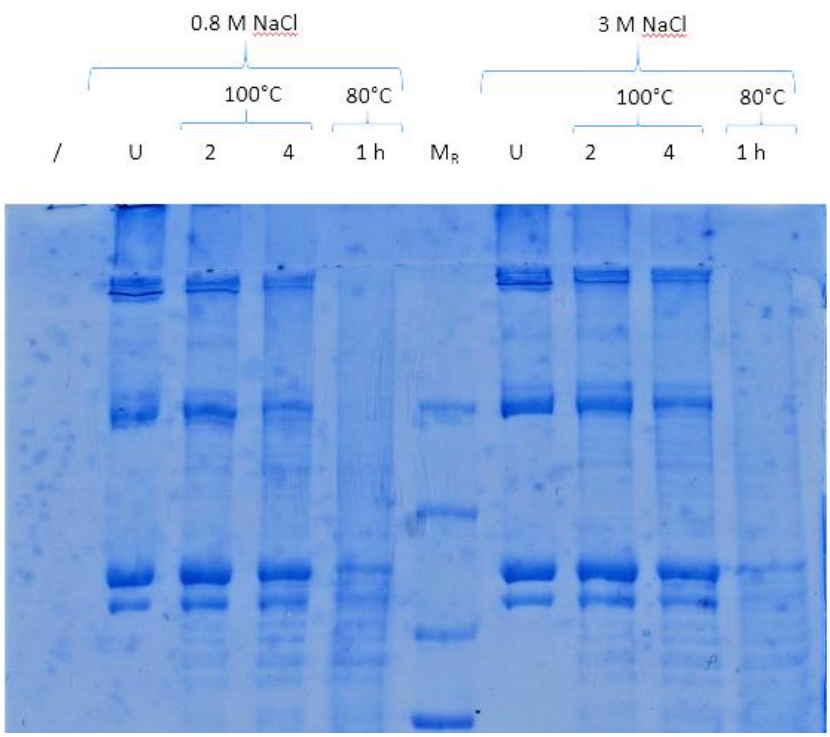


Figure 11. 5 % SDS-PAGE gels of collagen from cow skin after pepsin digestion at 25° C and precipitation with 0.8 and 3 M NaCl. Samples were diluted 10-fold in 0.5 M acetic acid prior to loading. U is unboiled samples. 2 and 4 are with boiling for 2 and 4 mins respectively prior to loading. 1 h samples were heated to 80 °C in a heating block for 1 h prior to loading.

fact that these bands only appeared upon boiling in sample buffer, the decision was made to load gels without boiling or heating of the sample in sample buffer, as the purpose of the gels is to ensure that the purification method has been successful.

3.3.4 Pre-treatment *vs* straight to pepsin

In order to increase the yield of collagen, many other groups included a pre-treatment of skins with NaOH to remove non-collagenous proteins as well as with alcohol to defat the skins prior to pepsin extraction of collagen. To test the validity of this claim, the following experiments were carried out. The method of Duan *et al* [158], involving immersion of 1 g of cow skin in 10 mL 0.1 M NaOH (w/v) with shaking for 24 h (changing the solution every 8 h), followed by immersion in 10 mL of 10 % butyl alcohol with shaking for 12 h (changing the solution every 6 h). Figure 12 shows the results of pre-treatment *vs* straight to pepsin extraction of collagen from cow skins.

As seen in Figure 12, very little difference can be discerned between the two extraction methods, with similar band intensities for each collagen form resulting from both methods.

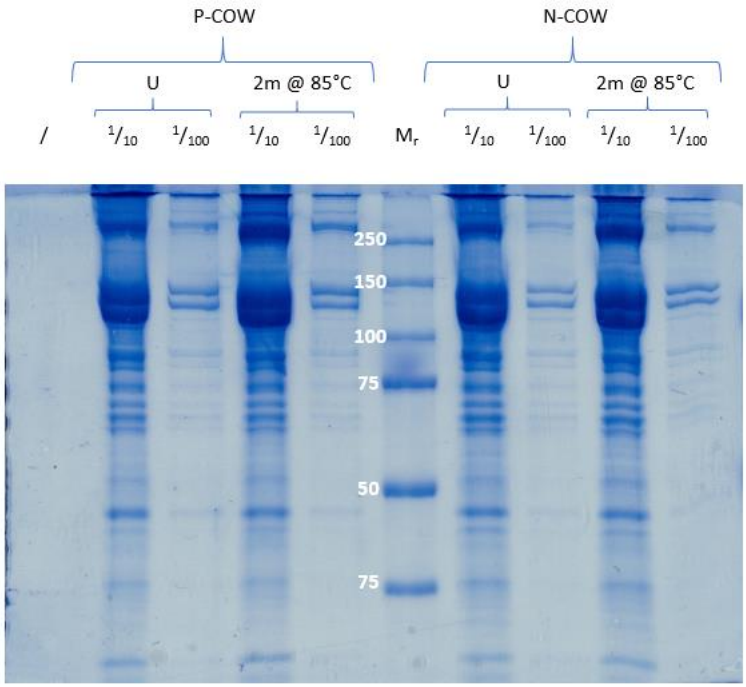


Figure 12. 15 % SDS-PAGE gels comparing "Straight to pepsin extraction" (P-COW) *vs* 0.1 M NaOH treatment of skin followed by defatting with 10 % Butyl Alcohol before pepsin extraction (N-COW). U is unboiled prior to loading, compared to heating for 2 mins at 85 °C. Fractions above lanes denote dilution of extractions in 0.5 M acetic acid.

Yields determined by lyophilised collagen/dry skin weight also showed little difference for pre-treated pepsin extracted skins (N-COW) compared to untreated pepsin extractions (P-COW); P-COW producing a yield of 9.73 % and N-COW a yield of 9.61 %. Because there was less than 0.5 % difference between the two yields, pre-treatment of skins was rejected for the simplicity and brevity of the 'straight to pepsin' method.

3.3.5 Acidic and neutral extractions to retain telopeptide regions

As enzymatic extraction results in tropocollagen lacking the C- and N- telopeptides, collagen was extracted using acidic and neutral extraction methods. Both of these retain the telopeptide regions which are necessary to study intermolecular crosslinks that involve one or more telopeptide Hyl/Lys residues. Acidic extraction was carried out using the same conditions as those used for pepsin extraction but without the addition of pepsin, and with the addition of cOmplete™ protease inhibitor + EDTA (Sigma-Aldrich) according to the manufacturer's instructions. One g each of cow and sheep skins were used for each extraction. The neutral (salt) extraction was carried out using a modified version of the method used by Yamauchi and Shiiba [52] (explained in detail in section 3.2). Figure 13 shows the results of all three extraction methods, in both sheep and cow skins. Comparison

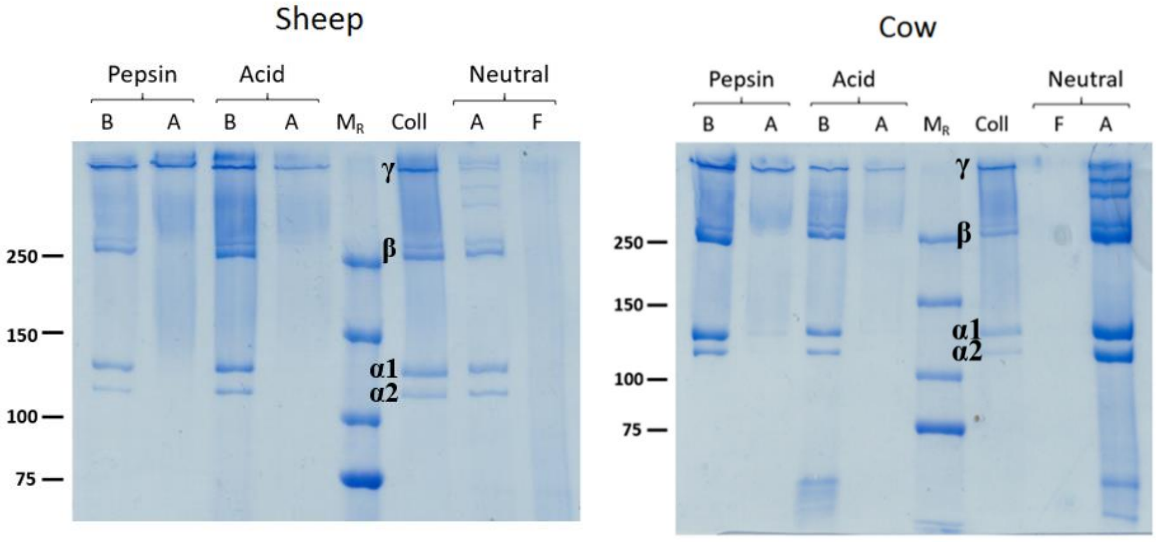


Figure 13. 5 % SDS-PAGE gels showing pepsin, acid and neutral extractions for both sheep and cow skin. In pepsin and acid extractions, B is before NaCl precipitation, A is after precipitation and before lyophilisation. For neutral extractions, A is after concentration and F is the flow through. α-, β-, and γ- collagens are labelled on collagen I standard lane (Coll).

of extractions with the collagen I standard on the gels shows little difference between the pepsin and acid extractions and the standard, whereas the neutral extraction of both sheep and cow collagen shows two extra distinct bands between the position of the β- and γ-collagen bands of the standard. These bands may correspond to β-strand collagen retaining the telopeptide regions, which may be lost in the acidic extraction for some as yet unknown reason. If acid hydrolysis is the cause of epimerisation of hydroxylysines as proposed by Hamilton and Anderson, and supported by Witkop [108, 112], according to Hudson's rule as described by Witkop, the bonds formed upon hyl1 residues may be more susceptible to breakage under other forms of stress, such as those afforded by the rigorous extraction methods involved in the extraction of collagen from skin. This could result in the differences seen in the β-bands seen in Figure 13 for neutral extractions with salt. As the salt extraction does not involve any acid, fewer centrifugations and fewer steps overall due to the lack of need for further precipitation, this "gentler" extraction may also preserve associations with other proteins or molecules, whether they be other collagen type I molecules, other accessory collagen types, other proteins altogether or associations with large proteoglycan molecules. In order to elucidate these structures, future work will involve the excision of these bands and their analysis with mass spectrometry to identify them.

3.3.6 Final extraction method

Using the results of these experiments, a modified method for extraction of the pepsin- and acid-soluble collagen samples for future analysis was determined. This method is summarised in Figure 14 and Figure 15 and explained in detail in section 3.2. These methods, along with an adaptation used by Yamauchi and Shiiba to extract collagen containing the telopeptides (also detailed in section 3.2), were then used to extract collagen from skin samples extracted under three different conditions. . The pepsin extraction method (PE) was used to extract pepsin-soluble collagen that did not contain the telopeptide regions. The acid extraction method (AE) was used to extract collagen that contained the telopeptide regions, and the guanidine extraction method (GE) was used to extract collagen containing telopeptides without exposing the collagen to acid. This "no acid" method is required to study any possible effect acid exposure may have on collagen PTMs, such as hydroxylysine diastereomer ratio and cross-link concentrations.

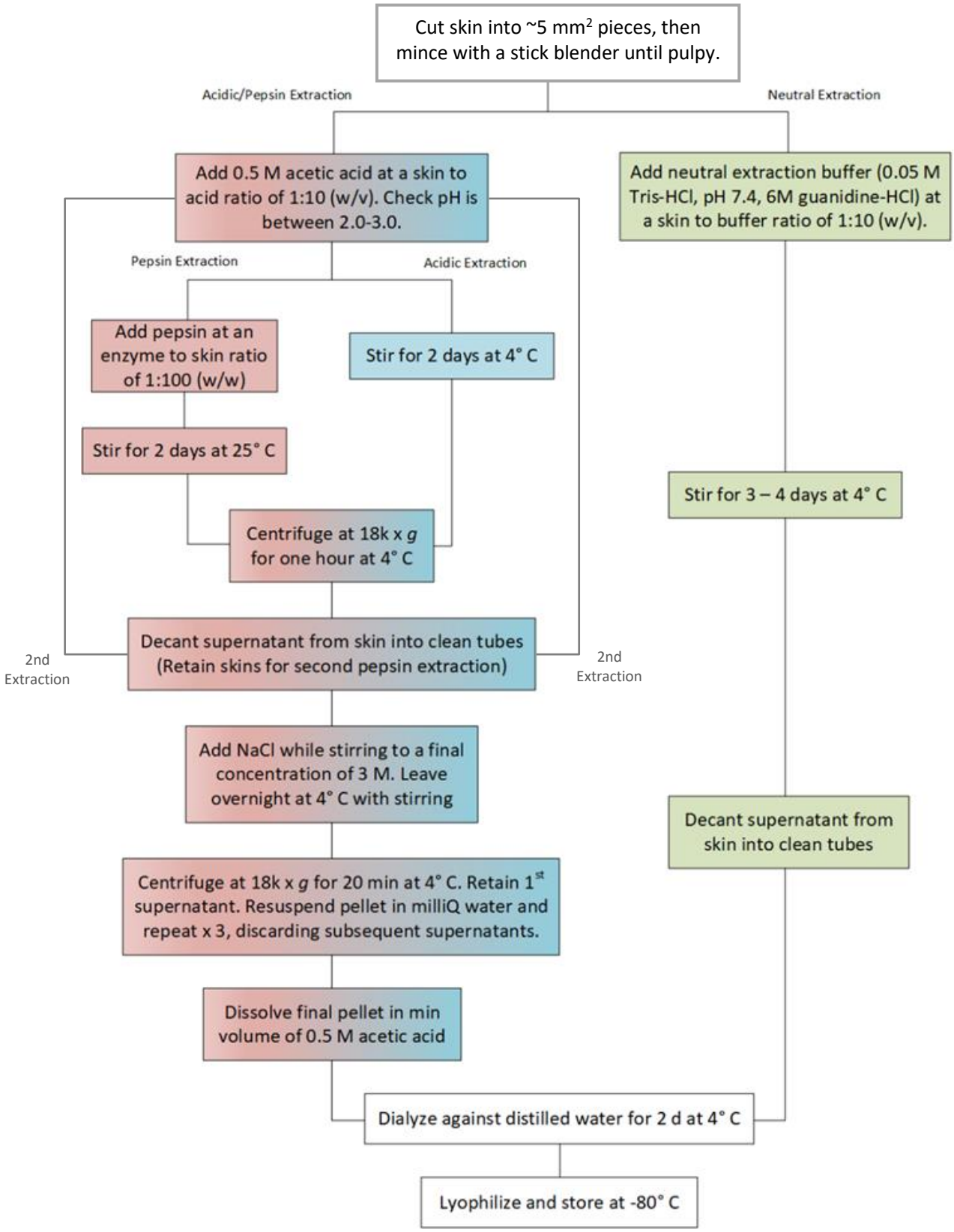
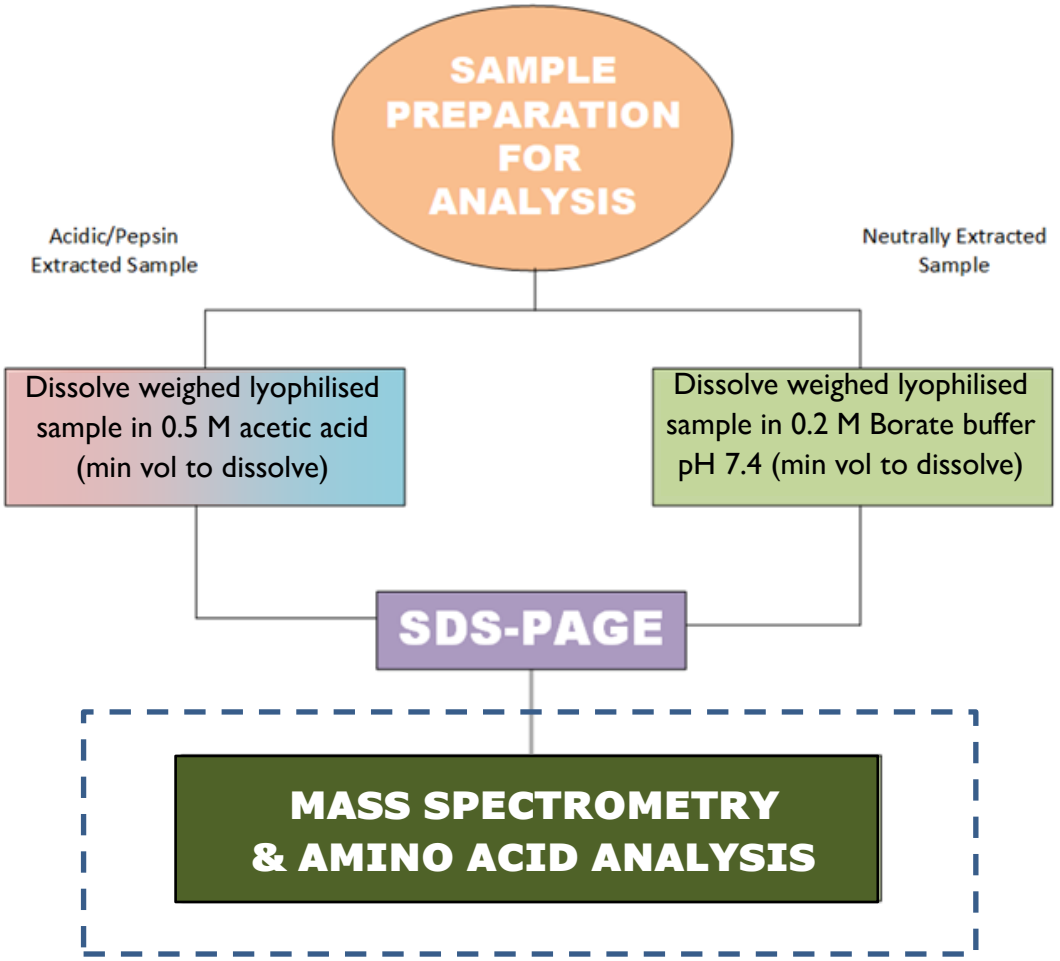


Figure 14. Extraction and purification of collagen from raw skin using pepsin, acidic, and neutral extraction methods, after optimisation of each stage of extraction and purification.



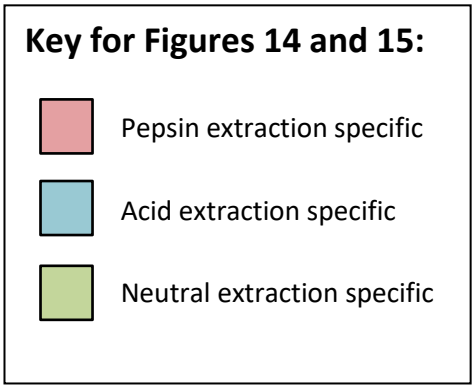


Figure 15. Method for preparation of pepsin, acidic, and neutral extractions of collagen for analysis on SDS-PAGE gels and subsequent future molecular analysis.

3.3.7 Putative collagen yield and purity

After extraction, purification and lyophilisation as per the schemes in Figure 14, samples were weighed using an analytical balance and the dry weights were divided by the skin sample weight to give a putative collagen yield (Table 6 and Figure 16). On average, the collagen yield from non-reduced skins was higher than that of the reduced skins by approximately 18-58 %. This is most likely due to the reduction of the cross-links causing the higher order structures of collagen to be retained, while the non-reduced forms were easily broken during the extraction process, resulting in a higher yield of soluble collagen. Both cow and deer collagen skins seem to be particularly affected by reduction, with cow skins showing the greatest difference between pepsin-extracted, reduced and non-reduced yields of collagen, at a ratio of 0.46 (reduced/non-reduced). The neutral non-reduced extractions from cow and deer skins also produced very different yields from their reduced counterparts, with ratios of 0.77 and 0.76, respectively. Overall it was the salt extraction which appeared to produce the greatest difference between the yields from reduced and non-reduced skins. The average ratio of reduced/non-reduced yields of collagen across all species when salt-extracted was 0.58, whereas collagen extracted with pepsin reduced/non-reduced ratio was 0.79.

Of all species and extractions, neutrally extracted, non-reduced cow skins produced the greatest average yield of collagen, at 16.45 %. Interestingly, although non-reduced cow and deer skins both produced the greatest yields using the pepsin and neutral extractions respectively, reduced cow skins produced the lowest yield from pepsin extraction at only 4.17 %, and reduced cow and deer both produced the lowest yields during neutral extraction, at 3.65 % and 3.19 % respectively.

While these results do not give accurate quantitation of collagen yield, as the lyophilised sample may contain trace amounts of proteins other than collagen, or collagens other than collagen type I, given the SDS-PAGE profile of these extractions showed only the typical collagen type I banding profile, with no other visible bands characteristic of smaller proteins, or other collagen types, it is likely that the sample is mainly collagen I. This does not, however, exclude contamination by other factors, such as salt, lipids or carbohydrates.

The purity will be more accurately determined through quantitation of collagen by the hydroxyproline concentration, determined by amino acid analysis, (section 4).

Table 6. Yields of collagen extracted from biological replicates of cow, sheep, deer, and goatskins. Skin weight is given as dry weight of the skin sample.

Code	*Extraction*	*Species*	*Reduction*	*Tech Rep #*	*Skin wt (g)*	*Collagen wt (g)*	*Yield (%)*
			Pepsin Extraction				
P-CN1	Pepsin	Cow	Non-reduced	1	7.69	0.4718	6.14
P-CN2	Pepsin	Cow	Non-reduced	2	6.67	0.4662	6.99
P-CN3	Pepsin	Cow	Non-reduced	3	5.10	0.5266	10.33
P-CR1	Pepsin	Cow	Reduced	1	5.44	0.0953	1.75
P-CR2	Pepsin	Cow	Reduced	2	5.76	0.3289	5.71
P-CR3	Pepsin	Cow	Reduced	3	6.36	0.3212	5.05
P-SN1	Pepsin	Sheep	Non-reduced	1	5.01	0.3396	6.78
P-SN2	Pepsin	Sheep	Non-reduced	2	5.04	0.2917	5.79
P-SN3	Pepsin	Sheep	Non-reduced	3	5.05	0.3729	7.38
P-SR1	Pepsin	Sheep	Reduced	1	5.05	0.2288	4.53
P-SR2	Pepsin	Sheep	Reduced	2	5.01	0.3076	6.14
P-SR3	Pepsin	Sheep	Reduced	3	5.05	0.2060	4.08
P-DN1	Pepsin	Deer	Non-reduced	1	6.89	1.0065	14.61
P-DN2	Pepsin	Deer	Non-reduced	2	6.44	0.8438	13.10
P-DN3	Pepsin	Deer	Non-reduced	3	7.11	0.7196	10.12
P-DR1	Pepsin	Deer	Reduced	1	5.52	0.4311	7.81
P-DR2	Pepsin	Deer	Reduced	2	5.59	0.3024	5.41
P-DR3	Pepsin	Deer	Reduced	3	5.63	0.1971	3.50
P-GN1	Pepsin	Goat	Non-reduced	1	5.01	0.2221	4.43
P-GN2	Pepsin	Goat	Non-reduced	2	5.05	0.2909	5.76
P-GN3	Pepsin	Goat	Non-reduced	3	5.02	0.1313	2.62
P-GR1	Pepsin	Goat	Reduced	1	5.01	0.4969	9.92
P-GR2	Pepsin	Goat	Reduced	2	4.99	0.2621	5.25
P-GR3	Pepsin	Goat	Reduced	3	5.01	0.2572	5.13
			Acid Extraction				
A-CN1	AcOH	Cow	Non-reduced	1	5.31	0.6647	12.52
A-CN2	AcOH	Cow	Non-reduced	2	5.18	0.7219	13.94
A-CN3	AcOH	Cow	Non-reduced	3	5.03	0.7053	14.02
A-SN1	AcOH	Sheep	Non-reduced	1	5.07	0.4658	9.19
A-SN2	AcOH	Sheep	Non-reduced	2	5.03	0.4034	8.02
A-SN3	AcOH	Sheep	Non-reduced	3	5.00	0.4080	8.16
A-DN1	AcOH	Deer	Non-reduced	1	7.54	0.1841	2.44
A-DN2	AcOH	Deer	Non-reduced	2	7.04	0.1386	1.97

Table 6. Continued...

Code	*Extraction*	*Species*	*Reduction*	*Tech Rep #*	*Skin wt (g)*	*Collagen wt (g)*	*Yield (%)*
A-DN3	AcOH	Deer	Non-reduced	3	7.47	0.4823	6.46
A-GN1	AcOH	Goat	Non-reduced	1	6.14	0.3056	4.98
A-GN2	AcOH	Goat	Non-reduced	2	5.99	0.2562	4.28
A-GN3	AcOH	Goat	Non-reduced	3	6.04	0.1743	2.89
			Guanidine Extraction				
G-CN1	Guanidine	Cow	Non-reduced	1	11.17	1.8897	16.92
G-CN2	Guanidine	Cow	Non-reduced	2	11.02	1.8498	16.79
G-CN3	Guanidine	Cow	Non-reduced	3	11.81	1.8489	15.66
G-CR1	Guanidine	Cow	Reduced	1	12.17	0.4298	3.53
G-CR2	Guanidine	Cow	Reduced	2	11.36	0.4429	3.90
G-CR3	Guanidine	Cow	Reduced	3	11.93	0.4208	3.53
G-SN1	Guanidine	Sheep	Non-reduced	1	9.31	1.2085	12.98
G-SN2	Guanidine	Sheep	Non-reduced	2	9.21	1.2338	13.40
G-SN3	Guanidine	Sheep	Non-reduced	3	9.06	1.2145	13.41
G-SR1	Guanidine	Sheep	Reduced	1	11.93	0.8709	7.30
G-SR2	Guanidine	Sheep	Reduced	2	11.36	1.0815	9.52
G-SR3	Guanidine	Sheep	Reduced	3	9.39	0.6368	6.78
G-DN1	Guanidine	Deer	Non-reduced	1	8.61	1.1713	13.60
G-DN2	Guanidine	Deer	Non-reduced	2	8.62	1.1852	13.75
G-DN3	Guanidine	Deer	Non-reduced	3	9.34	1.2426	13.30
G-DR1	Guanidine	Deer	Reduced	1	12.15	0.3264	2.69
G-DR2	Guanidine	Deer	Reduced	2	13.58	0.4325	3.18
G-DR3	Guanidine	Deer	Reduced	3	12.64	0.4665	3.69
G-GN1	Guanidine	Goat	Non-reduced	1	9.22	0.8873	9.62
G-GN2	Guanidine	Goat	Non-reduced	2	8.17	1.6839	20.61
G-GN3	Guanidine	Goat	Non-reduced	3	8.42	1.0393	12.34
G-GR1	Guanidine	Goat	Reduced	1	10.38	1.3606	13.11
G-GR2	Guanidine	Goat	Reduced	2	10.11	0.6911	6.84
G-GR3	Guanidine	Goat	Reduced	3	11.17	0.9154	8.20

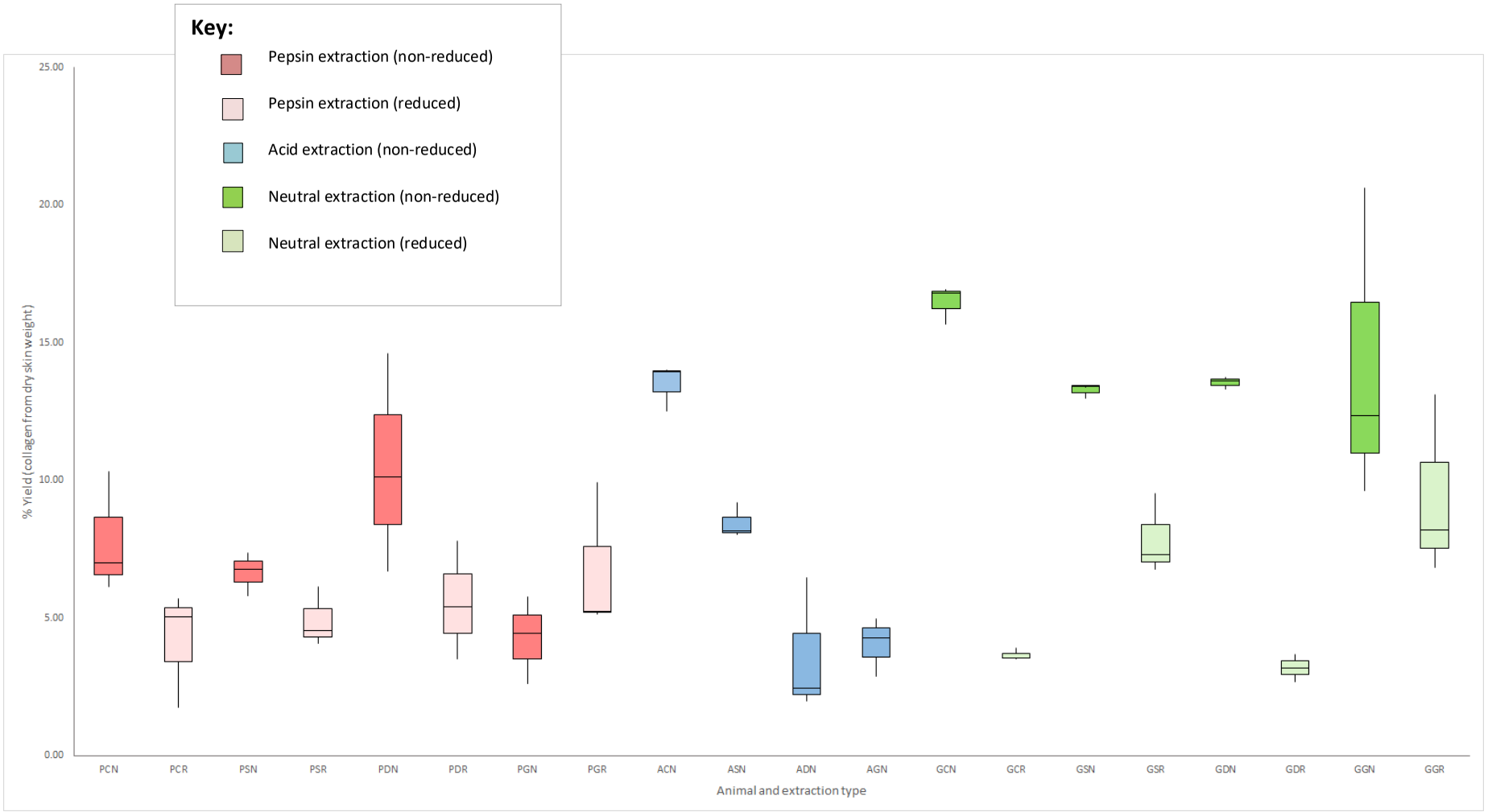


Figure 16. Yields of collagen from dry skin weight across extraction types and species. Pink represents pepsin extraction, blue is acid extraction, and green is neutral (salt) extraction. Pale versions of each colour represent samples reduced with $NaBH_4$ prior to extraction.

3.4 Conclusion

Methods of collagen extraction reported in literature vary widely, but there is a general consensus towards pre-treatment of skin involving removal of non-collagenous proteins with NaOH followed by defatting with alcohol, and subsequent extraction of collagen by acid followed by NaCl precipitation, dialysis and lyophilisation. While this method does produce purified collagen that displays SDS-PAGE bands consistent with the α- β- and γ-bands of collagen type I, this study found little to no difference between this standard method and simplified versions. Most notably, very little difference was found between the standard method involving pre-treatments of the skin prior to extraction, and no pre-treatment. Given that pre-treatment of skins involves a considerable increase in time and complication of method, and that it actually appears to lower the yield of collagen extracted, it may be more beneficial to omit this step.

In addition, conducting extractions at 4 °C, pH neutralisation following acid extraction, and boiling of samples prior to loading on SDS-PAGE gels all appear to be additional steps which are commonly used by other researchers studying collagen, but which seem from this study to result in a decreased yield of collagen. Boiling of samples prior to SDS-PAGE in particular appears to break down the higher structures of collagen. The reasons for this observation are not obvious and identification of the lower molecular weight bands is essential before any conclusions can be made.

The differences in the yields of collagen from the different animal skins and the different extraction methods is difficult to explain. The fact that the replicates are relatively consistent suggests this is not due to experimental inconsistencies, but rather is a property of the skins themselves and their molecular structure.

4. Amino acid analysis

4.1 Introduction

Amino acid analysis is a technique that can be used to accurately determine the concentration of peptides and proteins. It can also be used to detect atypical amino acids that might be present in a protein or peptide, and to evaluate fragmentation strategies for peptide mapping. Proteins are classified as large molecules that generally exist as folded structures with specific conformations, while peptides may consist of only a few amino acids. Because proteins and peptides are macromolecules which consist of amino acid residues covalently linked to form a linear polymer, they require hydrolysis into their individual amino acid constituents before amino acid analysis. Once hydrolysed, the resulting "free" amino acids of the test sample are typically derivatised for analysis by binding 'detectable' molecules to the amino acids to molecules which allows their identification after chromatographic separation. Current techniques typically use reversed phase high-pressure liquid chromatography (RPHPLC), to separate fluorescently derivatised amino acids. A detector then transforms the analogue signal from the detector to an output for analysis. High purity reagents are necessary to prevent background contamination and buffers are filtered through a 2 μ filter before use to remove any particulate matter and prevent possible microbial or environmental contamination.

Calibration of the instrumentation prior to amino acid analysis is typically achieved through the creation of a standard curve using a commercially available amino acid standard. These consist of an aqueous mixture of amino acids appropriate for comparison to the test sample. For example, while an amino acid standard containing only the common 20 amino acids may be appropriate for some proteins, collagen also contains the unique amino acids hydroxyproline and hydroxylysine, and as such the amino acid standard for calibration also needs to contain these residues. Once prepared and derivitised, the standard is analysed at a range of concentrations (typically a minimum of 4–6 different concentrations) to obtain a spectrum of peaks which correspond to each of the derivatised amino acids. The retention and chemistry factors of the separation are then adjusted by varying the solvent (including the ion-pairing agent, pH and buffer composition) and temperature, until individual peaks are well resolved.

The area of each amino acid peak is plotted against the known molar concentration in the standard and used to determine a response factor for each amino acid in the standard (average peak area or peak height per nanomole of amino acid present in the standard). This response factor is then used to calculate the concentration of each amino acid present in the test sample, when the test sample is run under identical conditions to the amino acid standard. This is achieved by dividing the peak area recorded for a given amino acid in the test sample by the response factor for that amino acid in the standard. If many test samples are being analysed at one time the standard calibration is routinely run between sets of test samples to ensure that the response factor is still accurate.

Collagen concentration is commonly determined using reversed-phase high pressure liquid chromatography (RP-HPLC) of hydrolysed whole skin, or collagen extracts. RP-HPLC utilises a hydrophobic stationary phase to separate derivatised amino acids according to their hydrophobicity. Thus the more hydrophobic amino will have a greater affinity for the stationary phase, in this case C_{18}, and be retained on the column longer. The initial mobile phase is polar (aqueous), resulting in the more hydrophilic amino acids eluting from the column first. Hydrophobic molecules are then eluted by increasing the hydrophobicity of the mobile phase by the addition of an organic (non-polar) solvent, reducing hydrophobic interactions with the stationary phase. As the gradient of the organic solvent increases, the mobile phase binds more strongly to the stationary phase, displacing the more hydrophobic amino acids that are eluted last.

4.2 Methods

Amino acid concentrations of the collagen extractions obtained in Aim 1 were determined using 6-aminoquinolyl-n-hydroxysuccinimidyl carbamate (AQC) labelling (Cohen & Michaud, 1993; Molnár-Perl, 2005). Labelled amino acids were separated by RPHPLC and quantitated using a standard curve. All reagents used were HPLC and/or analytical grade as required and sourced from Sigma-Aldrich (NZ). For all protocols three technical replicates of each biological collagen I sample extracted from each type of animal were used.

4.2.1 Hydrolysis of collagen

Following extraction and purification, lyophilised collagen extracts were hydrolysed under either acid conditions (pepsin and acid extracted collagen), or by using enzymes for salt-extracted collagen, to yield free amino acids. 100 mg each of pepsin, acid and salt-extracted, lyophilised collagen were hydrolysed in Kimax® 13 x 100 mm phenolic screw-cap borosilicate glass culture tubes (Kimble Chase, Rockwood, TN, USA) using 5 mL of 6 M analytical grade HCl, (LabServ, NZ) containing 3 % (w/v) phenol for 24 hours at 115 °C. Acid hydrolysates were then filtered through glass wool to remove particles and lyophilised. Lyophilised samples were dissolved in 1 mL Milli-Q water and stored at -80 °C until derivatisation.

Salt-extracted collagen was sequentially enzymatically hydrolysed using actinidin, (a kind gift from Dr M Boland, Riddet Institute, Massey University) followed by proteinase K, (Sigma-Aldrich, NZ) then carboxypeptidase Y (Sigma-Aldrich, NZ). 20 µg of each salt-extracted collagen sample was first treated with actinidin dissolved in 0.05 M borate buffer pH 8.4, at a 1:10 (w/w) ratio of enzyme to collagen and incubated overnight at 25 °C (14 hours). Proteinase K was then added at a ratio of 1:40 (w/w) enzyme to collagen and incubated overnight at 55 °C. 0.5 M acetic acid was added to reduce pH to 6.5, and carboxypeptidase y was then added at a ratio of 1:50 (w/w) enzyme to collagen and incubated at 37 °C overnight. Samples were then centrifuged at 16,000 g for 5 min (Heraeus Biofuge Pico centrifuge, Kendro Laboratory Products, Germany) and the supernatant was retained and stored at -80 °C until needed.

4.2.2 AQC derivatisation

10 µL of each hydrolysate sample or 10 µL of amino acid standard for collagen hydrolysate (2.5 nmol μL^{-1} of each amino acid except for cysteine (1.25 nmol μL^{-1}), hydroxylysine (12.5 nmol μL^{-1}), and hydroxyproline (12.5 nmol μL^{-1})), (Sigma-Aldrich), were added to 60 µL 0.2 M borate buffer pH 8.85 followed by 10 µL 100 pmol μL^{-1} norleucine, and 20 µL 15 pmol μL^{-1} 6-Aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) (SYNCHEM), mixed thoroughly by vortexing then heated to 55 °C for 10 mins to complete the reaction. AQC rapidly labels primary and secondary amines to form highly fluorescent asymmetric derivatives that are readily amenable to analysis by liquid chromatography Figure 17. Excess AQC in the mixture is rapidly ($t_{½}$<15 s) hydrolysed to 6-aminoquinoline N-

hydroxysuccinimide and carbon dioxide. AQC-derivitised amino acids were then subjected to RP-HPLC for analysis.

4.2.3 High performance liquid chromatography (HPLC) and amino acid analysis (AAA)

Collagen concentration was determined through quantification of hydroxyproline using HPLC and fluorometric detection, as previously described [159].0.5-8 µL of amino acid standard and 0.5–1 µL of each sample were injected onto a column (Phenomenex 150 x 4.6 mm 5 µ or 100 x 2 mm Gemini C18 High Performance 3 µ). The mobile phase was made up from 50 mM ammonium acetate pH 5.05 and 60 % acetonitrile and was optimised for the best resolution between hydroxlysine1 (Hyl1) and hydroxlysine2 (Hyl2). Eluted peaks were monitored using a fluorescence detector (Dionex RF 2000) with the excitation wavelength set at 245 nm and emission wavelength at 395 nm. Amino acid concentrations were determined from calibration curves calculated using Dionex CHROMELION

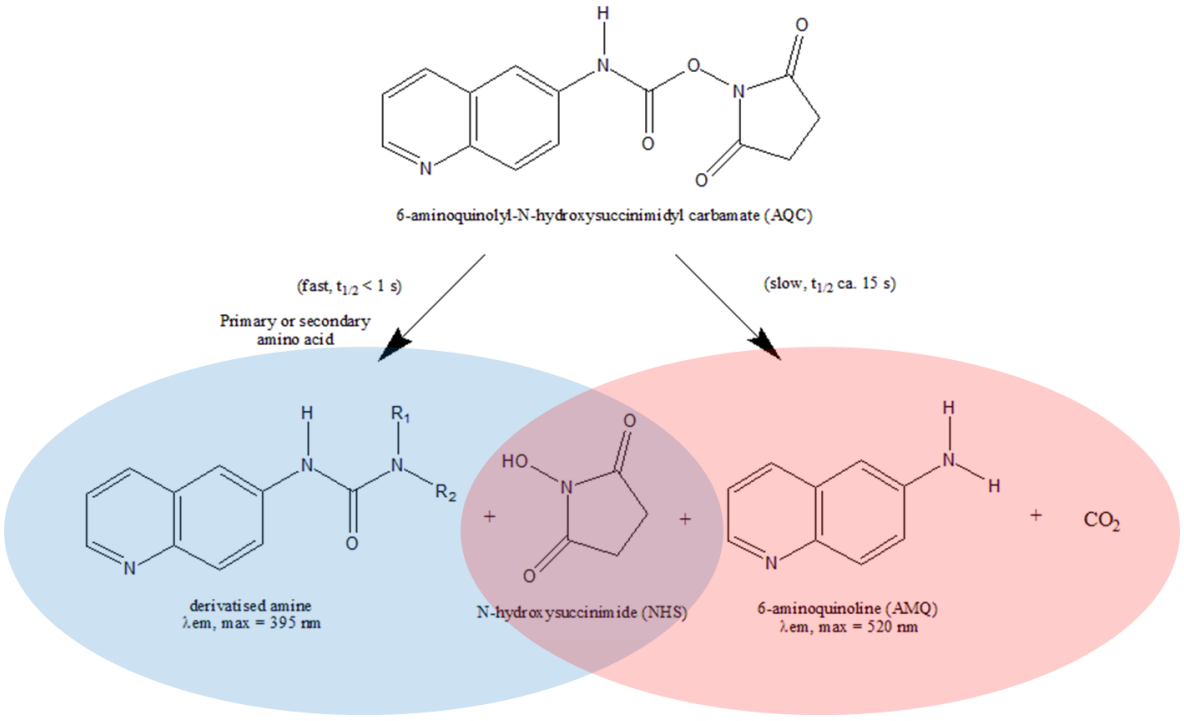


Figure 17. Amino acids and amines react with AQC to produce highly stable, fluorescent derivatives. The excess reagent reacts with water to form a free amine.

version 6.80 SR13 Build 3967. Hyl concentrations were expressed as the total number of residues per collagen molecule, assuming 300 Hyp residues per triple helix.

4.3 Results and discussion

4.3.1 Optimisation of the amino acid elution profile

Amino acid analysis of hydrolysed collagen extracts allows determination of the concentration of the unique amino acid hydroxyproline (Hyp). Because Hyp is only found in collagen, it is often used to estimate the amount of collagen present in a sample based on the understanding that it makes up 13.4 % (w/w) of collagen in mammals [88, 160]. Yields of collagen after each extraction method (pepsin, acidic and neutral) were determined for each animal and compared to determine any differences between collagen type and concentration. The identity of all of the amino acids in the extracted collagen type I samples of each animal were determined, including hydroxyproline and both diastereomers of hydroxylysine as they eluted from an HPLC column (Phenomenex® 150 by 4.6 mm Gemini C18 column). 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate

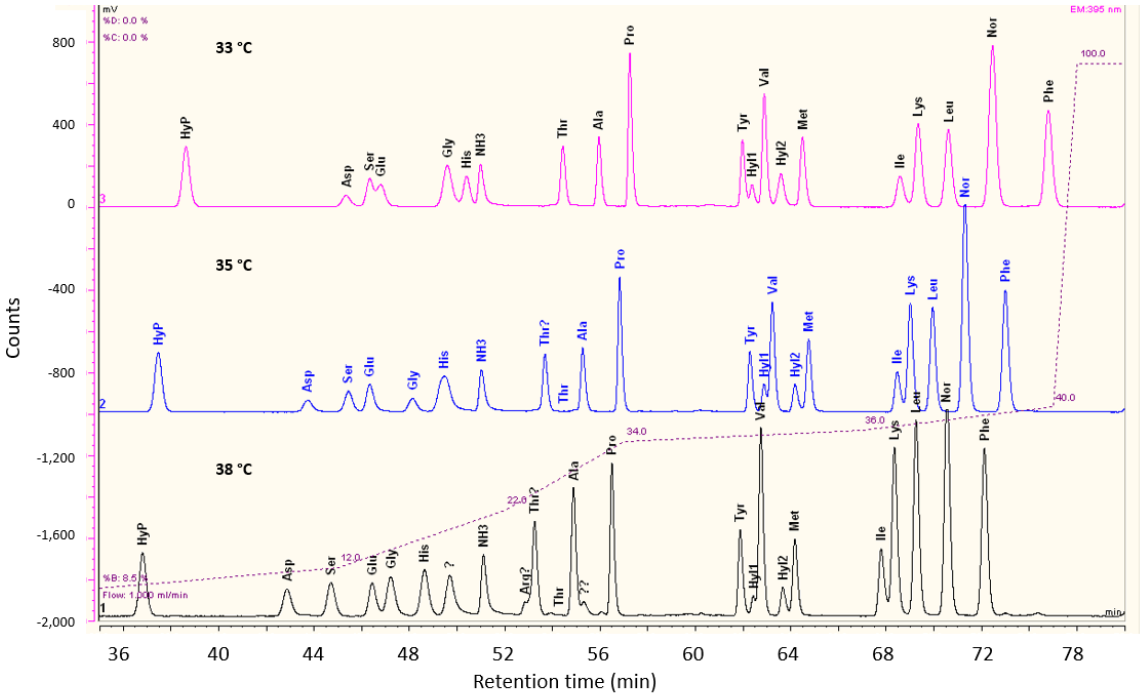


Figure 18. Effect of temperature change on elution profile of amino acids from C18 column. Mobile phase consisted of a gradient of 5 mM ammonium acetate pH 5.05, and 60% acetonitrile.

(AQC) was chosen as a pre-column derivative, due to its ease of use and excellent stability; derivatisation is a simple, one step procedure that requires no further clean up prior to analysis and the derivatised molecules remain extremely stable with no discernible loss detected after storage at room temperature for one week.

A standard containing amino acids commonly found in collagen was derivatised with AQC, then separated by HPLC to determine the retention time for each residue. Using an elution scheme previously established in this laboratory, good resolution of most of the amino acids was achieved using 5 mM ammonium acetate, pH 5.05 as the more hydrophilic component of the mobile phase. However, when the temperature was adjusted to improve resolution between Glu and Ser, and between Hyl1 and Tyr the resolution of the polar amino acids improved at the expense of the non-polar residues. This caused greater crowding in the second half of the chromatograph, particularly in the Y-V-M region, resulting in poorer resolution of Hyl diastereomers (Figure 18).

As ammonium acetate buffer produced an acceptable separation of the amino acids for quantitation, the decision was made to optimise the resolution by adjusting the gradient profile of the mobile phase. However, a change in instrumentation (Dionex Ultimate 3400 HPLC system with a FLD-3400RS Rapid Separation Fluorescence Detector and CHROMELION version 7.1), caused peaks that had previously been resolved using the same mobile and stationary phases, to merge. Various changes were attempted to regain resolution including a change to a shorter, narrower column with a smaller particle size (Phenomenex® 100 by 2 mm Gemini C18), and changes to the mobile phase and column temperature, all of which had little effect. Next the AQC was investigated as it is known to degrade slowly upon repeated exposure to atmospheric water, a possibility as a stock solution was used to prepare derivatives. On exposure to moisture, AQC undergoes minor hydrolysis into its by-products AMQ, CO_2 and *N*-hydroxysuccinimide (NHS), and as the concentration of AMQ builds up, it slowly reacts with another molecule of AQC, to form *bis*-aminoquinoline urea [161] that can form a precipitate over time due to its poor solubility. Another possibility was that the original AQC, which is purchased in crystal form and reconstituted in acetonitrile prior to use, was not fully dissolved when the reagent was prepared. To eliminate these possibilities, fresh AQC crystals were dissolved into

acetonitrile (MeCN) and briefly sonicated (< 5 mins) [161]. However, they were again difficult to dissolve, and even with sonication and subsequent gentle heating at 30 °C could not be fully dispersed into the MeCN. Further investigation using chemspider (www.chemspider.com) stated that dimethyl sulfoxide (DMSO) is the appropriate solvent for AQC. A new stock was prepared by reconstituting AQC into DMSO, which rapidly dissolved all crystals. The DMSO-AQC was then used to derivatise the standard-amino acid mix and a chromatogram of this standard is shown in Figure 19.

While the DMSO-AQC preparation appeared to give sharper peaks with shorter retention times, several additional peaks were detected which had not been present in the MeCN-AQC preparations. This raised concern about the possible presence of artefacts, either already present in the DMSO or formed by an unknown reaction between the amino acid standard and DMSO. While other groups have used DMSO as a solvent for similar precolumn derivatisation experiments [162], most use MeCN, following the original method from Cohen [163]. In order to eliminate the possibility that the crystallised AQC

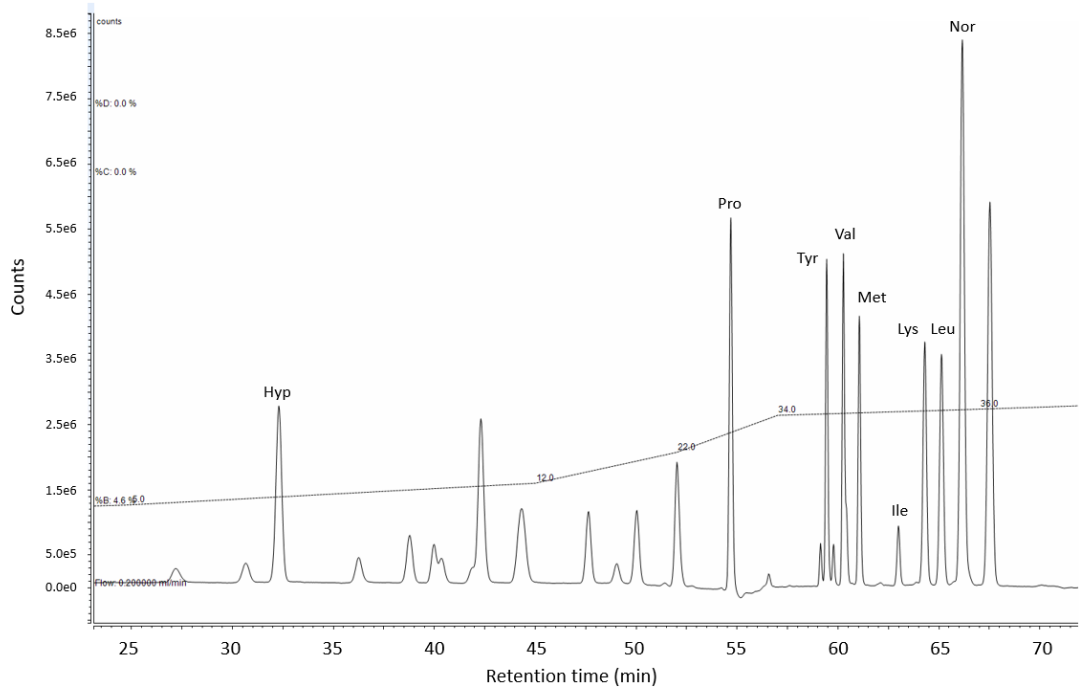


Figure 19. HPLC chromatogram of the separation of amino acid standards derivitised with AQC dissolved in DMSO. Buffer A was 5 mM ammonium acetate pH 5.05 and Buffer B was 60% acetonitrile and the total run time was 120 mins. All amino acids were eluted from the column after 70 mins.

used may have taken on water thus making it difficult to dissolve in MeCN, a fresh vial of reagent was opened and reconstituted again in MeCN. This time it completely dissolved in the solvent. Analysis of amino acid standards derivatised with this new batch of reagent still showed complete loss of peak resolution.

Next, both buffers A and B were tested to ensure that these were not causing the loss of resolution. The buffer lines to the HPLC system were flushed with 50% isopropanol (IPA) (J.T. Baker Chemical, Phillipsburg, NJ), to remove possible microbial contamination that continued to be seen in the buffer A solution after a period of 2+ days. After flushing of the buffer lines and preparation of new buffer A and B the standard was run again, this time showing some improvement to the chromatograph, however, after ~2 days the spectrum worsened, and particles could again be seen in the bottom of buffer A.

As the ammonium acetate used to prepare the buffer was several years old and appeared to have taken on water, and there was the possibility that this could be the source of the microbial contamination. Buffer A was instead prepared by mixing ammonia (28-30%, Merck, Germany) with glacial acetic acid to a final concentration of 5 mM and adjusting the

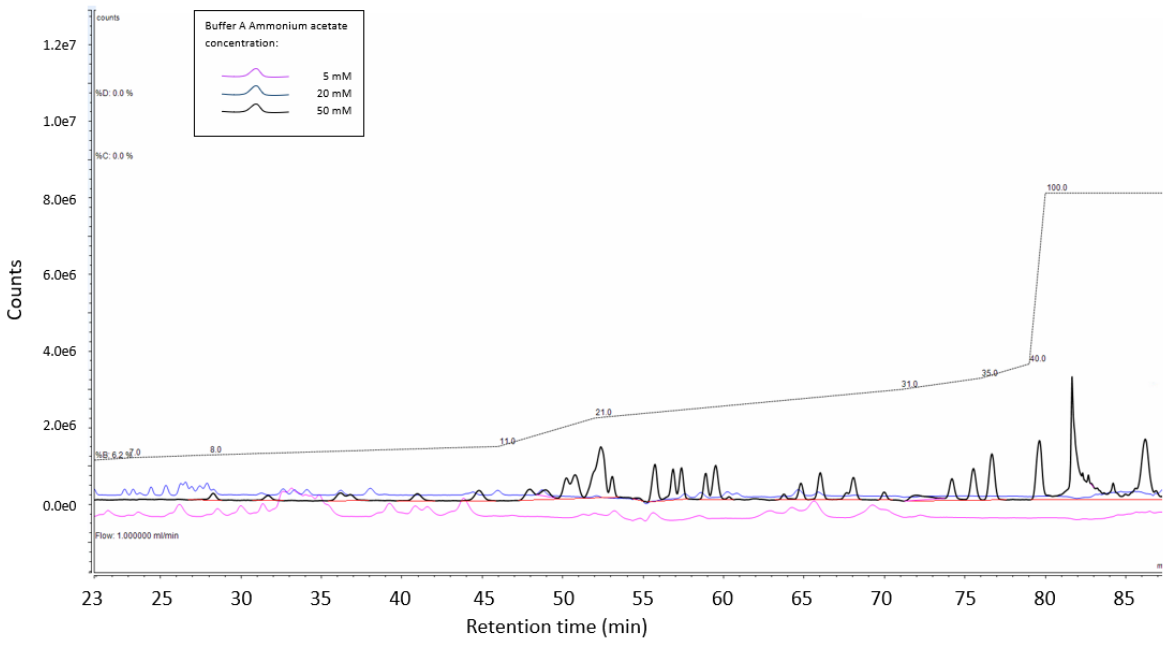


Figure 20. Comparison of the chromatographic separation of amino acid standards with 5 mM, 20 mM and 50 mM buffer A solutions.

pH to 5.05 by adding microlitre volumes of glacial acetic acid. This buffer produced a similar spectrum to the crystallised ammonium acetate preparation, and after 5 + days showed no visible signs of contamination in the bottle and the spectrum did not worsen.

A search of the literature showed that a some groups had achieved successful amino acid analysis using concentrations of ammonium acetate as high as 0.1 M therefore an increase in the concentration of ammonium acetate in buffer was trialled [164, 165]. The chromatographic separations using 5 mM, 20 mM and 50 mM ammonium acetate buffers are compared in Figure 20. When 50 mM ammonium acetate was used, fluorescent detection, appeared to be more sensitive, and background noise appeared to be less compared with the 20 mM ammonium acetate. While the elution profile of the amino acids did not look typical at this stage, many of the peaks were grouped in a similar pattern to the typical profile as seen in Figure 18. The column was equilibrated with the new 50 mM buffer A for 2 hours and then the elution was run 6 times with 20 min washes in 60% MeCN followed by equilibration for 10 mins between every second run. The elution profile

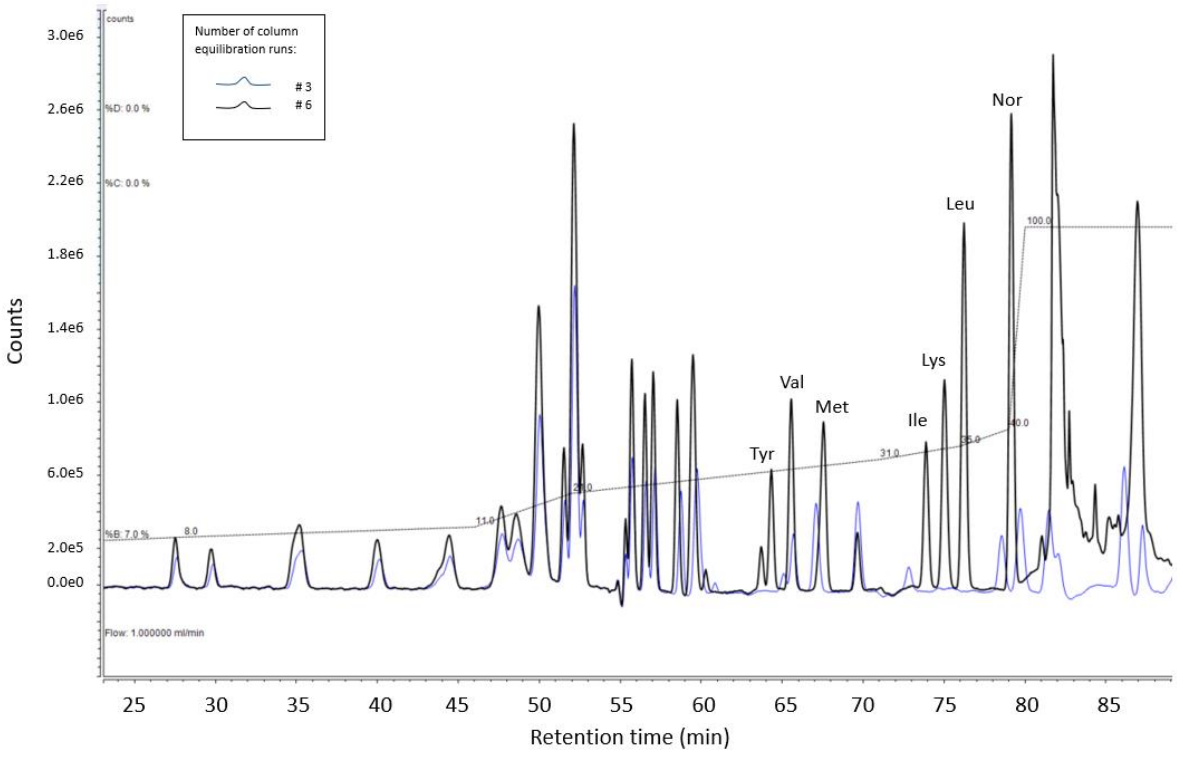


Figure 21. Example of improvement to chromatographic separation over several equilibration runs. Shown here are run number 3 (blue line) and run number 6 (black line).

of the amino acid peaks gained in resolution and sensitivity of the fluorescent detection increased. An example of the improvement is shown in the two overlaid spectra in Figure 21.

As seen in Figure 21, the region of the spectrum where the final hydrophobic amino acids are eluted around the 80+ minute retention time is crowded to the point where the Phe peak seems to be eluting at approximately 82 mins. This is during the 100 % buffer B period where the column is purged of any remaining hydrophobic molecules and the Phe peak is co-eluting with the purge. In order to ensure that Phe and the adjacent amino acids such as Nor, Leu and Lys are separated from the purge and well resolved from one another, the program run time was extended by 10 mins over the period that the gradient of buffer B increases from 21% to 31%, while the adjacent gradients were kept the same. The results are shown in Figure 22. The increased time over this gradient increase allowed all amino acids to elute prior to the purge period, where the gradient rapidly increases from 40% to 100% buffer B over a 1 min period. This change, however, resulted in the merging of some peaks, most notably the Lys and Leu peaks, while producing excellent resolution between

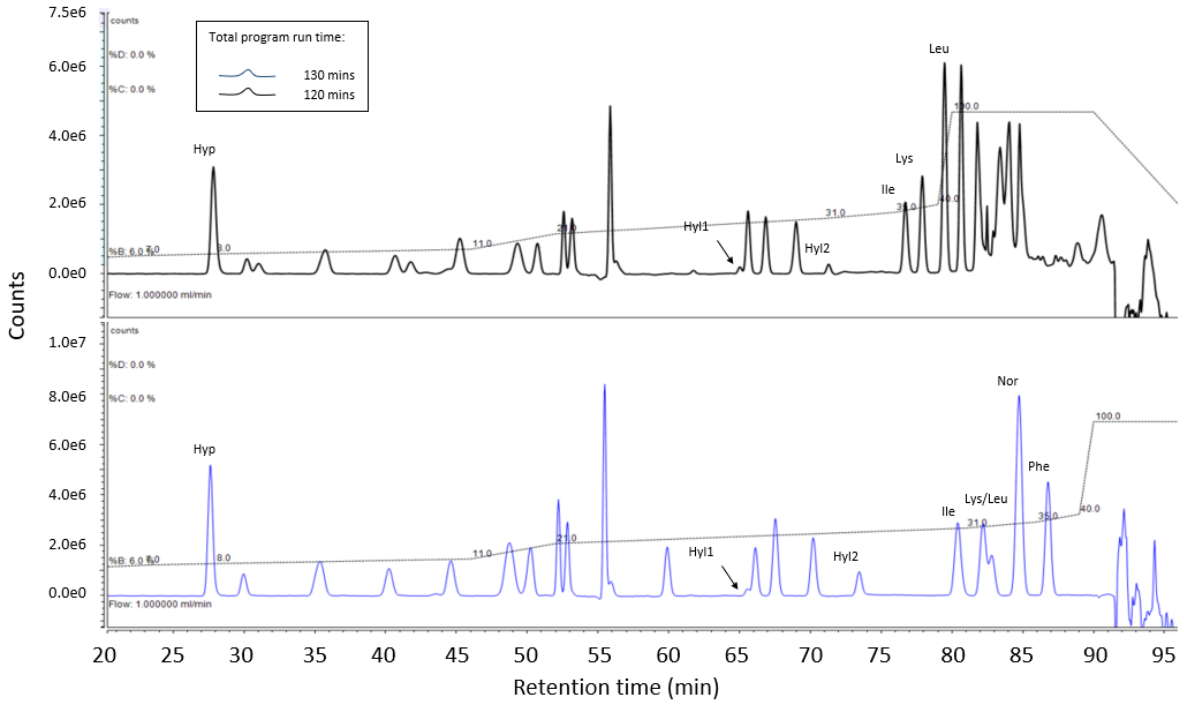


Figure 22. Improved resolution of hydrophobic amino acids in HPLC of amino acid standard eluted over a 130 minute run time.

Nor and Phe, and greater separation of the amino acids in the grouping containing the Hyl1 and Hyl2 diastereomers.

4.3.2 Improving the resolution of the hydroxylysine diastereomers

Now that most amino acids showed good separation from one another, focus was given to improving the resolution of the hydroxylysine diastereomers Hyl1 and Hyl2. While the Hyl2 had excellent separation from surrounding amino acids, the Hyl1 co-eluted with the adjacent Tyr peak. It was thought that changes to the column temperature may cause Hyl1 to elute at a different retention time enabling optimisation of its resolution, as the Hydroxylysines were both seen to move significantly with temperature change in previous runs. The same 130 min run program as seen in Figure 22, which was run at 31 °C column oven temperature, was run again at 27, 29, 31, 33, and 35°C. The results are shown in Figure 23. The spectra are normalised to the Hyp and Phe (first and last) peaks, in order to give a clearer indication of the shift of peaks relative to one another. There is a general trend of lower temperatures increasing the retention of all amino acids, as shown by their

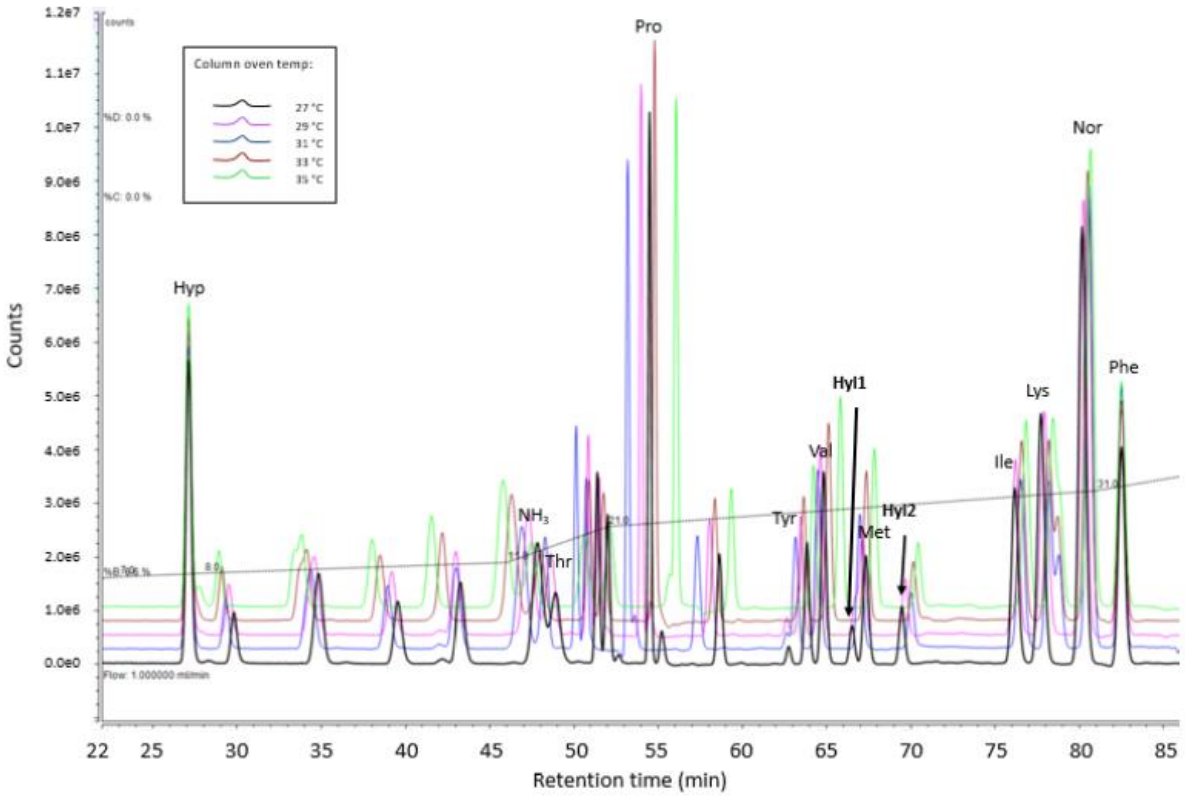


Figure 23. Comparison of different column oven temperatures on amino acid elution profile.

compressed elution profile compared to those at higher temperatures. This is most clearly seen in Figure 23 for the polar amino acids in the early portion of the spectra. The 35 °C spectrum shows the first 6 amino acids eluting closer together, with the spacing between them increasing as the temperature decreases.

The elution profile of the amino acids at 27 °C showed the best separation of the Hyl diastereomers, with Hyl2 position greatly unchanged, but Hyl1 position now between Val and Met and well resolved for quantitation. Most other amino acids were also well resolved at this temperature, with the exception of NH_3 and Thr, however, as these amino acids were not necessary for characterisation of the collagen content or modifications this was deemed acceptable. Correct identification of peaks representing Hyp, Gly, His, NH_3, Arg, Thr, Hyl1, Hyl2, and Nor were confirmed by individual injection and comparison with the mixed standard. All other amino acid peaks were identified based on order of elution and peak positions in the literature. Details of the optimised gradient used for all subsequent elutions are shown in Table 7.

Table 7. Binary gradient for elution of AQC derivatised amino acids using RP-HPLC.

Time	*Flow rate (mL/min)*	*% Buffer A*	*% Buffer B*
Initial	1.0	100.0	0.0
0.5	1.0	99.0	1.0
15.0	1.0	96.0	4.0
23.0	1.0	93.0	7.0
28.0	1.0	92.0	8.0
46.0	1.0	89.0	11.0
52.0	1.0	79.0	21.0
81.0	1.0	69.0	31.0
86.0	1.0	65.0	35.0
89.0	1.0	60.0	40.0
90.0	1.0	0.0	100.0
100.0	1.0	0.0	100.0
110.0	1.0	100.0	0.0
120.0	1.0	100.0	0.0

4.3.3 Calculating the response factors of each amino acid

A calibration curve was created to determine the response factor of each amino acids, which could then be used to determine the amount of each residue present in the test samples. Figure 24 shows the standard curve of 8 injections of the amino acid standards across concentrations from 1-8 µL. The calibration curve is shown in Figure A. 1. The R-squared values of all amino acids are calculated at 1.00, with the exception of Thr at 0.9997 and both Hyl diastereomers at 0.9999. The response factors are calculated by taking the area of each amino acid peak and averaging them across all 8 injections, then dividing the concentration of each amino acid in pmol by the averaged area. This gives the area or 'response' that 1 pmol of each residue generates from the fluorescent detector. These response factors are shown in Table A. 1

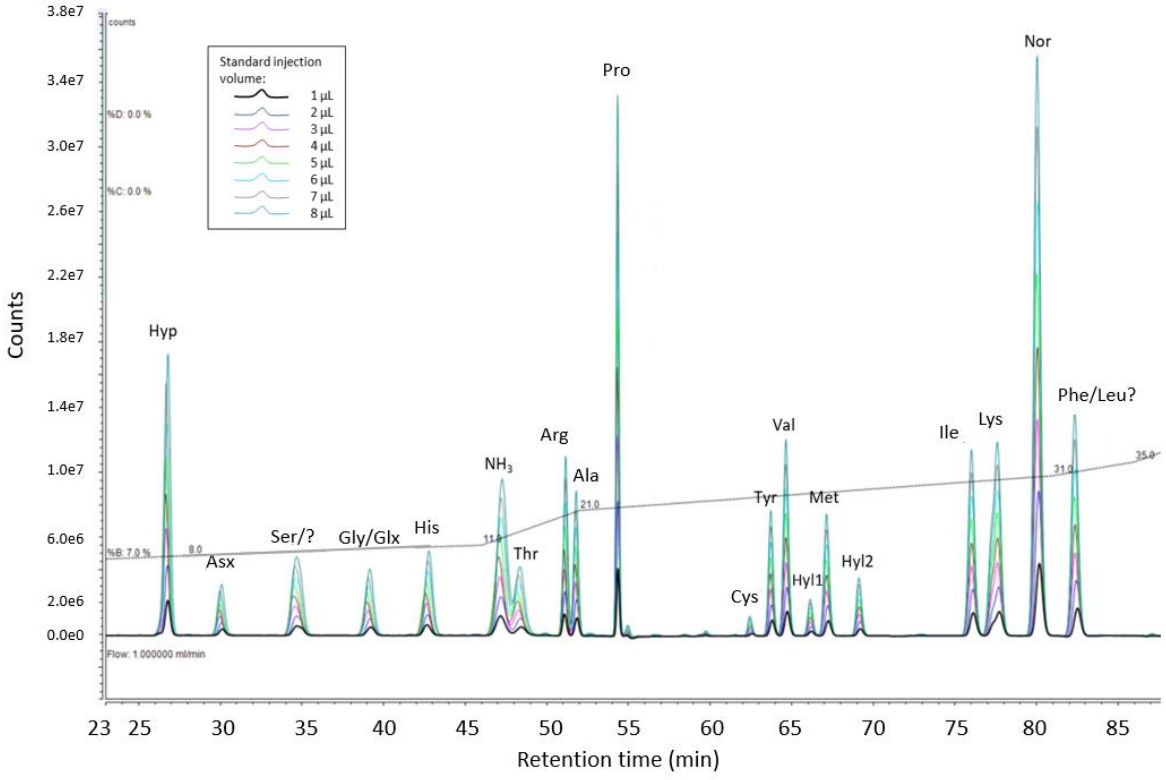


Figure 24. Standard curve of amino acid standards as eluted from a C18 column with RP-HPLC.

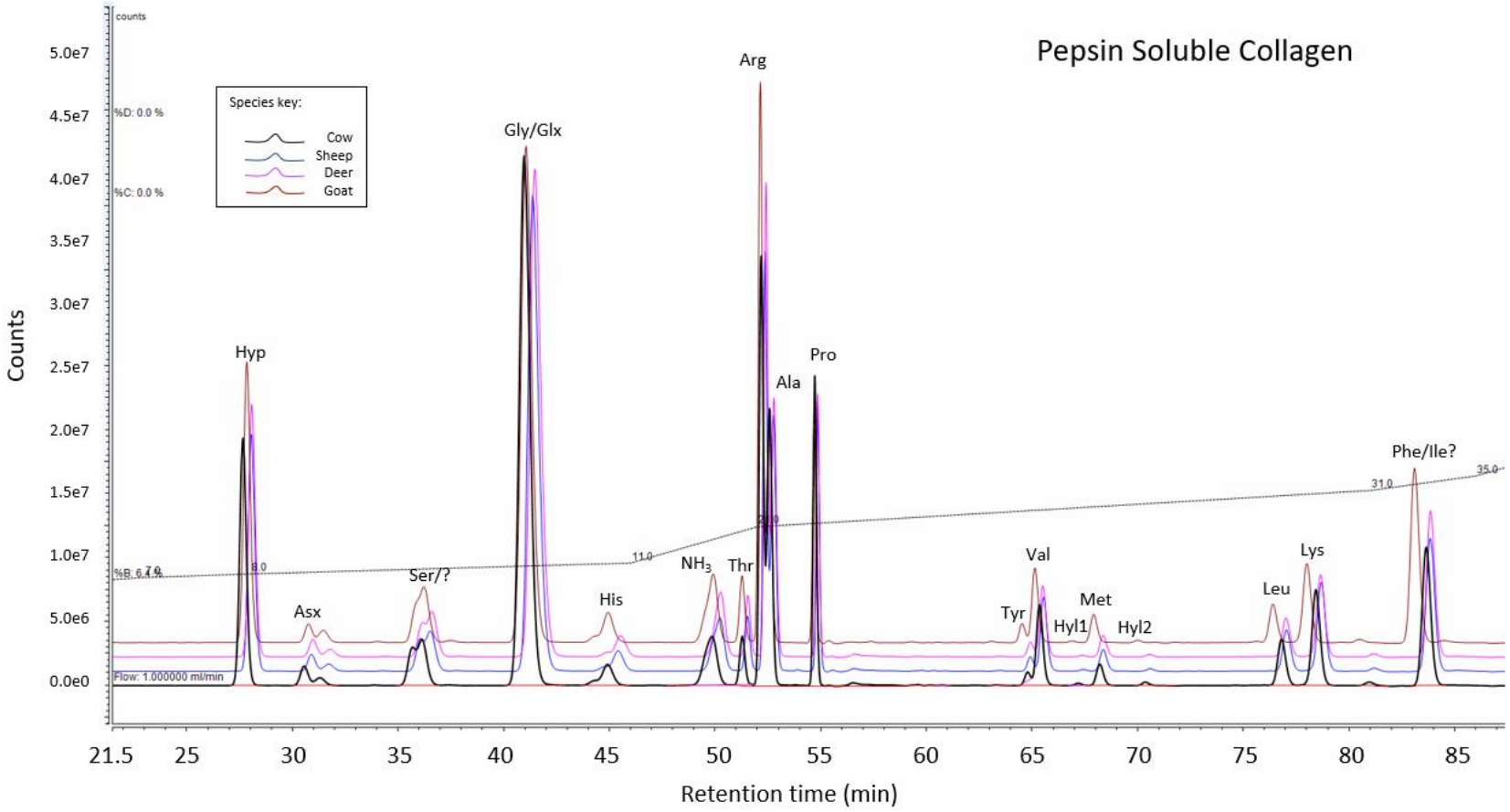


Figure 25. RP-HPLC of the amino acids extracted from four different species skins using pepsin as eluted from a C18 column.

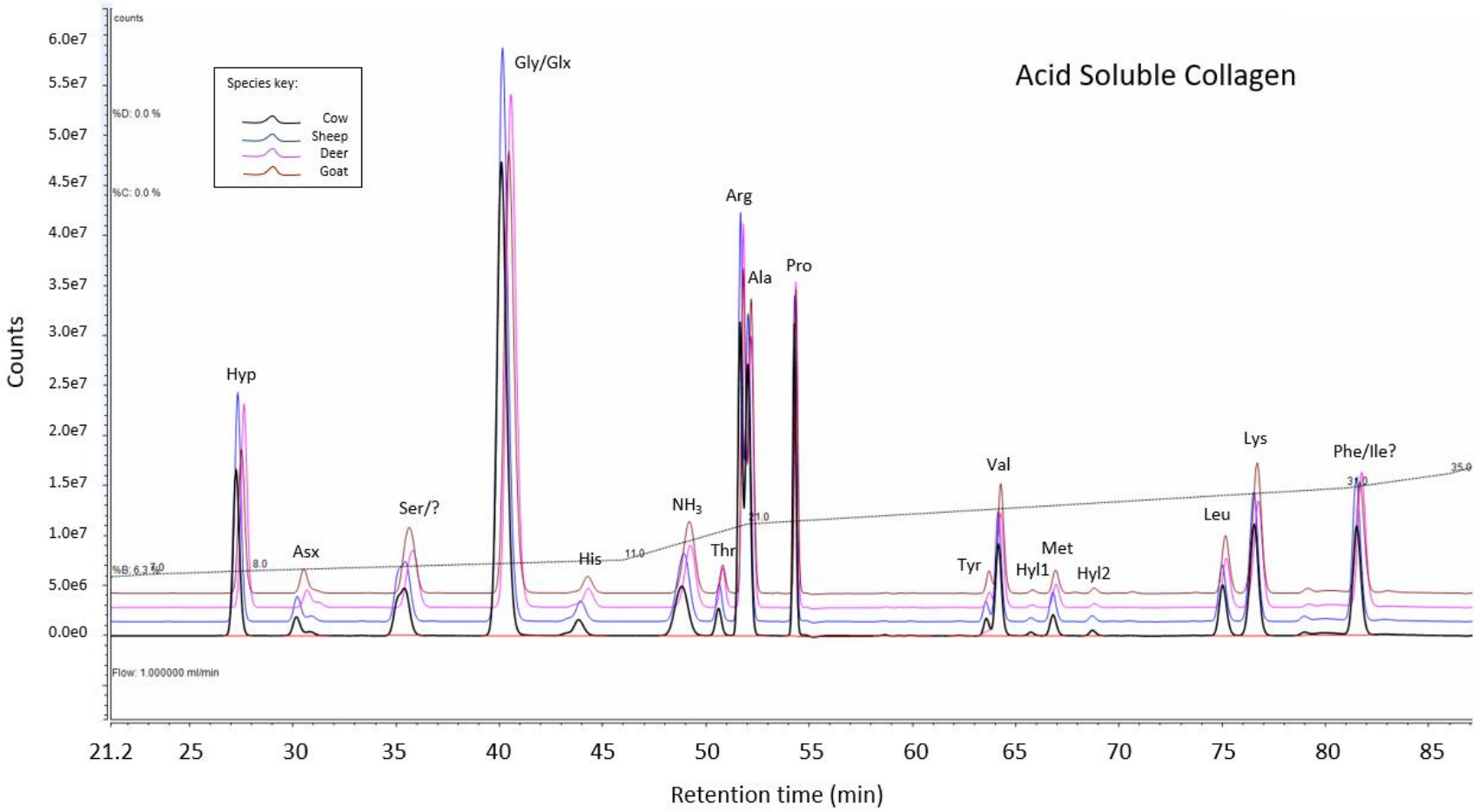


Figure 26. RP-HPLC of the amino acids extracted from four different species skins using acetic acid as eluted from a C18 column.

4.3.4 Hydroxyproline correction

An unfortunate consequence of the optimisation of the hydroxylysine diastereomers is the loss of the AMQ peak as it co-elutes with Hyp. AMQ is one of the artefacts of the AQC derivatisation reaction shown in Figure 17. Several sources state that AMQ fluorescence should not interfere with the amino acid detection, as, while AMQ and AQC share an excitation maxima at 248 nm, they have very different emission maxima, at 520 nm and 395 nm respectively [161, 163, 166]. In previous elution schemes, however a visible AMQ peak was detected by fluorescence that was confirmed by UV detection at 248 nm. Van Wandelen *et al* [167] reported that the resolution between AMQ and Hyp could be improved by reducing the pH of Buffer A., On the other hand Deb Choudhury and Norris [168] found that reducing the pH resulted in Hyl eluting in with the non-polar residues, resulting in poor resolution. As a major focus of this project is hydroxylysine, the elution scheme was optimised for best resolution of the majority of the amino acids with an emphasis on the Hyl diastereomers. Given that each run took 120+ minutes and there were time restraints for the project completion it was decided that a series of previous injections in which the AMQ peak was well separated from the Hyp peak could be used as a calibration to determine the ratio of AMQ to total amino acid content. The calculation for this is shown in the appendix. Briefly, the peak areas for AMQ from a series of injections were summed then divided by the total peak areas of all amino acids from those injections then multiplied by 100 to give a percentage of AMQ relative to all the AAs present. This gave a response factor for AMQ, which was then used to calculate the Hyp concentration. The standard deviation of AMQ peak areas was ±5%, and while this cannot guarantee that the Hyp values reported here are accurately assessed, relative collagen concentrations will be obtained. As the response factor for AMQ was determined using the collagen standard and applied uniformly across all species and extractions it should not affect the comparison of the Hyp and subsequent collagen concentrations found for different species and extraction methods.

4.3.5 Determination of the amino acids in collagen for each species

The amino acids in each sample were determined by multiplying the area of the peak detected for each residue by the response factor calculated for that residue. All calculations can be found in the appendix (Table A. 1). Two replicate injections were performed for

each sample to test the accuracy of the injection, the HPLC machinery and the detection. Each sample had an average coefficient of variation (CV) in peak area across all amino acid replicates of <5 %. If the CV of any given replicate was >10 %, the injection was repeated until a variation of <5 % was achieved. Chromatograms for acid extracted collagen and pepsin extracted collagen for each species are shown in Figure 25 and Figure 26 respectively.

Table 8. Amino acid profile of pepsin extracted collagen expressed as molar percentages.

	% mol/mol in pepsin extraction			
Amino acids	*Cow*	*Sheep*	*Deer*	*Goat*
Gly/Glx	42.00	42.88	42.25	40.99
Hyp	15.40	16.70	16.29	17.76
Arg	10.20	10.08	11.12	11.53
Ala	9.50	8.63	8.33	7.56
Ser/?	3.90	3.46	3.76	3.69
Phe/Leu	3.01	3.23	3.33	3.61
Pro	2.97	2.64	2.49	2.11
Valine	2.47	2.27	2.03	1.99
Asx	2.15	1.88	1.79	1.87
Lys	1.98	1.81	1.65	1.57
Thr	1.45	1.70	1.77	1.97
His/?	1.41	1.56	1.54	1.88
Ile	1.32	1.20	1.13	1.07
Met	1.03	0.97	1.05	1.04
Hyl2	0.42	0.28	0.27	0.25
Hyl1	0.38	0.26	0.25	0.26
Tyr	0.32	0.34	0.81	0.72
Cys	0.10	0.11	0.13	0.14

Table 9. Amino acid profile of acid extracted collagen expressed as molar percentages.

% mol/mol in acid extraction				
Amino acids	*Cow*	*Sheep*	*Deer*	*Goat*
Gly/Glx	43.66	47.97	35.97	41.42
Hyp	12.27	11.47	12.87	10.14
Ala	10.06	8.42	10.97	10.19
Arg	9.52	8.80	12.12	9.90
Ser/?	3.97	4.52	4.37	4.59
Pro	3.20	2.49	3.71	2.99
Phe/Leu	3.10	2.67	3.85	3.27
Valine	2.74	2.50	3.44	3.26
Lys	2.41	2.12	2.72	3.15
Asx	2.11	2.56	2.16	2.96
Ile	1.49	1.25	1.68	1.84
Thr	1.35	1.21	1.70	1.24
His/?	1.31	1.60	1.78	1.53
Met	1.00	0.97	1.28	1.29
Tyr	0.60	0.39	0.60	1.27
Hyl1	0.56	0.46	0.36	0.50
Hyl2	0.52	0.42	0.32	0.45
Cys	0.13	0.20	0.11	0.00

The raw and calculated data for each sample is shown in the accompanying spreadsheet '*Amino acid calculations.xlsx*'. Briefly, the area of each peak as recorded by fluorescence was multiplied by the response factor for that amino acid, to give the picomoles (pmol) of that amino acid present in the injection volume. This was then multiplied by the dilution factor of the sample to give the μmol of amino acid present in the total extract. These amounts were then divided by the total μmol of all amino acids present and multiplied by 100 to give the % mol/mol values for each sample. The results are shown in Table 8 and Table 9 and have been calculated after the correction for hydroxyproline (section 4.3.4).

Concentrations are presented in order from highest to lowest amino acid concentration, determined with respect to the cow amino acid profile (second column of each table). In both the acid and pepsin extractions, glycine is the dominating amino acid with a range of ~35–48%. The theoretical percentages put glycine at approximately 33% (±0.5%). This is significantly lower than the experimental amounts shown in Table 8 and Table 9. Acid

extraction of sheep collagen in particular gave a glycine percentage of 47.97%, accounting for nearly half of the total amino acids in the collagen molecule. Hydroxyproline was the second most highly represented amino acid, and the value obtained is in better agreement with the theoretical values. The number of glycines that should be detected by AAA can be determined using the translated gene sequence for bovine Col1A (P02453) retrieved from the UniProtKB database (www.uniprot.org). The FASTA protein sequence was downloaded and the ExPASy ProtParam tool used to calculate the percentage of each amino acid present in the sequence with and without the telopeptides. Pepsin extracted collagen should contain a greater proportion of Gly, Hyp and Pro than acid extracted collagen, in which the telopeptides (which do not contain this motif) are retained. These differences, and the theoretical molar percentages for each amino acid are shown in Table 10.

Table 10. Theoretical molar percentages of amino acids based on FASTA data for the COL1A gene in *Bos taurus* (bovine).

Theoretical % amino acids of total collagen α1-chain		
Amino acids	*With telopeptides*	*Without telopeptides*
Gly	32.9	33.5
Hyp/Pro	22.7 (11.4)[1]	23.3 (11.7)[1]
Ala	11.1	11.4
Arg	5.0	5.0
Glu	4.6	4.6
Ser	3.7	3.4
Lys	3.6	3.6
Asp	3.2	3.1
Gln	2.7	2.7
Leu	1.9	1.9
Thr	1.9	1.9
Val	1.7	1.7
Phe	1.4	1.2
Asn	1.2	1.3
Ile	0.9	0.8
Met	0.8	0.7
Tyr	0.4	0.0
His	0.3	0.2
Cys	0.0	0.0

[1]Brackets denote the value of each Hyp and Pro if they are in a 1:1 ratio.

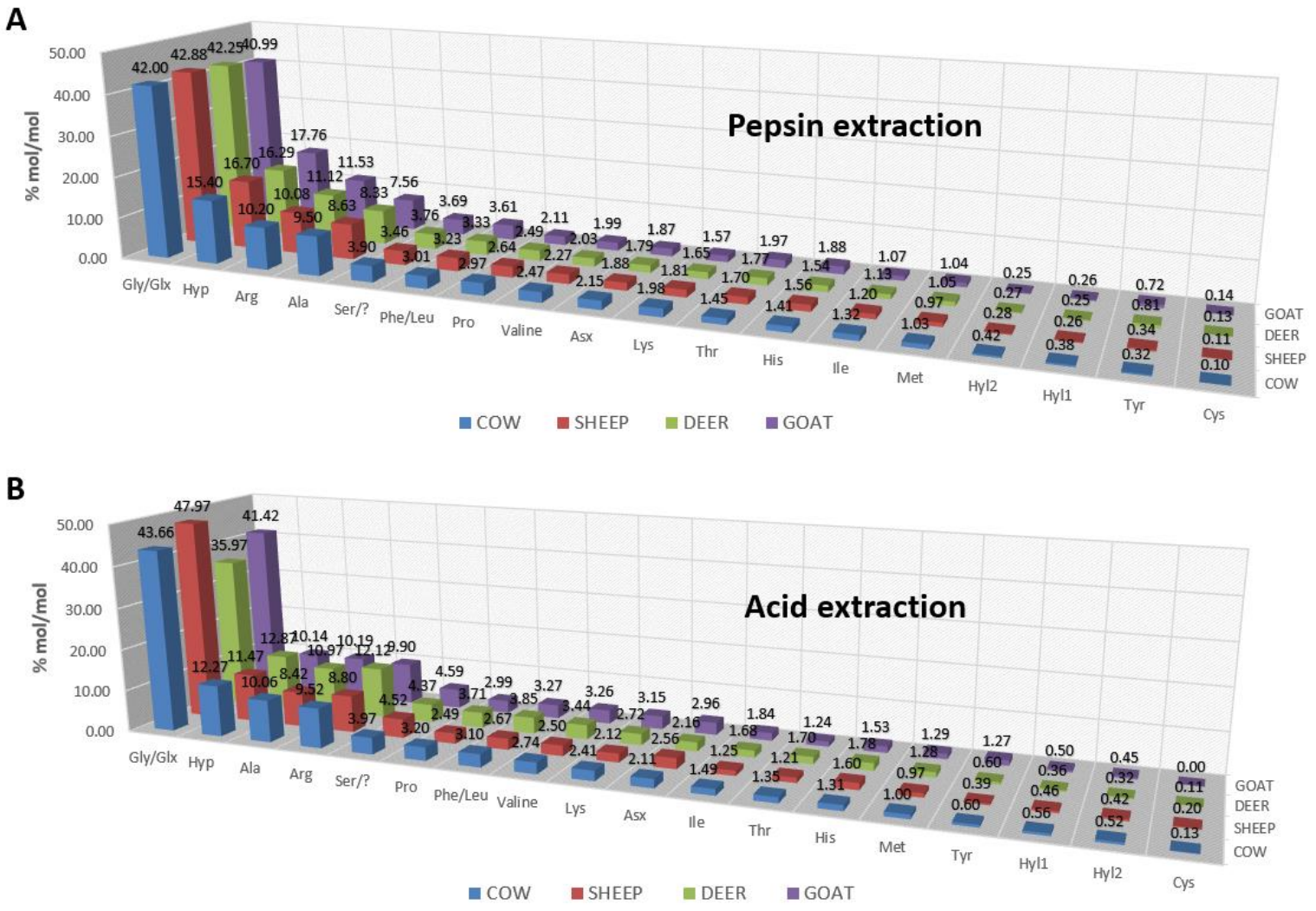


Figure 27. Amino acid molar % across all four species, as determined by RP-HPLC amino acid analysis. A is the amino acid profile of pepsin extracted collagen; B is from acid extracted collagen.

The proportion of glycine in a single collagen polypeptide should be 33% (±0.5%), a value that is significantly lower than the 48% obtained experimentally (Table 8 and Table 9). However, the Glu/Gln (Glx) peak is not present, and the theoretical values for Glx and Gly combined total to ~38% and as the Glx peak typically elutes just prior to Gly, it was assumed that the Glx peak may be co-eluting with Gly.

It is impossible to calculate the number of Hyp residues in a collagen polypeptide chain as the gene sequence cannot predict which proline residues will be hydroxylated. Numerous studies on collagen structure suggest that Hyp and Pro should occur in a 1:1 ratio, each making up 1/6 of the total amino acids. If the number of Pro residues in the sequence is halved, the fraction of Pro and Hyp residues in the collagen sequence with the telopeptide (acid extraction) should be 11.4 % and without the telopeptide (pepsin extraction) domain should be 11.7 %. In this study the proportion of Hyp obtained using acid extraction made up 10–13 % of the sequence, in agreement with the theoretical value. However, when the collagen was extracted from the four animal skins with pepsin, the proportion of Hyp was much higher at 15–18 %. Could this anomaly be due to interference from AMQ giving an abnormally high reading for Hyp? If this is the case then the number of Hyp residues measured in acid extracted collagen would also be overestimated making it well below the theoretical estimation. There may be another factor which is contributing to the increased number of Hyps in the pepsin extracted samples. If Hyp and Pro are truly present in a 1:1 ratio, then Pro is greatly underestimated. In acid extracted collagen Pro makes up 3–4 % of the sequence, while in pepsin extracted collagen, it makes up 2–3 %. These values are well below the theoretical value of approximately 11 %. Figure 28 summarises the differences seen between the theoretical and the actual mole percent values for hydroxyproline and proline combined. Because the FASTA data cannot give an indication of the ratio of Hyp to Pro, the results from this study can only be compared to amino acid analyses of collagen by other groups. An extensive search of the literature found no mention of possible under-detection of proline from AQC-derivatised HPLC amino acid analysis. However, this is not surprising, given that most proteins do not contain the levels of proline shown in collagen. A thorough examination of the collagen literature found no other amino acid analyses that reported Pro levels as low as those found in this study. Perhaps more interestingly, the ratios of Hyp to Pro measured in collagen appear to vary greatly among

groups, even among those using similar methodologies. Table 11 and Figure 29 compare the number of individual amino acids measured in bovine collagen (often crude skin preparations) by other research groups to the results found in this study. Eastoe [169] and Gallop and Seifter [170] used post-column derivitisation with ninhydrin, while Cohen and van Wandelen [171], Deb Choudhury and Norris [168], and Naffa *et al* [144] used pre-column derivatisation with AQC. The Eastoe, Gallop and Naffa groups used whole skin samples, as Naffa hydrolysed whole skin pieces, and Eastoe and Gallop hydrolysed non-differentiated whole extracts from skin and gelatin respectively. Deb Choudhury and Cohen used collagen type I standard obtained from Sigma. According to the Sigma website, the calf skin collagen type I standard used in this study was prepared using a modified version of the Gallop and Seifter method. Unless this method has been vastly altered from the original, it involves extraction from minced calf skin using sodium acetate and sodium citrate buffer at a pH of 3.7 followed by deionisation and lyophilisation. There appears to be no differential protein precipitation such as 'salting out' using ammonium sulfate or sodium hydroxide, and the authors provide no indication as to the purity of the collagen in the extract, such as could be demonstrated with an SDS-PAGE gel. This appears to be the case for all of the studies used to compare the results obtained in this study. As such it cannot be known whether there are high concentrations of proteins other than collagen contaminating these samples, and if there are, these could have a significant impact on the amino acid profile reported. Nevertheless the results reported for Gly and Pro in this study are at odds with what is expected from knowledge of the DNA sequence, as are other amino acids in the sequence.

Numerous studies on acid hydrolysis of proteins have also reported differences in the hydrolysis of the peptide bond of proline. Hill *et al* [172], gave a comprehensive review of all forms of protein hydrolysis, and found that with acid hydrolysis free prolines showed increased destruction relating to the length of time of the hydrolysis experiment. This concurred with the findings of Gordon *et al* [173] who demonstrated that although the proline peptide bond, particularly on the amine side was broken late in the hydrolysis process and in some cases did not break at all. When subjected to extended hydrolysis over 19 days at 37 °C (known as partial hydrolysis) free proline levels decreased due to the proline degrading in the acid, particularly in protein solutions which contained high levels

of sugars. Hill *et al* also showed that results seen for proline hydrolysis were not the same for hydroxyproline, which was released from the peptide bond earlier than proline, and under less intensive procedures [172]. This may account for the large differences in the proline concentrations seen between different analyses shown in Table 11 and Figure 29. One possibility is that proline has been fully hydrolysed from the peptide bond, but that the free proline has degraded at a greater rate than the other amino acids, as found by Gordon *et al* [173].

When the levels of Hyp and Pro are combined, the levels determined in this study can be compared ignoring possible overestimation and underestimation of Hyp and Pro

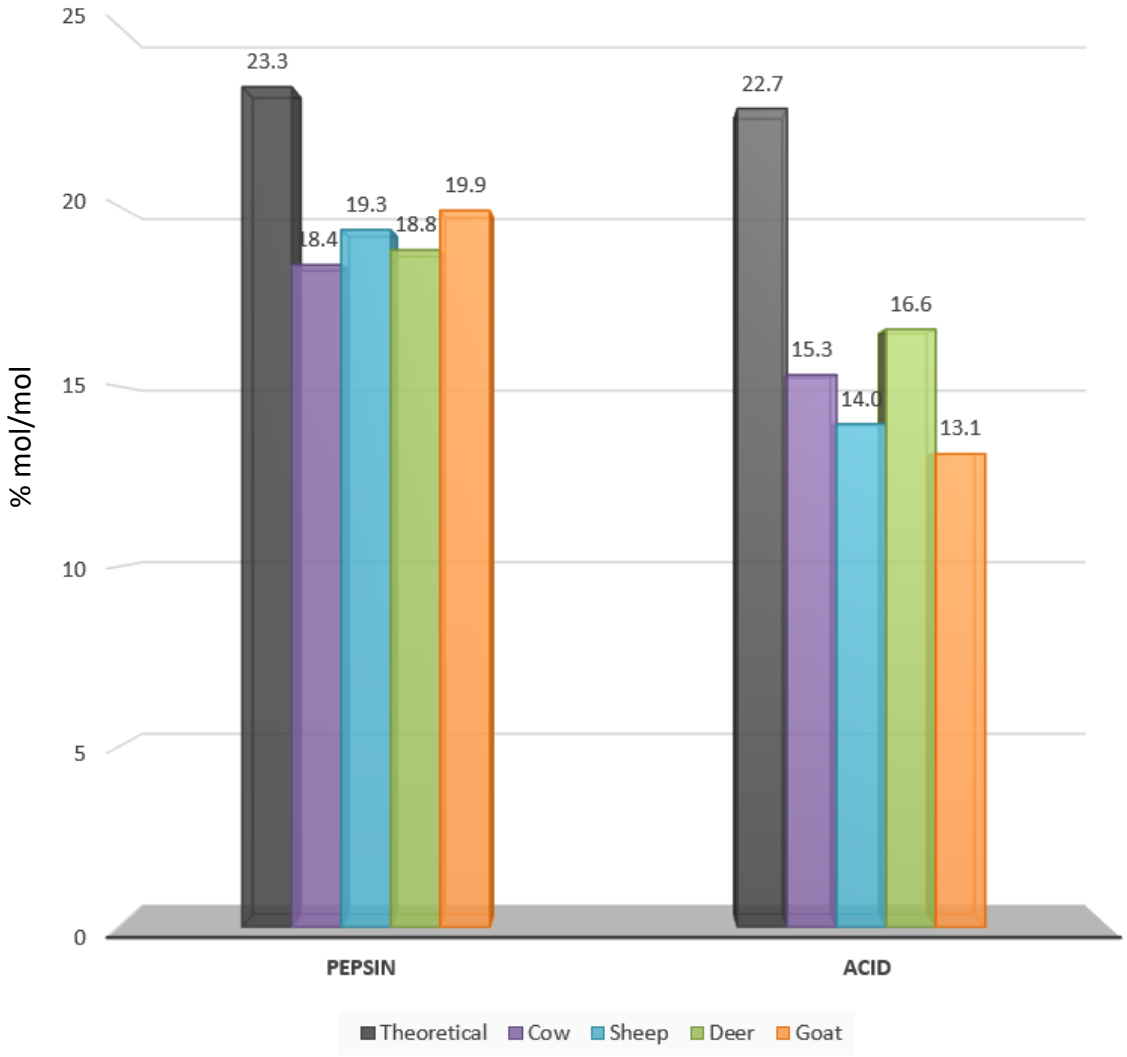


Figure 28. Combined proline and hydroxyproline % molarity measured for each animal versus the theoretical combined values determined by the FASTA sequence.

respectively. Table 11B shows that all of the previous studies compared here underestimated the level of total proline (Hyp or Pro, from now on referred to as *Pro) when compared with the FASTA sequence, except for Deb Choudhury who reported higher levels. The results of *Pro for the current study have a ratio of approximately 0.67 of the averaged results from other groups for the acid-extracted collagen, and 0.80 for the pepsin-extracted collagen.

When other amino acids are compared, there are some other significant differences seen between the current study and previous groups. Asp concentrations are significantly lower for both acidic and pepsin extractions, with Asp showing approximately half the number of residues as expected. Inversely, Arg levels were ~2 fold greater than those detected by other groups, Met was ~1.5 fold higher than the expected theoretical number, as was Val. The peak detected for His in the present study was 13–14 residues per 1000 residues for acidic and pepsin extractions respectively, yet for other groups His peaks yielded a result of 3–6 residues, a ~3 fold increase on the experimental results found here, and ~6.5 fold higher than the expected theoretical number. One Cys residue was also detected in both the acid and pepsin extractions, while the FASTA sequence contains no Cys residues in collagen type I.

These differences are disturbing, and possible reasons for the differences need to be addressed, particularly the differences with the theoretical results. The standard deviations for each of the amino acids between separate injections to the column was in most instances <1%. This shows that these differences can not be accounted for by errors in the machine fluorescent detection, the injection apparatus or in the column itself.

However, when the standard deviations of the technical replicates (the three separate pieces of skin from the same hide processed in parallel for each animal) are compared, the differences are quite significant. Table 12 summarises the standard deviations for each amino acid across each species and extraction type. At a glance it is clear that specific extraction and species combinations produce greater deviations between the results than others, although this is not generally the case. For example, A-Deer and A-Sheep show the greatest average variation across all amino acids at 38.54, 20.56 and 15.73 respectively.

However, P-Sheep and A-Cow show the lowest average variation at 4.37 and 7.94, respectively. Medium variation between the extremes were shown for P-Goat, P-Cow, P-Deer and A-Goat at 15.73, 13.33, 11.21 and 8.96 respectively.

Table 11. Comparison of amino acid residues/1000 in bovine collagen or skin as reported by previous groups, in this experimental study, and theoretical (FASTA sequence). A: All amino acids. B: Hyp and Pro combined. 1. Cohen & Seifter [171], 2. Eastoe [169], 3. Gallop et al [170], 4. Deb Choudhury & Norris [168], 5. Naffa [144]

A	*Previous studies (Cow, ox or calfskins or non-purified collagen)*					*Current Study (Cow collagen)*		*FASTA file (Cow collagen)*	
Amino acids	*1.*	*2.*	*3.*	*4.*	*5.*	*Acid*	*Pepsin*	*Acid*	*Pepsin*
Hyp	107	98	91	173	88	123	147	114[1]	118[1]
Asp	44	46	42	36	47	21	22	42	41
Ser	34	36	30	29	34	40	39	36	33
Glu	85	70	70	61	83	-	-	75	73
Gly	316	333	350	333	311	437[†]	423[†]	324	335
His	3	5	5	4	6	13	14	3	2
Arg	54	46	48	47	62	95	103	50	50
Thr	26	17	16	20	10	13	15	16	16
Ala	99	112	111	79	108	101	96	116	119
Pro	113	129	124	91	124	32	30	114[1]	118[1]
Cys	2	0	0	0	0	1	1	0	0
Tyr	3	1	3	6	7	6	3	5	0
Hyl1	24[‡]	5[‡]	7[‡]	16[‡]	2	6	4	-	-
Val	19	20	19	20	24	27	25	16	16
Hyl2	-	-	-	-	3	5	4	-	-
Met	6	5	-	6	6	10	10	7	7
Lys	16	28	25	24	28	24	20	36*	36*
Ile	9	12	11	14	14	15	13	9	8
Leu	26	23	24	28	29	-	-	21	19
Phe	13	12	12	17	15	31[°]	30[°]	12	12

[†]Glycine and glutamate coeluted together. [‡]Hydroxylysine diastereomers reported as one Hyl peak. [°]Phenylalanine and leucine coeluted together.

B	*Previous studies (Cow, ox or calfskins or non-purified collagen)*					*Current Study (Cow collagen)*		*FASTA file (Cow collagen)*	
Amino acids	*1.*	*2.*	*3.*	*4.*	*5.*	*Acid*	*Pepsin*	*Acid*	*Pepsin*
Hyp	107	98	91	173	88	121	154	117[•]	117[•]
Pro	113	129	124	91	124	32	30	123	119
Total	220	227	215	264	212	153	184	240	236

[•]Levels estimated from total Pro based on the preference of Pro for the X position and Hyp for the Y position of the Gly-X-Y motif.

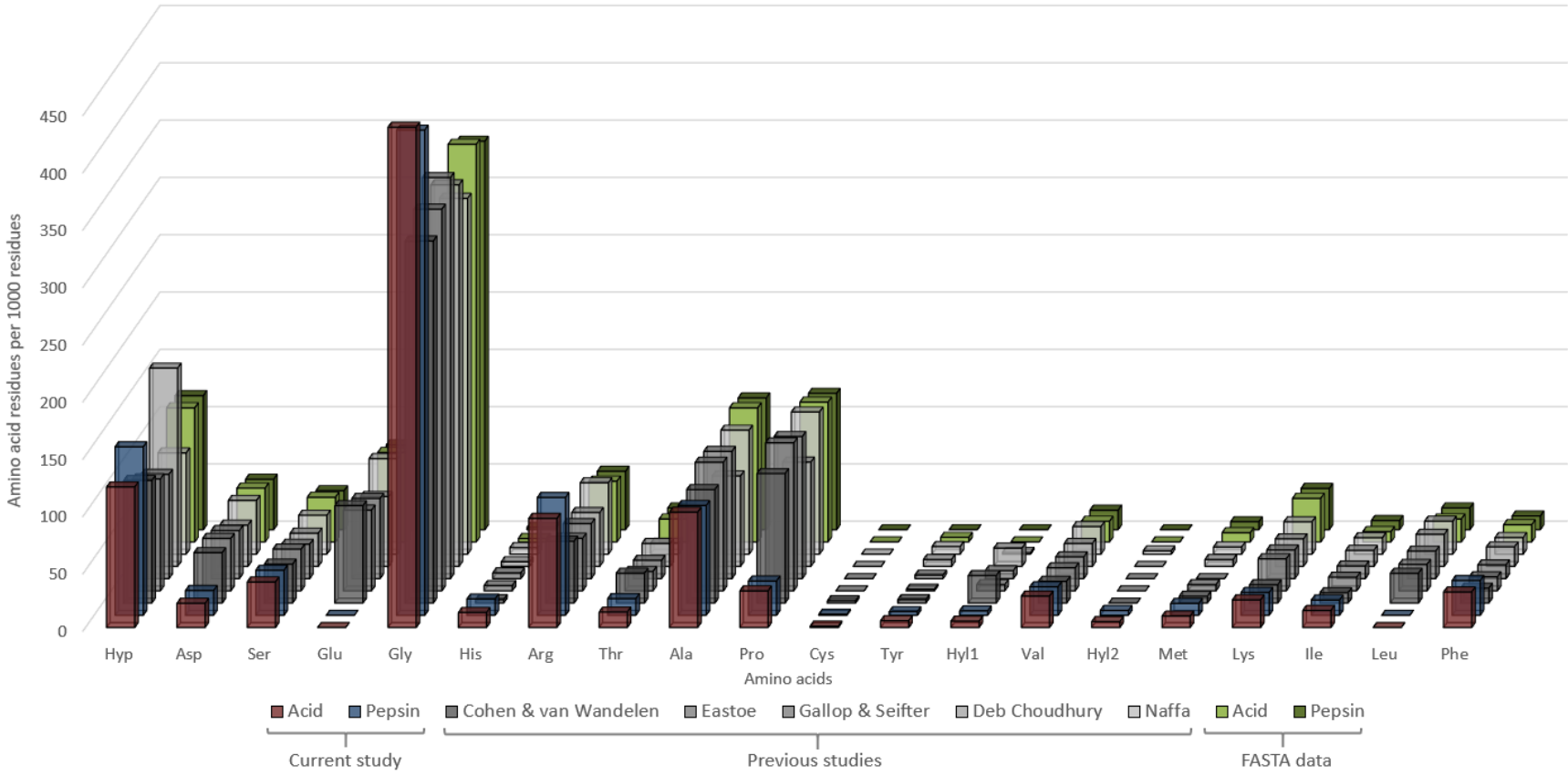


Figure 29. Comparison of number of amino acid residues found experimentally and in the literature for either bovine skin or bovine collagen.

Table 12. Comparison of coefficient of variation (CV) between technical replicates for each species and extraction type. P or A before the species denotes pepsin and acid extractions, respectively.

	CV% of % mol/mol							
AAs	*P-Cow*	*A-Cow*	*P-Sheep*	*A-Sheep*	*P-Deer*	*A-Deer*	*P-Goat*	*A-Goat*
Hyp	6.40	5.24	1.57	19.58	6.94	0.99	7.55	7.65
Asx	13.03	4.19	4.55	7.92	6.67	28.56	10.25	8.72
Ser/?	9.23	3.89	1.66	9.21	5.30	38.42	2.87	3.99
Gly/Glx	7.85	0.58	0.53	17.78	2.84	64.29	4.67	1.07
His/?	1.69	17.37	4.63	15.19	17.61	39.70	17.67	11.11
Thr	6.42	4.66	4.48	21.74	6.89	40.04	10.48	7.80
Arg	1.79	4.92	1.67	24.52	4.60	42.71	13.91	6.20
Ala	14.21	6.98	2.31	20.51	7.67	35.27	11.89	9.04
Pro	14.64	6.18	2.22	16.87	10.00	37.65	18.05	7.16
Cys	9.49	17.34	4.86	9.39	9.69	15.87	53.47	0.00
Tyr	27.94	28.43	18.57	32.14	38.13	42.12	11.82	22.29
Valine	14.21	7.67	1.94	22.16	4.08	47.37	8.00	4.23
Hyl1	25.24	5.77	6.88	29.33	26.31	55.44	27.26	17.77
Met	17.95	2.47	5.54	30.51	12.10	41.33	8.92	11.78
Hyl2	29.54	6.98	11.20	32.12	23.10	57.88	33.90	16.91
Ile	15.35	5.18	1.01	20.14	5.62	37.18	12.64	7.59
Lys	20.14	8.23	2.29	21.10	5.18	31.21	17.11	8.06
Phe/Leu	4.74	6.79	2.71	19.80	9.09	37.65	12.59	9.85
Average variation	13.33	7.94	4.37	20.56	11.21	38.54	15.73	8.96

This difference between combinations of species and extractions may relate to the conformations of the higher order collagen structures within these skins. Sheep and deer collagen fibrils have been found to have similar structural conformation with regards to their fibril orientation in the skin and the fibril diameter size through previous work undertaken in this lab [144]. Cow skins were found to be in a mostly vertical orientation in relation to the skin direction, and with the largest fibril diameter size, however they were also found to have the greatest deviation across all data collected. This is in direct contrast to sheep and deer skins which were found to have fibrils of a mostly horizontal orientation, and with the smallest diameter sizes, with both also displaying the narrowest deviation across their results. Goat fibrils displayed mixed vertical and horizontal fibrils of medium diameter with a medium deviation of the data.

This is interesting when taken in conjunction with the results shown here. Both sheep and deer samples displayed similar properties of large variations across their acid extractions, and small to medium variations across their pepsin extractions, while in contrast to the previous study, cow collagen produced medium variation across the pepsin extractions and small variation across the acid extractions. This may relate to the accessibility of the collagen fibrils to solvent and enzyme. It is possible that the specific orientation and small size of the sheep and deer fibrils gives greater access to the soluble collagen portion of the skin, resulting in a more efficient extraction with acid. In contrast, the larger fibril size of cow skin may result in a less efficient acid extraction, with less resistance to pepsin digestion. This may provide different balances of acid and pepsin extracted collagen within the pepsin extracts, which would result in variation of the amino acid profile due to different mixes of telopeptide and non-telopeptide containing collagens.

4.3.6 Hydroxylysine and its diastereomers

4.3.6.1 In an acidic environment

Both hydroxylysine 1 and 2 were seen in all samples and their molar percentages are summarised in Table 13 and Figure 30. Interestingly, Hyl1 is predominant in all acid extracted collagen, and Hyl2 is predominant in all pepsin extracts with the exception of goat. The highest molar percentage of Hyl2 is, however found in cow collagen, which also contains the highest level of Hyl1, and therefore the highest level of both Hyl diastereomers combined.

Table 13. Molar percentages of hydroxylysine diastereomers as determined by RP-HPLC.

	Pepsin extraction (% mol/mol)			
Amino acids	*Cow*	*Sheep*	*Deer*	*Goat*
Hyl1	0.38	0.26	0.25	0.26
Hyl2	0.42	0.28	0.27	0.25
Ratio (Hyl1/Hyl2)	0.91	0.96	0.94	1.05
	Acid extraction (% mol/mol)			
Amino acids	*Cow*	*Sheep*	*Deer*	*Goat*
Hyl1	0.55	0.46	0.36	0.50
Hyl2	0.51	0.42	0.32	0.45
Ratio (Hyl1/Hyl2)	1.07	1.08	1.13	1.12

Collagen fibrils in cow skin have the largest diameter of all skins tested and are stabilised by more crosslinks than those in the other animal skins analysed in this study. Hyl and Lys residues involved in forming crosslinks will not show up in amino acid analysis Furthermore Hyl can also be glycosylated in which case it will not coelute with either Lys or Hyl1 or Hyl2 in AAA. These modifications which cannot be predicted, will impact on the levels of Lys and Hyl measured by AAA resulting in an underestimation of the number of these residues in the sample as found in this study.

Both Hyl1 and Hyl2 concentrations are lowest in pepsin extracted goat collagen, yet second highest in acid extracted goat collagen. Interestingly, goat collagen was the only one to have higher concentrations of Hyl2 in both acid and pepsin extracted collagen which could relate to the structure of its higher order structure. It is possible that the higher level of Hyl2 in goat skin collagen influences the cross-directional matrix of collagen fibrils seen in the skin. The presence of both Hyl diastereomers in one collagen fibril would result in multi-

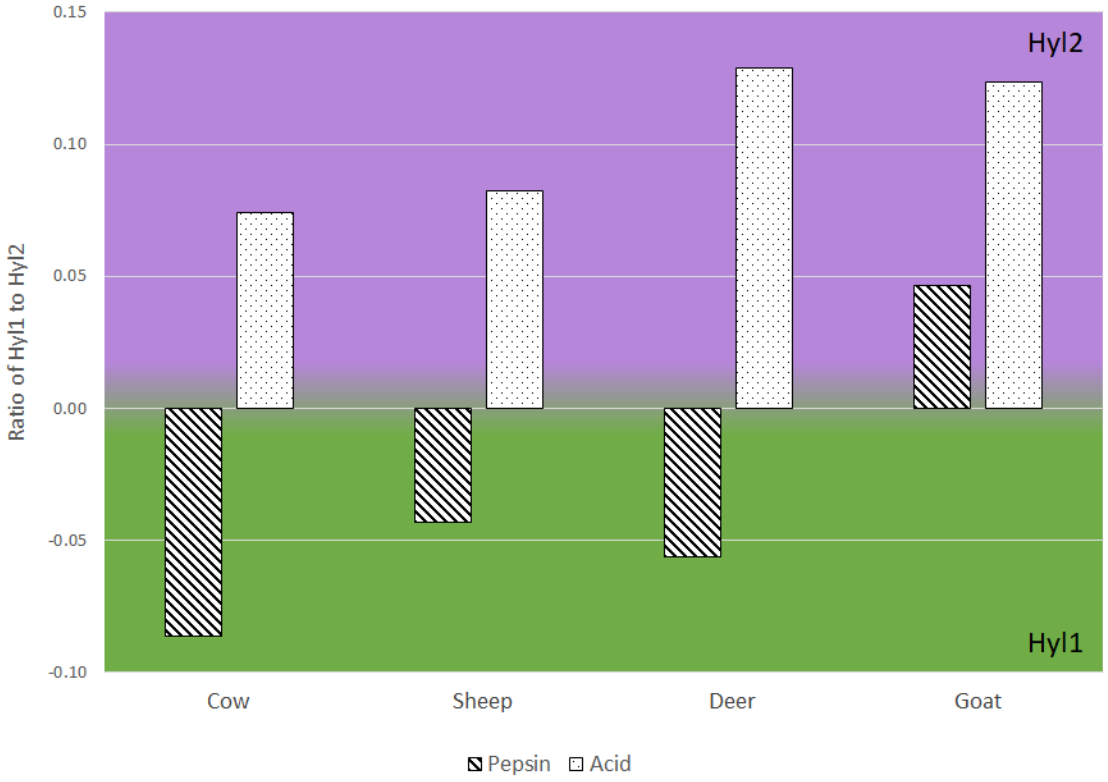


Figure 30. Ratio of Hyl1 to Hyl2. Positive bars rising above the 0.00 line indicates a higher level of Hyl2 present, while bars falling below the 0.00 line indicate higher levels of Hyl1.

directional cross-linking due to the angle of rotation of the -OH bond causing cross-links formed at these centres to point in opposing directions.

The bias towards in Hyl2 in the more soluble acid extracted collagen may indicate a differentiation between the orientation of Hyl hydroxylation in younger versus older skin. Soluble collagen which is readily removed by acid is has been found to be in lower in concentration as the skin ages. Schnider *et al* [174] found that as age increased the ratio of soluble to insoluble collagen decreased, as soluble collagen was predominantly composed of newly synthesised procollagen molecules, and production of new collagen decreased with age. Further studies need to be undertaken to compare Hyl diastereomers in skins from both juvenile and mature animals to determine if there is indeed a correlation.

4.3.6.1 In a neutral environment

All of the above results are based on the hydrolysis of collagen in acid conditions and questions have been raised as to whether these two epimers of Hyl exist naturally or are artefacts of acid hydrolysis. Some researchers are convinced under acid conditions the -

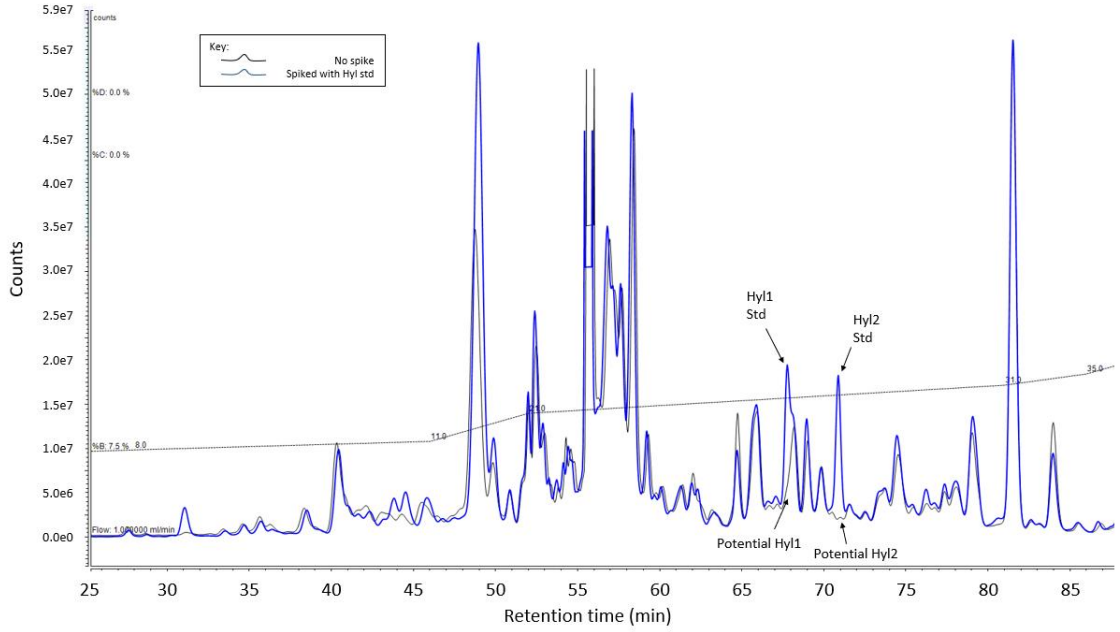


Figure 31. HPLC chromatogram of salt-extracted collagen type I enzymatically hydrolysed sequentially with actinidin, proteinase K and carboxypeptidase y. The black line denotes the chromatogram of the enzymatically hydrolysed sample (no spike), and the blue line is this same sample with the addition of a racemic Hyl standard containing both diastereomers, added to the sample prior to hydrolysis (spiked sample). Arrows denote potential Hyl diastereomer peaks.

OH bond on the Hyl residues is subject to epimerisation resulting in the formation of the second epimer. In order to determine if two epimers could be detected by RP-HPLC when the collagen molecules had not been exposed to acid at any stage in their handling, an enzymatic hydrolysis was attempted. Unfortunately, the elution profile of the amino acids could not be resolved adequately, as the spectrum produced multiple additional peaks, most likely due to incomplete hydrolysis of the peptides. Even following four treatments with crude actinidin and subsequent proteinase K digestion for 48 hours the hydrolysis still appeared to be incomplete. Figure 31 shows the elution profile of neutrally extracted, enzymatically hydrolysed cow collagen. While hydrolysis is clearly incomplete, spiking with racemic Hyl standard shows two potential peaks which could be Hyl peaks from the hydrolysate. The large peak showing in black beneath the Hyl1 standard appears to have a rider on its left slope as it is asymmetrical. This could be the Hyl1 peak in the hydrolysate which should be present in either acid or neutrally hydrolysed collagen. However, there also appears to be a small peak directly beneath the Hyl2 standard, which could correspond to Hyl2 in the hydrolysate. If this is in fact Hyl2, it would put to bed the argument that Hyl2 is an artefact of acid hydrolysis, as this sample was not exposed to an acidic environment at any stage of its processing. Time restraints did not allow for it, but future work should include LC-MS/MS analysis of this sample to confirm definitively that Hly2 is present.

4.4 Conclusion

When the amino acid profiles obtained in the current study are compared with those obtained by other groups the general pattern is similar. However this study found increased levels of Gly, Val, and His and reduced levels of Pro, Arg, Asp compared to the numbers expected from the sequence. They were also different to what others had reported for skin. It should be noted that in the study by Naffa [144], it was found that the analyses carried out on skins from sheep, goat, deer and cow contained 40%, 28%, 27% and 21 % non-collagenous material that included other proteins. Thus comparison with this study is not really valid. While reduction in the numbers of Pro per 1000 residues can be rationalised due to its PTM into hydroxyproline which was increased, the other differences are difficult to explain due to the fact that every effort was made to isolate pure collagen I as evidenced by SDS PAGE (Figure 13). This is a key difference between this and other studies on collagen I. The most likely cause for these differences is a lack of conistency with either the

extraction of the collagen, the hydrolysis or the labelling. As all extracts were subjected to SDS PAGE to check the purity of the preparation it is difficult to attribute the differences seen to contamination as from the gels, the preparation was estimated to be at least 95% pure. It is therefore highly unlikely that these results are due to contamination by other proteins. It is possible however that incomplete hydrolysis could result in the variations seen. In this preparation samples were hydrolysed in 6M HCL, 3 % phenol at 115 °C for 24 hours. The solution was not gassed and covered with nitrogen as is traditionally done. However, the method used here has been successfully deployed by others in our laboratory [144]. Traditionally acid hydrolysis slowly destroys Ser and Thr, and Val and Ile bonds are hard to break, leading to their under-estimation. In this study however, Val, His and Arg appeared to be over estimated suggesting contamination rather than incomplete hydrolysis. Asp on the other hand was most likely understimated. These anaomalies will all affect each other and the totals and ratios obtained. The fact that tyrosine was seen at 2 fold the number in the pepsin extraction as in the acid extraction makes sense as Tyr is only found in the telopeptide region of collagen, however, this means that in the pepsin extraction Tyr should not be seen at all, as the pepsin extracted collagen is presumed to be lacking the telopeptide regions. The last obvious cause of the inconsistancies in the results is the labelling of the amino acids with AQC. This could occur through a lack of reagent, breakdown of reagent, or lack of solubility of the labelled amino acids. There was nothing to suggest that any of these possibilities had occurred. Firstly, the presence of an AMQ peak was indicative of excess reagent. Secondly, there was no buildup of pressure during the runs, indicative of insoluble particles blocking the column. Thirdly, when individual amino acids were labeled , they eluted with the expected retention times overlaying with the standard peaks. No internal norleucine standard was used prior to hydrolysis however to check for losses as a result of this relatively destructive method. The addition of this step to future extractions would allow corrections for losses due to hydrolysis.

One other question asked in this study was is the ratio of proline to hydroxyproline 1:1 in all skin samples, and does it always constitute 13.5 % of total amino acids? These questions impact on the traditional method for calculating collagen concentration in a sample. The results show that there are differences in the Hyp concentration between animals although whether these are significant needs to be addressed.

The unequivocal proof that diastereomers of hydroxylysine occur naturally still remains to be verified despite best efforts being made to show that they do. Tantalising evidence was obtained to suggest they do exsist but this needs to be verified with mass spectrometry.

5. Cross-links and glycosylation

5.1 Introduction

Mass spectrometry allows the identification and characterisation of the compounds within a complex sample with high sensitivity. It can be used to determine the identity of compounds using their monoisotopic masses and comparing them to a known standard or database entry, to sequence peptides, and to identify and characterise post translational modifications (PTMs) on proteins. A mass spectrometer typically includes at its most basic level an ion source, a mass analyser, a detector and a computer. Ions are produced in the gas phase by the ion source, then either pass through the mass analyser to the detector or are selected for fragmentation by collision with highly charged gaseous molecules, which break the analyte into smaller, charged fragments that then pass into the mass analyser. The mass analyser then separates ions according to their mass to charge ratio (m/z), and the detector counts the number of ions for every m/z ratio (intensity). The data from the detector is then converted into a mass spectrum using proprietary software. There also needs to be a method to introduce the sample to the mass spectrometer. It is common for electrospray mass spectrometers to be coupled to a High Pressure Liquid Chromatography (HPLC) instrument.

The cross-link profile of collagen was analysed by first extracting acid- hydrolysed cross-links from the collagen hydrolysate using CF-11 column chromatography. They were then lyophilised, rehydrated in water and formic acid and separated on a silica hydride column prior to mass spectral analysis (see Section 5.2 for details). Cross-link standards which had been prepared by the method of Naffa *et al* [169] were be used to calculate the response factors for each of the seven different collagen cross-links in the prepared standard; the three divalent cross-links DHLNL HLNL, and LNL, and the four mature cross-links HHMD, HHL, PYD, and DPYD. The cross-links were identified by using mass spectrometry after separation through a silica hydride column. The same collagen extractions were also analysed for glycosylation of hydroxylysine residues by mass spectral analysis following in-solution tryptic digestion.

In the interest of brevity, throughout this chapter the collagen extracted in section 3 will be referred to based on the properties of the collagen molecule and its cross-links that were thought to be extracted. As pepsin-extracted collagen is presumed to lack the telopeptide regions, and acid- and salt-extracted collagen is presumed to have retained the telopeptides, all pepsin-extracted collagen will be referred to as non-telopeptide collagen, and acid- and salt-extracted collagen as telopeptide collagen. Both reduced and non-reduced, non-telopeptide collagen were extracted using pepsin, while telopeptide collagen was extracted from non-reduced skins using acid, and reduced skins using salt. This results in four categories of samples as illustrated in Figure 32.

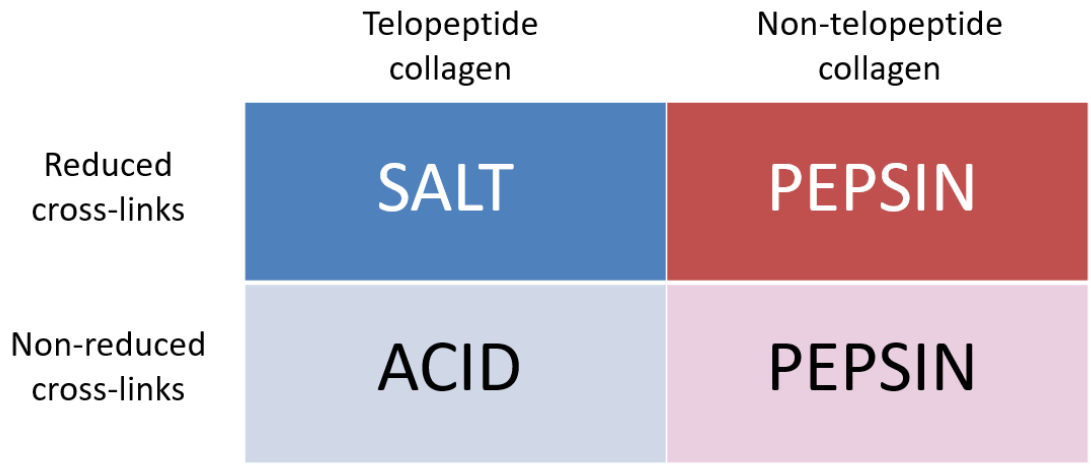


Figure 32. The four classifications of collagen subjected to cross-link analysis.

5.2 Methods

5.2.1 Hydrolysis of collagen

Cross-links were isolated from the collagen extracted from all four animals (cow, sheep, deer, and goat) using three different extraction methods on both reduced and non-reduced skins (see Section 3.2 for details). The purpose of reduction with sodium borohydride ($NaBH_4$) was to stabilise the acid-labile cross-links dihydroxylysinonorleucine (DHLNL), hydroxylysinonorleucine (HLNL), and histidinohydroxymerdesmosine (HHMD). The samples were hydrolysed as described in section 4.2.1.

5.2.2 Cross-link enrichment by CF-11 column

Cross link enrichment was carried out using the method of Naffa *et al* [175]. Fifty µL of hydrolysate from the step above was removed and stored at -80 °C for amino acid analysis. The remaining 950 µL was lyophilised to concentrate the solution and to remove acid prior to crosslink extraction and enrichment using CF-11 column chromatography. Ten mL plastic syringes were packed with glass wool, washed with one volume of MilliQ water and the glass wool compressed with the plunger to a final depth of ~1 cm. A 5% slurry of CF-11 fibrous cellulose powder (Whatman, Maidstone, United Kingdom) in 4:1:1 butanol (But), acetic acid (AcOH) and water (v/v/v) was used to fill the syringes (5 g in 100 mL). After the slurry had settled, it was washed 3 times with 5 mL But/AcOH/water. Lyophilised hydrolysate was rehydrated in 1 mL of MilliQ water, mixed into 2 mL of CF-11 slurry, then carefully loaded onto the top of the column so that the main bed was not disturbed and allowed to drain. The initial eluent was discarded, and the column was washed 10 times with 5 mL But/AcOH/water (4:1:1) to remove the free amino acids. The columns were then eluted with 3 x 5 mL volumes of water to give 15 mL eluent that contained the cross-links. This was then lyophilised and rehydrated in 200 µL mass spectrometry grade water (Fisher Chemical, Belgium), and stored at -20 °C prior to cross-link separation and quantitation by mass spectrometry.

5.2.3 Cross-link separation and quantitation

Chromatography was carried out using a Cogent Diamond Hydride™ HPLC column (150 mm × 2.1 mm; particle size, 2.2 µm; pore size 120 Å) with a Cogent diamond hydride guard column (20 mm × 2.1 mm; particle size, 2.2 µm; pore size, 100 Å) purchased from Microsolv Technology, (Leland, NC, USA) [175]. All samples were filtered through a 0.2 µm ReliaPrepTM syringe filter (Bärenstein, Germany) before they were applied to the column. Sample analysis was carried out in positive mode and MS scans (100–600 m/z) were obtained with resolution of 70, 000. Ionisation energy was set to 3.3 kV, sheath gas pressure of 35, and auxiliary gas pressure of 8. S lens was set to 50, capillary temp was set to 325 °C and auxiliary gas temp was 275 °C.

5.2.4 Crosslink separation

The crosslinks were separated using isocratic elution, with 70% Solvent A (0.1% formic acid in water (v/v)) and 30% solvent B (0.1% formic acid in 100 % methanol (v/v)). A flow rate of 400 µL min^{-1} and column temperature of 20 °C were used. The total run time was 5 minutes and a flow rate of 400 µL/ml was used.

5.2.5 Software & instrumentation

5.2.5.1 Cross-link analysis

The mass spectrometer used was a QExactive Focus liquid chromatography mass spectrometry (LC/MS) system (Thermo Fisher Scientific, San Jose, CA) with a heated electrospray ionization (HESI) source. The HPLC was a Dionex UltiMate™ LPG-3400RS Rapid Separation system with a Quaternary Pump (Thermo Fisher Scientific, San Jose, CA). Thermo Scientific Freestyle™ software version 1.3.115.19 (Thermo Finnigan LLC, San Jose, CA) was used for data visualisation and quantitation was conducted using Thermo Scientific Xcalibur software version 4.1.31.9 (Thermo Finnigan LLC, San Jose, CA).

5.2.5.2 Collagen PTM analysis

The mass spectrometer used was a QExactive Plus and Proteome Discoverer version 2.2.0.388 was used for data visualisation and quantitation (Thermo Fisher Scientific, San Jose, CA). The Percolator node was used for non-parametric estimation of q-values and posterior error probabilities, and all peptides used for PTM analysis were assigned a high confidence level. Modifications were determined by label-free peptide spectral matching (PSM). Peptides containing modifications matching the masses of mono-glycosylation with galactose, di-glycosylation with glucosyl-galactose, and hydroxylation of proline or lysine were identified by the software, and then the residues were manually identified from each peptide for each animal and technical replicate. These were used to identify the percent prolyl hydroxylation in Figure 39, and the glycosylation map in Figure 38. Label-free peptide spectral matching was also used to determine the relative ratios of the collagen proteins found in the samples in Figure 40.

5.2.6 In-solution tryptic digestion of collagen

One mg of each sample of collagen was dissolved in 50 µL of 0.1 M acetic acid, then diluted to 500 µL with 20 mM Tris-HCl, pH 7.8. Trypsin (proteomics grade; Sigma-Aldrich, NZ) was added to 20 µL of each sample at a concentration of 1:40 enzyme/protein ratio (v/w) and incubated overnight at 37 °C.

5.3 Results and discussion

5.3.1 Mass spectral identification of collagen type I cross-links

The LC/MS total ion-current profiles of seven of the cross-links of collagen are shown in Figure 33. All crosslinks were eluted from the column in 3 minutes with a peak width less than 1 minute. The total area of each cross-link profile was used to determine a response factor for that cross-link, in the same manner as for amino acid analysis in Section 4.2. These response factors were then used to determine the concentration of each cross-link in the experimental samples. Concentrations were related to the concentration of collagen (µmoles of crosslink per µmol collagen), which was calculated by determining the collagen amount from hydroxyproline as found in section 4. The results are summarised in Table 14.

All skin samples contained DHLNL, HLNL, HHL, and HHMD, although they were present in different ratios (Table 14). LNL, DPYR and PYR were also found in all samples, but only in trace amounts (Table 14 and Figure 33). Cow skin had the highest total crosslink concentration in all extractions except for the salt extraction, which produced higher cross-link concentrations from deer skin collagen. Sheep skin collagen had the lowest total cross-link concentration in samples produced using every extraction method except the non-reduced pepsin extraction Pepsin extraction was least efficient for goat hide which, interestingly, was the only skin sample that produced a higher concentration of total cross-links from reduced skin compared to non-reduced skins. Collagen lacking its telopeptides (NT), extracted from reduced cow, sheep and deer skins, had a significantly lower ratio of total cross-link concentration compared to collagen extracted from skins that weren't reduced (ratios of 0.24, 0.09 and 0.08 respectively). Conversely, all animals showed a 3-5 % increase in total cross-link concentrations in reduced *versus* non-reduced telopeptide-containing collagen (T, salt and acid extractions respectively). Robins *et al* found that increases in the concentration

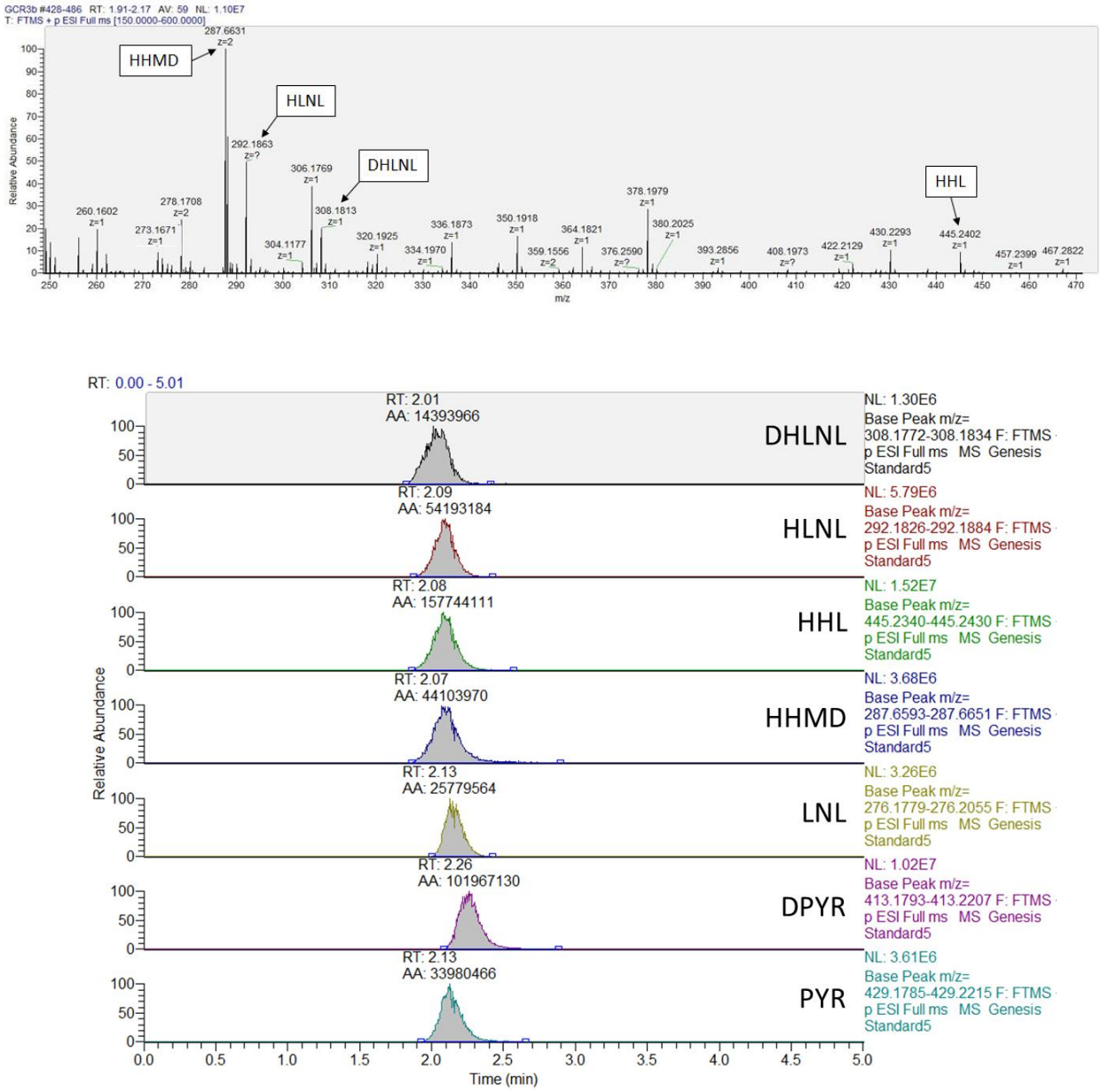
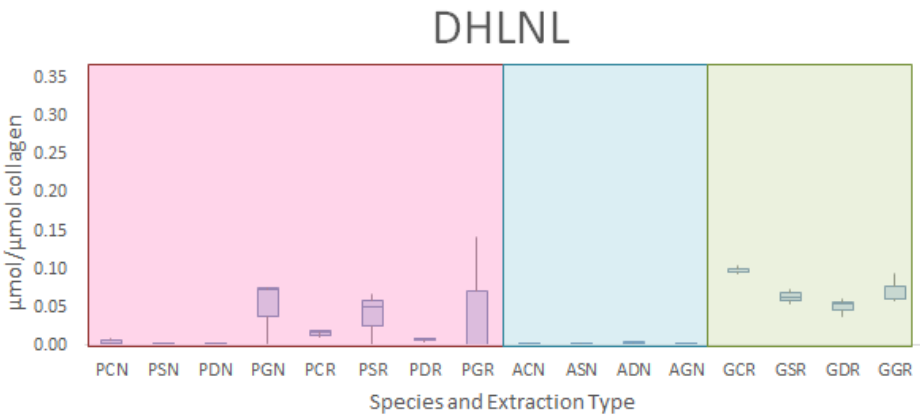


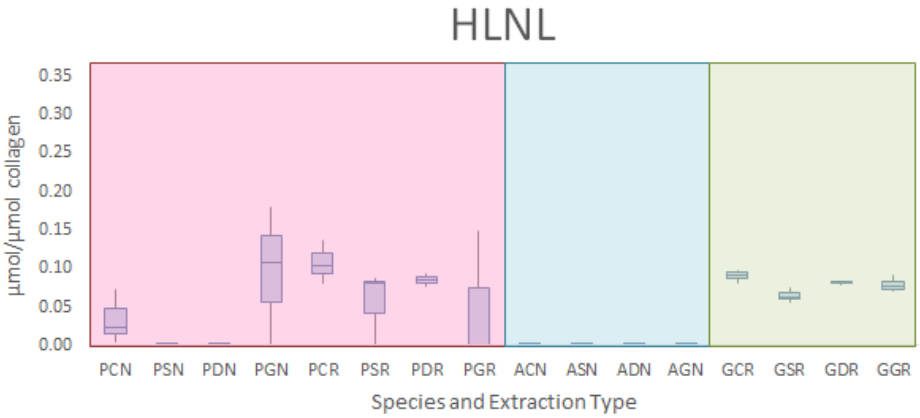
Figure 33. Mass spectral identification the standards of seven cross-link types found in collagen type I; DHLNL (dihydroxylysinonorleucine), HLNL (hydroxylysinonorleucine), HHL (histidinolysinonorleucine), HHMD (histidinohydroxymerdesmosine), LNL (lysinonorleucine), DPYR (deoxypyridinoline), and PYR (pyridinoline).

Table 14. µmol cross-links per µmol collagen from cow, sheep, deer, and goat skins. NT- before the species name denotes non-telopeptide collagen extracted with pepsin. T- denotes telopeptide containing collagen extracted using either salt (reduced cross-links) or acid (non-reduced cross-links). Cross-links names in the first column that are preceded by R- have been reduced with sodium borohydride. µmol/ µmol collagen was calculated using the values for the molarity of collagen in each sample calculated in section 4 determined by the hydroxyproline content.

		Cross-links µmol/µmol collagen															
		Non-telopeptide containing collagen								*Telopeptide containing collagen*							
		Cow		Sheep		Deer		Goat		Cow		Sheep		Deer		Goat	
Cross-links		*NT-Cow*	*%mol*	*NT-Sheep*	*%mol*	*NT-Deer*	*%mol*	*NT-Goat*	*%mol*	*T-Cow*	*%mol*	*T-Sheep*	*%mol*	*T-Deer*	*%mol*	*T-Goat*	*%mol*	*Average*
HHL	*Non-reduced*	0.13	21.74	0.10	94.22	0.14	96.47	0.04	5.98	0.23	90.56	0.12	94.47	0.25	95.24	0.16	90.79	*33.32*
	Reduced	0.14	6.46	0.08	6.48	0.15	8.13	0.04	8.28	0.06	9.17	0.04	9.25	0.11	8.40	0.09	17.01	*3.79*
HHMD	*Non-reduced*	0.41	71.49	4.1E-03	3.77	3.6E-03	2.48	0.52	73.10	0.02	8.01	4.71E-03	3.68	0.01	3.54	0.01	7.79	*11.14*
	Reduced	1.97	70.00	0.96	74.58	1.63	80.74	0.30	26.18	0.43	45.99	0.24	45.14	1.05	67.72	0.29	41.29	*27.82*
HLNL	*Non-reduced*	0.03	5.90	9.6E-04	0.88	3.8E-04	0.26	0.10	13.72	1.3E-03	0.51	0.00	0.79	1.0E-03	0.40	9.4E-04	0.54	*1.51*
	Reduced	0.11	3.81	0.08	6.49	0.08	4.18	0.05	4.43	0.09	9.71	0.06	12.10	0.08	5.24	0.08	11.22	*3.11*
DHLNL	*Non-reduced*	4.9E-03	0.85	1.0E-03	0.95	1.1E-03	0.77	0.05	7.16	2.2E-03	0.87	1.3E-03	1.00	2.1E-03	0.80	1.4E-03	0.80	*0.83*
	Reduced	0.02	0.55	0.06	4.57	0.01	0.34	0.05	4.19	0.10	10.59	0.06	11.85	0.05	3.29	0.07	10.17	*2.39*
PYR	*Non-reduced*	1.4E-04	0.02	1.7E-04	0.16	2.4E-05	0.02	1.6E-04	0.02	1.3E-04	0.05	7.4E-05	0.06	5.0E-05	0.02	1.1E-04	0.06	*0.02*
	Reduced	1.3E-04	4.6E-03	2.2E-04	0.02	3.1E-05	1.5E-03	2.0E-04	0.02	6.9E-05	0.01	8.8E-05	0.02	3.3E-05	2.2E-03	9.8E-05	0.01	*4.5E-03*
LNL	*Non-reduced*	1.5E-05	2.5E-03	5.8E-06	0.01	6.4E-06	4.4E-03	5.9E-05	0.01	3.0E-06	1.2E-03	6.0E-06	4.7E-03	3.2E-06	1.2E-03	3.6E-06	2.1E-03	*1.8E-03*
	Reduced	5.5E-05	2.0E-03	4.7E-05	0.00	5.5E-05	2.7E-03	4.7E-05	4.1E-03	1.0E-04	0.01	4.4E-05	0.01	9.3E-05	0.01	5.3E-05	0.01	*2.6E-03*
DPYR	*Non-reduced*	2.7E-06	4.7E-04	9.8E-06	0.01	1.9E-06	1.3E-03	1.4E-05	2.0E-03	2.6E-06	1.0E-03	4.7E-06	3.7E-03	3.7E-06	1.4E-03	8.7E-06	0.01	*1.3E-03*
	Reduced	3.1E-06	1.1E-04	1.3E-05	1.0E-03	4.4E-07	2.2E-05	1.6E-05	1.4E-03	2.0E-06	2.2E-04	3.7E-06	6.8E-04	7.1E-07	4.6E-05	8.1E-06	1.1E-03	*2.4E-04*
N-TOTAL		0.58		0.11		0.15		0.71		0.25		0.13		0.26		0.17		*0.29*
R-TOTAL		2.23		1.18		1.88		0.43		0.68		0.41		1.29		0.53		*1.08*
Total		2.81		1.29		2.02		1.14		0.93		0.54		1.55		0.70		*1.37*

	$p < 0.05$





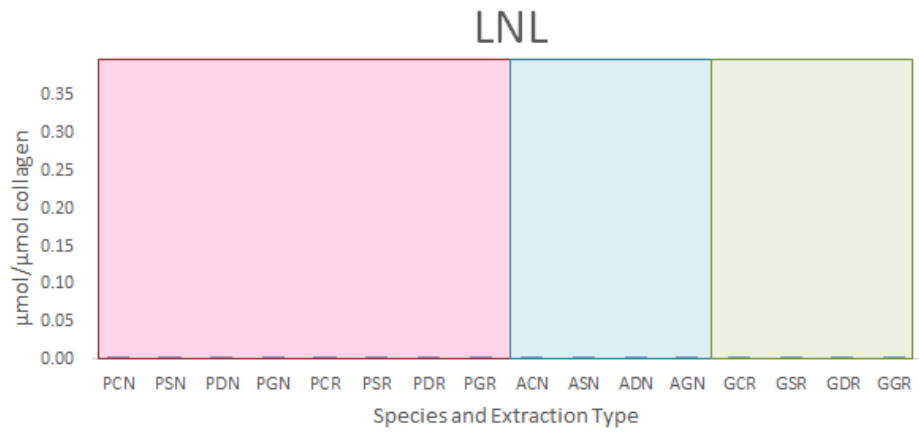


Figure 34. Comparison of the molar percent of each type of intermediate, reducible cross link across species and extraction type. Pink boxes denote pepsin-extracted collagen, blue boxes are acid-extracted collagen, and green boxes are salt-extracted collagen. Data labels are as follows: the first letter denotes extraction type, the second letter is species (C: Cow, S: Sheep, D: Deer, G: Goat); the third letter denotes whether the skins were reduced with $NaBH_4$ (R) or non-reduced (N).

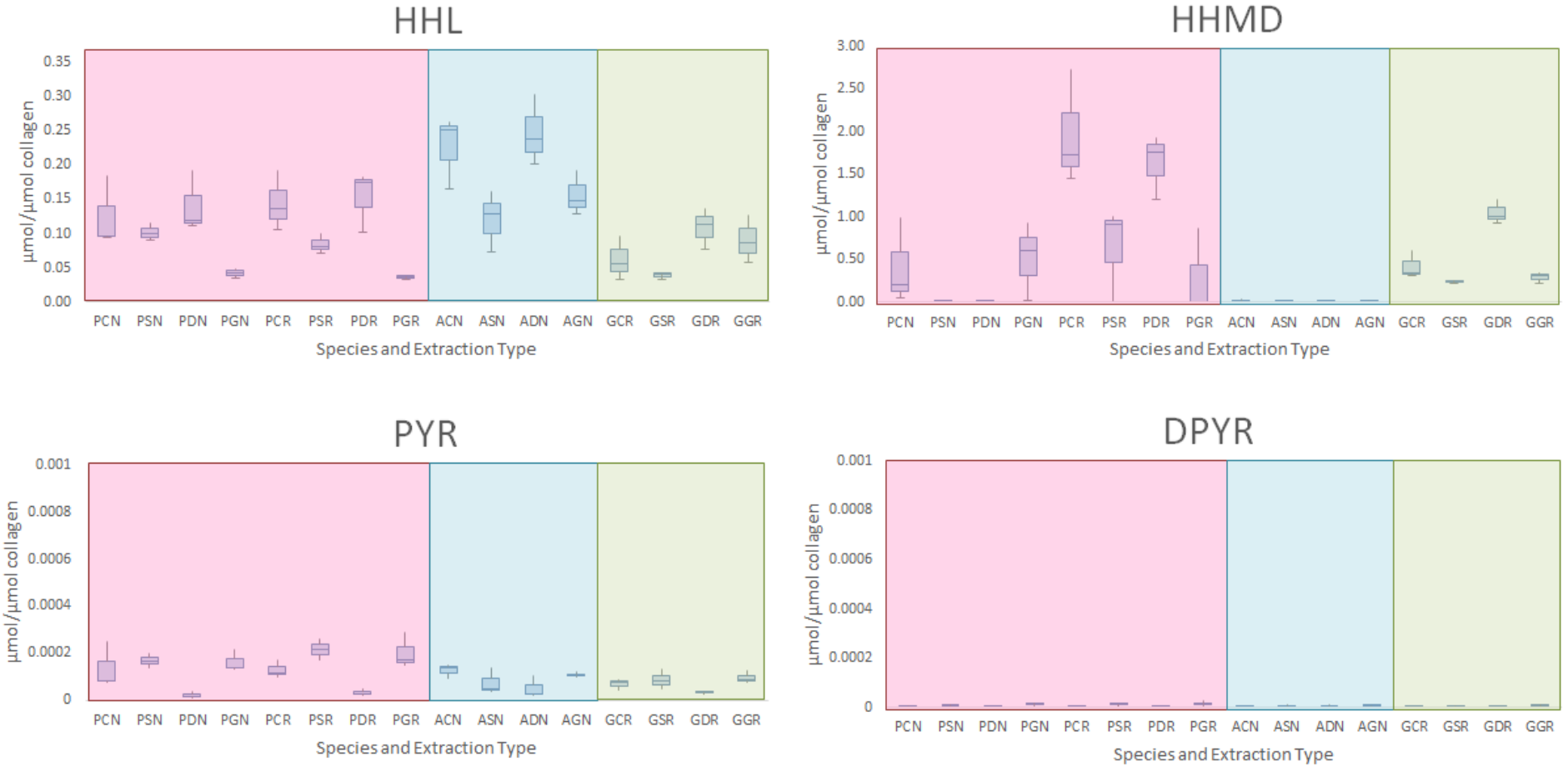


Figure 35. Comparison of the molar percent of each type of mature cross link across species and extraction type. Pink boxes denote pepsin-extracted collagen, blue boxes are acid-extracted collagen, and green boxes are salt-extracted collagen. Data labels are as follows: the first letter denotes extraction type, the second letter is species (C: Cow, S: Sheep, D: Deer, G: Goat); the third letter denotes whether the skins were reduced with $NaBH_4$ (R) or non-reduced (N).

of the histidine adduct cross-link, HHL, could be correlated with increased resistance to solubilisation of the skin using pepsin [176]. If this is correct, the concentrations should be lower in reduced samples where intact crosslinks are likely to make collagen extraction less efficient. This study, however, found that collagen extracted from reduced sheep, deer and cow skins using pepsin contained the highest concentrations of HHL (an average of > 90 % of total crosslinks). On the other hand, pepsin-extracted goatskin contained the lowest concentration of HHL in both reduced and non-reduced samples, (0.43 μmol/ μmol collagen). A study published early this year by Eyre *et al* [177] suggested that HHL was an artefact produced as a result of acid hydrolysis by the reaction of a C-telopeptide allysine aldol and glcgalHyl at α1(I)K87. While the higher concentrations of HHL in telopeptide containing collagen compared with non-telopeptide containing collagen in goat and cow skins would fit with this theory, in both telopeptide and non-telopeptide containing sheep and deer collagen HHL makes up ~95 % of the total crosslinks. As both the residues thought to be implicated in the production of HHL as an artefact are located in the telopeptide regions, non-telopeptide containing collagen should not contain high levels of this cross-link. The fact that they do suggests that acid hydrolysis is not responsible for HHL formation, and that it does indeed exist *in vivo*.

The immature cross-link LNL was only present in trace amounts in all samples but both DHLNL and HLNL were measured at ~2 % mol/mol in all non-reduced, telopeptide containing collagen samples (acid extracted). As only monoisotopic masses of the reduced forms of DHLNL and HLNL were monitored, this is surprising, as these samples were extracted from non-reduced skin. Therefore, the origin of these peaks is unclear. It is unlikely that contamination is the source, as the reduced and non-reduced samples were prepared from separate pieces of skin that were at no stage in contact with one another. However, HLNL is found in even higher quantities in non-reduced, non-telopeptide collagen (pepsin extracted) at ~6 % of the total cross-links found in in cow skins and ~14 % of the total cross-links in goatskins which also contain ~7 % DHLNL. Collagen extracted with pepsin from the skins of these animals contains masses matching the reduced form of immature cross-links, whereas acid

extracted collagen from these same skins does not. As the masses were not fragmented further with MS/MS, it is likely they belong to some other molecule with a similar mass.

Further fragmentation is absolutely necessary in future work to confirm the identity of these masses. Schnider *et al* found that acid soluble collagen contains predominantly newly synthesised procollagen molecules (as discussed in section 4.3.6), which contain higher concentrations of immature cross-links. However, this does not agree with the results found here, as the acid extractions contained the mature cross-links HHMD at ~3-8 % mole percent, and HHL at > 90 % mol/mol. In fact, over all species and extraction types, HHL was present in the highest concentration in the acid extractions. Salt extraction was the only extraction type to produce low mole % of HHL across all species. Pepsin extraction of skins across the different species showed varied concentrations of HHL in the different species (Table 14 and Figure 35). The ratio of mature cross-links (HHMD+HHL) to immature cross-links (HLNL+DHLNL) was determined and is shown in Figure 36. In all collagen samples except for those isolated from NT-Cow and NT-Goat the ratio of mature to immature cross-links is increased in samples from non-reduced skins compared with those from reduced skins. Collagen isolated from deer and sheep skins both contain similar ratios between non-telopeptide (NT) and T collagens in both reduced and non-reduced samples.

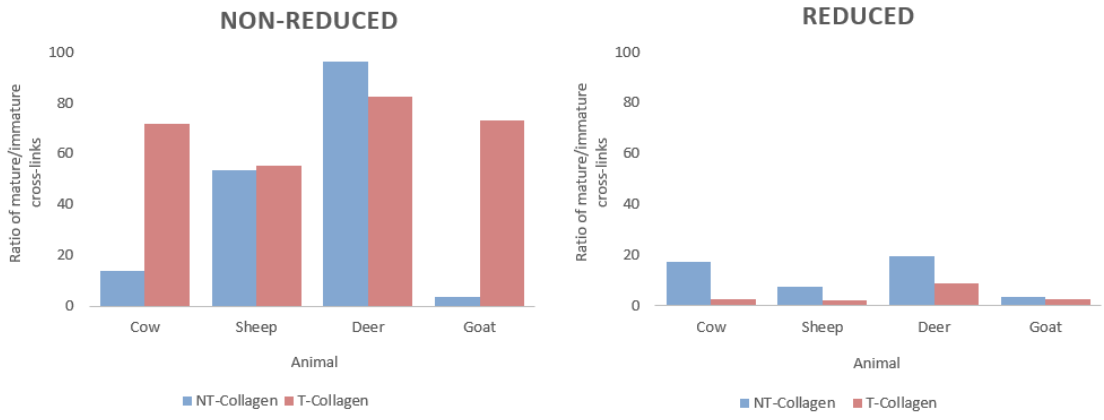


Figure 36. Ratio of mature cross-links (HLNL+DHLNL) to immature cross-links (HHMD+HHL) across all four species in reduced skins and non-reduced skins.

There is a general trend towards reduced skins containing higher ratios of mature cross-links in NT versus T collagen, while the opposite is seen in non-reduced skins, with the exception of Deer, which shows higher ratios of mature cross-links in all collagens including NT collagen whether reduced or not. While goat had the lowest ratio of mature to non-mature crosslinks in non-telopeptide containing collagen sheep had the lowest ratio in telopeptide-containing collagen.

Although these results are interesting, the coefficient of variation (CV%) varies greatly between them, with many samples CVs well over 10 % (Table A. 2). A one-way ANOVA (analysis of variance) followed by a post-hoc Tukey's HSD (honestly significant difference) was therefore used to interpret significance (see accompanying spreadsheet '*crosslink analysis.xlsx'*). Goat collagen had a significantly higher concentration of HHL in both reduced and non-reduced NT collagen ($p < 0.05$), and HHMD in non-reduced NT samples ($p < 0.05$). Sheep had significantly lower concentrations of both HLNL and HHL cross-links ($p < 0.05$) in all telopeptide containing collagen and had the lowest ratio of mature to immature crosslinks in samples taken from telopeptide-containing collagen, indicative of a younger animal. Samples taken from sheep and goat collagen also contained low concentrations of HHMD and DHLNL and the lowest total cross-link concentration over all extractions 1.83 and 1.84 μmoles per μmol collagen, (Table 14). Conversely, cow and deer collagen contained approximately twice this concentration at 3.74 and 3.57 μmoles/ μmol respectively. These results concur with the work of Naffa *et al* [144] which showed cow to have the highest total cross-link concentration and sheep the lowest.

5.3.2 Glycosylation of collagen type I

Glycosylation of acid extracted collagen type I was analysed using mass spectrometry as described in section 5.2. These samples contained collagen with the telopeptide regions intact. From the map in Figure 38 it is clear that the α1 chain is more glycosylated than the α2 chain, with 19 (1.8 % of the 1053 α1 chain amino acids) glycosylation sites identified compared with 8 (0.8 % of the 1037 α2 chain amino acids) respectively.

Monoglycosylation is predominant in both the α1 and α2 chains, with an average ratio of 5 times as much galactosylation to glucosylgalactosylation in the α1 chain and 3 times in the α2 chain (Figure 37). Glycosylation across each chain also appears to be conserved to specific regions which do not overlap between chains. This gives the appearance of a "lock and key" pattern between the chains, where the glycosylation-rich regions of one chain correspond with the glycosylation-poor regions of the other chain, and *vice versa*.

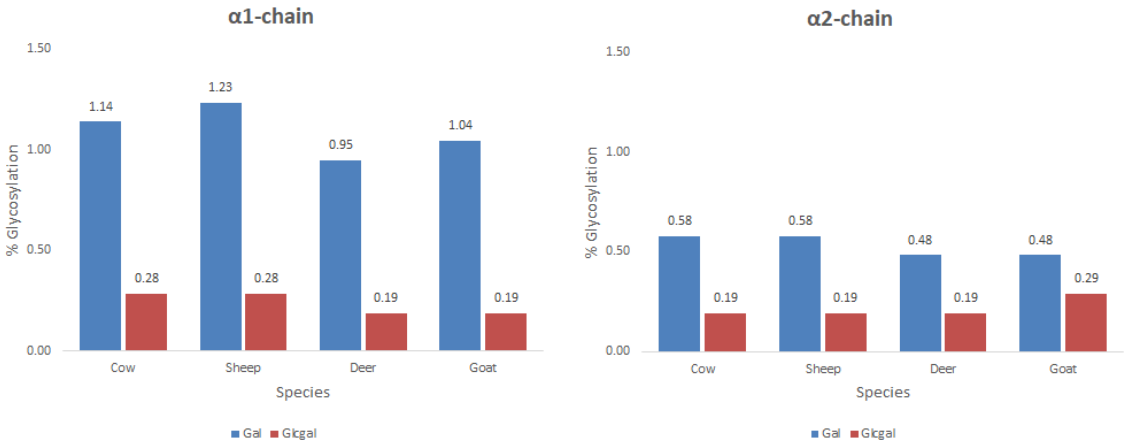


Figure 37. Percent glycosylation of collagen α chains across four different species. Blue bars denote galactosylation (Gal), red bars denote glucosylgalactosylation (Glcgal).

Table 15. Comparison of the FASTA sequence for Col1A (bovine) against the number of hydroxylysine residues found in chapter 4 and the number of glycosylated residues detected by mass spectrometry. Calculated based on glycosylated sites that were detected in ≥2 out of 3 technical replicates.

Hyl/glycosylation comparison	
	Number of residues
Lys	15
Hyl	11
Total experimental	26
Total FASTA	36
Theoretical available Hyl residues for glycosylation (FASTA -Total)	10
Average glycosylation of α1 + α2	11

When considered in the context of collagen biosynthesis this is interesting. Glycosylation occurs in the ER concomitantly with the formation of the triple helix, and that once the triple helix is fully formed, no further glycosylation takes place. It is possible therefore, that this "lock and key" arrangement of the two α-chain glycosylations may somehow direct the alignment of the α2 chain to the α1 chains, although the two collagen α1(I) chains are tethered to one another by a disulfide bond in their C-propeptides, mutational studies have shown that the cysteine residues in the C-propeptide of the α2(I) chain do not form inter-chain disulfide bonds. No evidence could be found explaining how the α2(I) chain is aligned with the α1(I) chains during the formation of the triple helix, indicating that further investigation of any possible interaction between the α-chains and the glycosylated residues of adjacent chains is required.

Interestingly, glycosylation of K1021 in α2 was detected in all four animals with cow, deer and goat having this modification in all three replicates and sheep having it in two. This residue falls outside of the predicted telopeptide-containing α2 chain region and therefore should not be present in collagen after the propeptide regions have been cleaved shortly after synthesis, suggesting that newly synthesised collagen must be present in all of the samples, although without quantitation it is impossible to know at what concentration. If present at high concentrations, it may have an influence on amino acid determination, as this region has a different amino acid composition and additional modifications such as disulfide bridges that are not usually present in collagen. Since little is currently known about these regions, they may contain other modifications as well, and as they constitute 384 residues (approximately a quarter of the total residues of procollagen) they could have a large impact on the results. The presence of this region in all samples reinforces the suggestion in chapter 4 that acid extraction results in a higher concentration of newly synthesised collagen. In future work it would be interesting to analyse the residues detected in the pepsin and salt-extracted collagen as well.

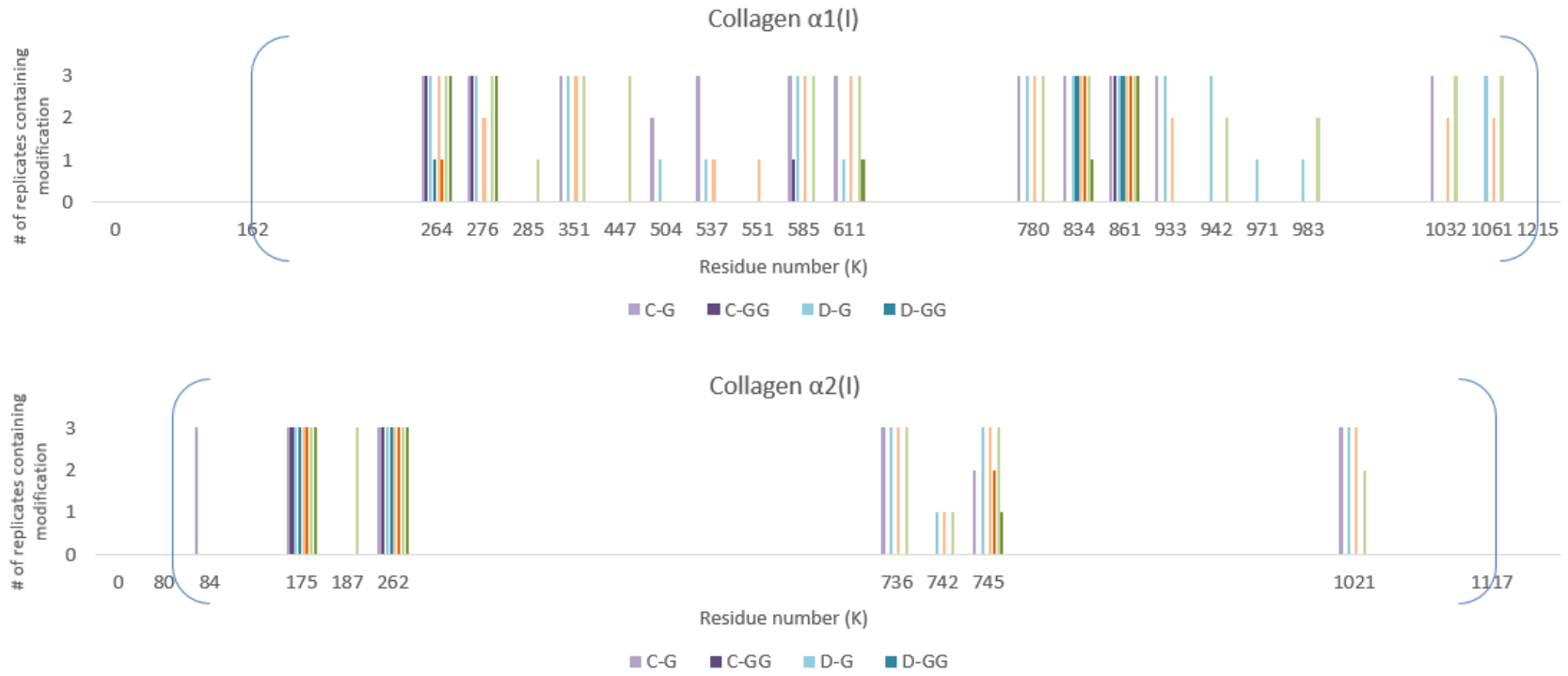


Figure 38. Map showing glycosylation of hydroxylysine residues in collagen α1(I) and α2(I).The left axis shows the total number of technical replicates (n=3) that the modification was found on. The first letter of each bar label relates to the animal (c, d, g, s are cow, deer, goat and sheep, respectively), the second letter/s relate to the sugar attachment (-G is galactosyl, -GG is glucosylgalactosyl). Blue brackets represent the borders of the telopeptide containing collagen molecule, areas outside of the brackets belong to the propeptide and signal peptide regions, which are cleaved during/after secretion from the cell.

5.3.3 Glycosylation differences between species

See *glycosylation analysis.xlsx* on supplementary compact disc for calculations relating to glycosylation analysis between species. Sheep collagen contained the highest number of glycosylated residues with 19 residues in α1 and 10 residues in α2. Cow collagen had the least with 16 and 7 respectively, followed closely by goat which had one extra glycosylated residue in the α2 chain. Several residues were glycosylated with consensus across all species and technical replicates. In α1 these were K264, K351, K585, K780, and K834, which were all galactosylated, and K861 which could be either galactosylated or glucosylgalactosylated in all species across all technical replicates. K276 was also galactosylated in all animals and in nearly all replicates with the exception of goat which lacked the modification in one replicate. K264 was also glucosylgalactosylated in all technical replicates in cow and sheep, but only one replicate in deer and goat. In collagen α2, residue K736 was galactosylated across all animals and replicates, while K175 and K262 could be either galactosylated or glucosylgalactosylated.

Interestingly, in collagen α2, K84 was galactosylated in all three collagen extracts from cow skin replicates, but not in any other animal. This residue is very close to the N-terminus of the telopeptide region. Glycosylation in the telopeptide regions has been implicated in controlling fibrillogenesis and in particular directing the quarter-staggered arrangement of collagen molecules. It is possible that this unique glycosylation in cow collagen may influence the direction of the collagen molecules influencing the mostly vertical arrangement of the collagen fibrils as found by Naffa *et al* [144].

5.3.4 Prolyl hydroxylation

To determine the accuracy of the high molar percentages of hydroxyproline detected by amino acid analysis in chapter 4, the mass spectral data were also analysed for hydroxyl modifications to proline residues. The results are summarised in Figure 39.

The collagen α1(I) chain appears to contain the highest percentage of proline hydroxylation, and this is conserved across all four species, ranging from 68.9-72.1 % of proline residues. Hydroxylation of the α2(I) chain, however, varies greatly between species, with the percent hydroxylated in cow and sheep similar at 55.8 % and 53.4 % of total proline

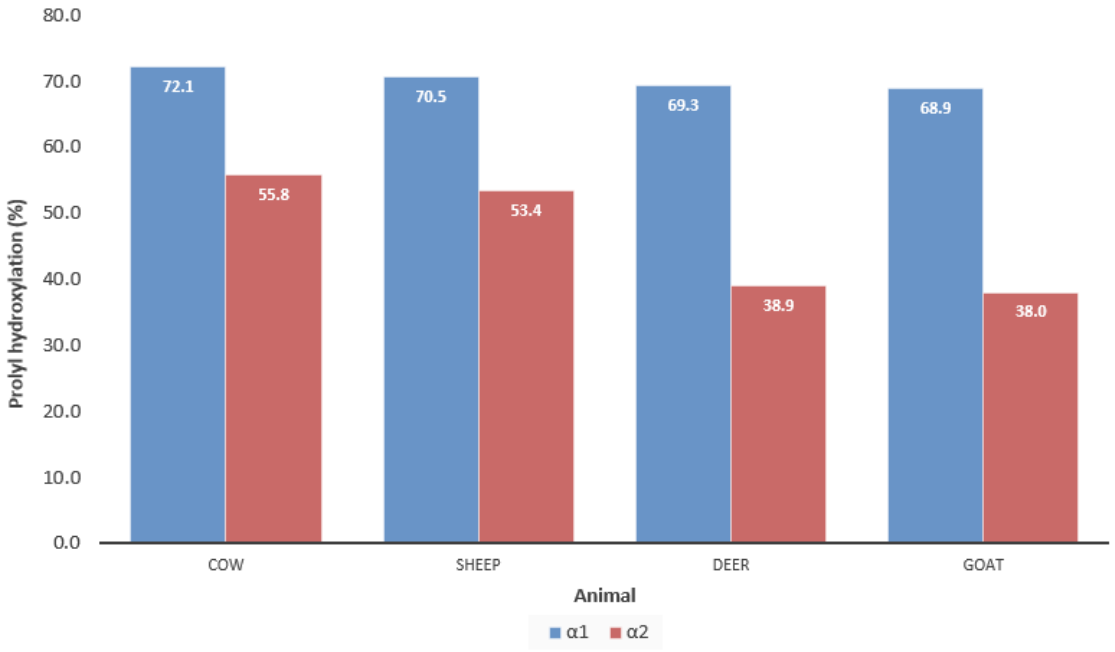


Figure 39. Percent of proline residues in collagen type I that are detected by mass spectrometry as modified by a hydroxy group (hydroxyproline). Blue columns denote proline residues originating from the collagen α1(I) chain and red columns denote prolines from the collagen α2(I) chain.

residues respectively, and deer and goat showing a much lower percent of hydroxylation at 38.9 % and 38 % of total prolines respectively. This puts the average percent of glycosylation across the tropocollagen molecule in cow at 66.6 % (or two-thirds of all proline residues which is much higher than the 1:1 ratio of proline to hydroxyproline assumed in most collagen reviews and textbooks.

This may provide an explanation for the high molar percentage of hydroxyproline and the low percentage of proline determined by amino acid analysis in section 4.3. If proline is hydroxylated at the percentages determined here, the standard equation for calculating collagen concentration devised by Neuman [160] which has been in practice since 1950 needs to be questioned. If hydroxylation of proline is truly greater than the 13.4 % molar percent which has been accepted for the past 70 years, and, according to the results of this study vary widely between species and extraction methods, then this method of calculating collagen concentration should no longer be used universally and a better method of collagen concentration determination needs to be devised.

5.3.5 Ratios of collagen types and α-chains

The mass spectral data were also analysed to determine the relative ratios of both collagen types and α-chains in each animal. The results are summarised in Figure 40. It is immediately apparent that while 10 different types of collagen/collagen α-chains were detected, only three collagens dominate in the samples of all four animals; collagen type I (α1), collagen type I (α2), and collagen type III (α1). All other collagens were found in only trace amounts. Interestingly, collagen type I (α2) is the predominant protein in all samples and animals, ranging from 63 % in cow skin to 82 % in goatskin, and in sheep and deer at similar percentages of 78 % and 79 % respectively. Astonishingly, collagen type III was detected at abundances of an almost equal percentage to collagen α1(I). This brings into question the assumed purity of the SDS-PAGE results in Section 3.3, as the collagen α1(III) chain is known to migrate to approximately the same region of the gel as collagen α1(I). These results are still surprising, however, as they imply that collagen α2 (I) is present at over twice the concentration of the α1 chains of collagen I and III combined, and yet when the SDS-PAGE of the extractions is seen in Figure 13 (Section 3.3), the bands show the

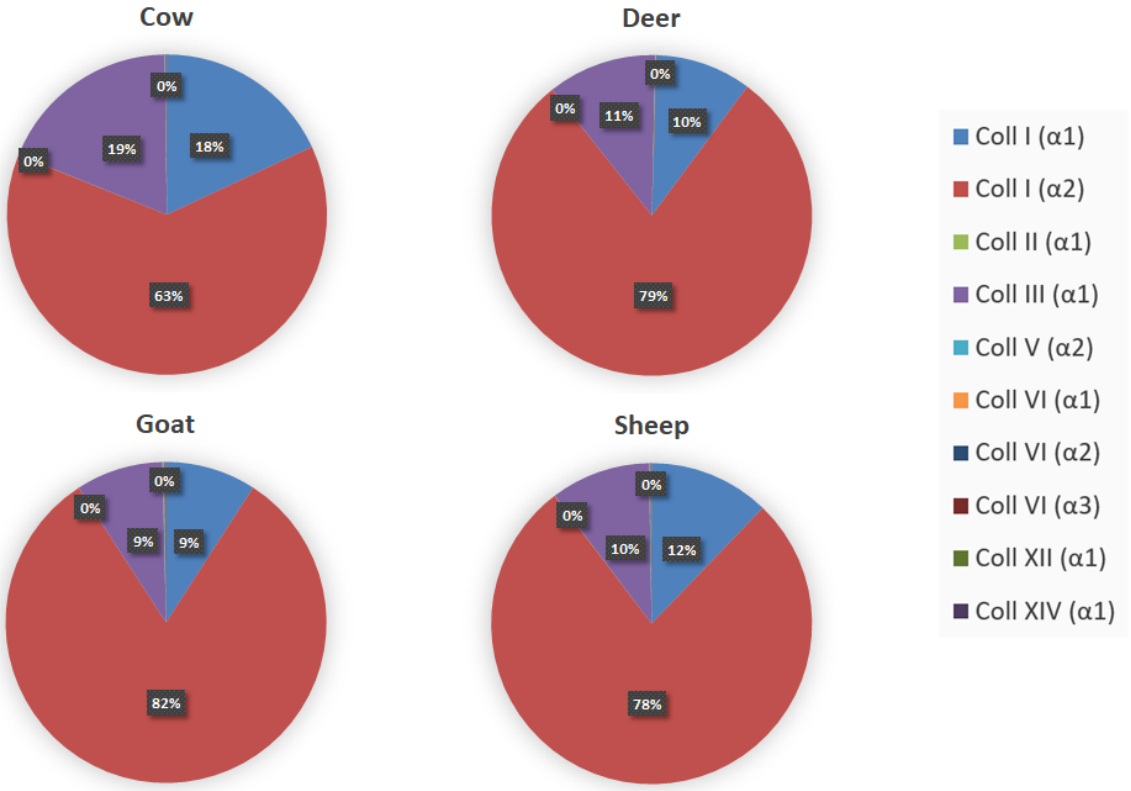


Figure 40. Percent of collagen types (chains) across species as determined by mass spectroscopy. All proteins were detected with high confidence (low false detection rate (FDR)).

typical pattern of a denser higher molecular weight band belonging to the α1(I) chain, and a less dense band at a lower molecular weight belonging to the α2(I) chain. It would be interesting in future work to cut these bands for in-gel tryptic digestion and analysis by mass spectrometry to confirm their identity.

In total, 180 proteins with greater than 2 unique peptides and FDR < 1 % and identified with high confidence were detected across the 4 species. While all of these other than the three collagen chains identified above were detected in only trace amounts, these, combined with the influence of the large contribution of collagen type III may have led to the unusual results seen in the amino acid analysis in section 4.3. A multiple sequence alignment was carried out against collagen α1(I), α2(I), and α1(III) in ClustalX version 2 software (Figure A. 2), and the UniProt Clustal Omega align tool was used to compare collagen α1(I) and α1(III) for sequence homology. This showed a sequence identity of only 45.153 % between α1(I) and α1(III). With collagen III being detected in percentages from 63–82 %, these differences in homology may very well have led to the unusual amino acid results found in section 4.3.

5.4 Conclusion

In agreement with the literature, DHLNL and its subsequent mature cross-links Pyr and Dpyr were found in only trace amounts. These cross-links are much less common in tissues such as skin, as due to a lack of lysine hydroxylation in the telopeptide regions, favouring the formation of HLNL cross-links instead. HLNL cross-links form predominantly in skin and cornea between lysine in the telopeptide region of collagen and a helical-Hyl residue from a neighbouring collagen molecule. HLNL was found in low mole percent in all reduced skins except for telopeptide containing collagen isolated from goat skin where a concentration of 0.06 μmol/μmol which related to 12 mole percent was measured ($p < 0.05$). In other skins the mole percent of HLNL ranged from 3-11 %, however these varied greatly between replicates and were not statistically significant. These relatively low molar percentages of HLNL may be due to a high degree of conversion into their mature cross-link HHL. Studies have shown that HHL increases rapidly from birth and throughout the maturational stage, followed by a gradual but steady increase over the lifetime in both

human and cow skin collagen. As the skins used for this study were sourced from the meat industry the age of the animals is unfortunately unknown.

It is clear from Figure 36 that the general pattern is for higher ratios of mature collagen in non-reduced skins compared with reduced skins. This is consistent with the literature consensus that sodium borohydride reduction stabilises these immature, reducible cross-links that are otherwise acid-labile (and would therefore be destroyed during acid hydrolysis in the preparation of the cross-links for analysis). While HHMD is also reducible and is found in significantly higher concentrations in reduced non-telopeptide containing deer and telopeptide containing goatskins compared with non-reduced skins ($p<0.05$), there has been some recent controversy as to its origin. Interestingly, along with their conclusion that HHL was a product of acid hydrolysis Eyre *et al* also proposed that the mature cross-link HHMD is not a cross-link which occurs *in vivo* but is produced by Michael addition during sodium borohydride reduction and then released upon acid hydrolysis [177]. While HHMD is found in high concentrations in the reduced skins of all pepsin and salt-extracted concentrations and is present in low concentrations in all non-reduced, acid extractions, significant concentrations are found in pepsin extractions of both cow and goat skins which have not been reduced with sodium borohydride. In fact, HHMD is the predominant cross-link found in non-reduced goat skins. This is interesting, and brings into question again whether this cross-link is an artefact of preparation or does indeed exist *in vivo*. Non-reduced, enzymatically hydrolysed samples were prepared during the course of this study, but due to time constraints were not able to be subjected to cross-link analysis. These will be analysed and characterised in the future in an attempt to shed further light upon the source of both HHL and HHMD.

The results of glycosylation analysis by mass spectrometry revealed that in both α1 and α2 chains mono-galactosylation is predominant in all animals. It was also shown that cow appears to contain a unique glycosylation site at K84 on the α2 chain. While this was found in all three cow replicates and not found in any of the other animal replicates, it would be interesting in future work to analyse greater numbers of replicates of all four animals to determine if this is truly a modification unique to cow. It has been suggested that even one extra glycosylated residue per molecule can influence several steps in fibrillogenesis and

ultimately reduce the diameter of newly formed collagen fibrils [178, 179]. However, although cow was shown to have the second highest percent of glycosylation of both the α1 and α2 chains, it was shown using SAXS (small-angle x-ray scattering) [144] to have a collagen structure with the largest fibril diameter of these four animals which is at odds with this theory. However, it may be possible that the K84 glycosylation may affect cow fibril properties in a different manner, as this residue is located very close to the N-telopeptide terminus compared to other glycosylation sites. As skin cross-links are predominantly formed through the action of lysyl oxidase on lysines found in the N-telopeptides, and glycosylation occurs in the ER before the formation of cross-links in the extracellular space, there is a possibility that this sugar-attachment plays a role in directing cross-link formation. This possible role of collagen glycosylation has been suggested previously by several researchers [178, 180-183]. There is also a possibility that glycosylation serves to inhibit excess cross-linking of collagen molecules. As glycosylation of hydroxylysine residues occurs on the nascent procollagen molecule in the ER and triple helix formation is shown to inhibit further glycosylation, the cross-links characterised here cannot form on glycosylated residues.

There has also been some debate as to whether cross-linked hydroxylysine residues can be glycosylated themselves, which may also have an impact on directing cross-link formation. While this is outside of the scope of this project, future work may include the characterisation of these proposed glycosylated cross-link structures.

6. General summary and conclusions

6.1 Overall differences in collagen modifications between species and extraction types

Cow, sheep, deer, and goat are common domestic animals in New Zealand, as their meat is used to varying extents, sheep wool is a major export, and their skins are commonly used in the leather industry. Studies have shown that each animal's skin has different biomechanical properties that result in leather of different strengths; sheep leather for example is very soft, but unlike cow skin it has a low tensile strength and cannot be used to make bags or shoes. These differences occur even though the main component of the skin remains the same: collagen type I. While many studies have been conducted on differences in collagen between tissue types, and some research has been undertaken looking at differences between animals of different classes, very little is known about the differences in collagen type I from the same tissue type between animals of the same class (e.g. Mammals). Recent work in this laboratory has been undertaken to study the differences between total skin content of the four animals that are the subject of this study, and this revealed some intriguing differences between fibril organisation and structure [144]. However, the reason for this study was to take a closer look at collagen type I only, to determine if differences in its modification between animals existed and whether these may contribute to the higher structural differences as determined in the Naffa study [144].

The first part of this study involved determining the optimal method for purifying the collagen type I from the skins. The method needed to produce collagen type I to a level of confidence where this was the only collagen type visible on an SDS-PAGE gel. While other collagens may still be present the logic was that if these could not be visibly detected upon staining with Colloidal Coomassie Brilliant Blue G-250, which detects proteins to a minimum of 30 ng the contributions of these collagens to the results would be minimal. Of greater importance was the removal of other contaminating proteins such as elastin and proteoglycans, which are a significant feature of skin. After optimisation of extraction temperature and NaCl molarity to produce a high ratio of staining to bands consistent with the α- β- and γ- bands of collagen type I, and removal of the heating step which was

suspected to be causing breakdown to the collagen chains there were no visible bands present below the expected molecular weight of the collagen α2 chain.

Few differences could be seen in the number of bands on SDS-PAGE gel between the different animal skins and possible differences in molecular weight were difficult to decipher and may have been simply an effect of gel warping. Sheep appeared to produce collagen α-chain bands of higher intensity than the other animals suggesting a higher concentration of collagen may have been extracted, however this was not reflected in the final yields, in which cow skins consistently produced the highest yields of collagen across all extractions.

Amino acid analysis however, showed that while all animals pepsin extractions consisted of nearly pure collagen (determined using Neuman's method [160]), Deer had the highest concentration of collagen in acid extractions, followed by cow, sheep and then goat. This differs from the results found by Naffa *et al* [144] where it was determined that sheep had the lowest concentration of collagen and the cow the highest concentration, however it is important to note that their study was based on total skin samples rather than extracted collagen alone. These contrasting results, combined with the variations observed between amino acid profiles of skins produced by different groups indicates that differences in extraction methods, hydrolysis methods, derivitisation of amino acids, and other contributing factors such as the age of the animals probably has a large effect on the results obtained. The high molar percent of Hyp found in these samples also brings into question the long-held notion that Hyp makes up approximately 13.4% of the collagen molecule.

Comparisons between the ratio of mature to immature cross-links revealed that the deer was most likely the oldest of the animals and sheep the youngest. Cross-link analysis also revealed that both of the cross-links recently proposed to be artefacts of laboratory preparation were present despite the elimination of the hypothesised contributing factor. HHL which is proposed to be formed by the reaction of two telopeptide Lys and a glycosylated Hyl residue located within the N-telopeptide was present in the non-reduced pepsin (i.e. non-telopeptide containing) extractions of deer and sheep. It is interesting that reduced, non-telopeptide collagen from all animals contained significant concentrations of HHL's precursor; HLNL, which is formed between an N-telopeptide Lys and a helical Hyl

on a neighbouring molecule. These divalent cross-links may be present in collagen that have their telopeptide regions cleaved as reduction prior to pepsin cleavage would stabilise these cross-links, leaving helical collagen with a telopeptide-only region attached. This may account for the fact that it was difficult to see any difference between the full-length collagen chains and the cleaved collagen chains on the SDS-PAGE gels. However, this would not facilitate the formation of HHL in the manner that is proposed by Eyre *et al*, as another collagen telopeptide would have to be in close proximity with the correct lysine residue located at just the right distance during hydrolysis, and for this circumstance to occur at such a high incidence is extremely unlikely. Likewise, HHMD which was proposed to be produced by the reaction of sodium borohydride reduction was found in skins which had not been reduced. It is also interesting to note that the residue proposed to be involved in HHL formation was not found by the mass spectral analysis to be glycosylated in any of the animals. Glycosylation analysis also revealed a glycosylation site that appears to be unique to cow skins. There is a possibility that this may contribute to the formation of HLNL cross-links, as these were found in the highest concentration in cow skins and are formed in the N-telopeptide region located near to this site.

Mass spectral analysis also revealed that on average 70% of proline residues were in fact hydroxylated, supporting the results found by amino acid analysis. This further highlights the need for the method to calculate collagen concentrations in tissue samples to be revisited, as the method devised by Neuman over 70 years ago does not fit with this ratio of Pro/Hyp. Mass spectral analysis of trypsin digested collagen also revealed that contrary to the belief that the SDS-PAGE results from section 3.3 contained predominantly pure collagen type I, with α1(I) and α2(I) found at a ratio of (2:1), in actuality collagen (α1)III was found at similar percentages to collagen α1(I), and collagen α2(I) was the predominant collagen chain at percentages ranging from 6382 %. This was very surprising, and provides a possible explanation for the unusual results of the amino acid analysis found in section 4.3. If collagen type II is a major contributor to the results found in this study, with only ~45 % homology to collagen type I it may produce a very different amino acid and cross-link profile.

One of the shortcomings of this project is that only three technical replicates from one biological sample for each animal was able to be used, due to budget restrictions and sourceability. This limits the statistical significance of the findings, however given the variation found between technical replicates of the same skins, taken from the same OSP region of each animal it is likely that variations between biological replicates would be even greater. Covington [31] addresses this issue stating that comparisons between two different animals of the same species cannot be made without very large numbers of biological replicates, as each animal will vary in age, breed, sex, diet, husbandry, medical history and treatment, and many other aspects. This makes comparisons between the results difficult, particularly as the skins used in this study were by-products of the meat industry and as such all of these details are unknown. Without time and budget restraints, a large number of technical replicates from the same animal of a known age and breed would be ideal and would help to eliminate some inconsistencies observed across the data.

6.2 Future directions

There are several unanswered questions arising from this work which will be addressed into the future:

1. Bands produced upon heating of the collagen extracts prior to SDS-PAGE will be cut out and subjected to in-gel tryptic digestion followed by mass spectral analysis. This is required to determine whether these are truly breakdown products of collagen or if the bands are the product of the removal of accessory proteins.

2. Enzymatically hydrolysed, salt-extracted collagen samples will be analysed on LC/MS to determine if both diastereomers of Hyl are present. This will determine whether Hyl2 is an artefact produced by acid hydrolysis or truly exists *in vivo*.

3. LC/MS analysis of non-reduced, enzymatically hydrolysed, salt-extracted collagen samples will also be used to determine if the cross-links HHL and HHMD exist *in vivo* or are also artefacts of the combination of acid hydrolysis and or sodium borohydride reduction as proposed recently.

7. References

1. Di Lullo, G.A., et al., *Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen.* Journal of Biological Chemistry, 2002. **277**(6): p. 4223-4231.
2. Sikorski, Z.E., *Functional properties of proteins in food systems.* Chemical and functional properties of food proteins, 2001: p. 113-135.
3. Ashhurst, D.E. and N.M. Costin, *The development of a collagenous connective tissue in the locust, Locusta migratoria.* Tissue and Cell, 1974. **6**(2): p. 279-300.
4. Ashhurst, D.E. and A.J. Bailey, *Locust Collagen: Morphological and Biochemical Characterization.* 1980. **103**(1): p. 75-83.
5. Sutherland, T.D., et al., *A new class of animal collagen masquerading as an insect silk.* Scientific Reports, 2013. **3**: p. 2864.
6. Yasothornsrikul, S., et al., *viking: identification and characterization of a second type IV collagen in Drosophila.* Gene, 1997. **198**(1): p. 17-25.
7. Tziveleka, L.-A., et al., *Collagen from the Marine Sponges Axinella cannabina and Suberites carnosus: Isolation and Morphological, Biochemical, and Biophysical Characterization.* Marine drugs, 2017. **15**(6): p. 152.
8. Huxley - Jones, J., D.L. Robertson, and R.P. Boot-Handford, *On the origins of the extracellular matrix in vertebrates.* Matrix Biology, 2007. **26**(1): p. 2-11.
9. Granito, R.N., M.R. Custódio, and A.C.M. Rennó, *Natural marine sponges for bone tissue engineering: The state of art and future perspectives.* 2017. **105**(6): p. 1717-1727.
10. Yu, Z., et al., *Bacterial collagen-like proteins that form triple-helical structures.* Journal of structural biology, 2014. **186**(3): p. 451-461.
11. Rasmussen, M., M. Jacobsson, and L. Björck, *Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins.* 2003. **278**(34): p. 32313-32316.
12. Nocera, A.D., et al., *Development of 3D printed fibrillar collagen scaffold for tissue engineering.* Biomedical Microdevices, 2018. **20**(2): p. 26.
13. Chvapil, M., *Collagen sponge: Theory and practice of medical applications.* 1977. **11**(5): p. 721-741.

14. Rhee, S., et al., *3D Bioprinting of Spatially Heterogeneous Collagen Constructs for Cartilage Tissue Engineering.* ACS Biomaterials Science & Engineering, 2016. **2**(10): p. 1800-1805.
15. Koo, Y., et al., *3D printed cell-laden collagen and hybrid scaffolds for in vivo articular cartilage tissue regeneration.* Journal of Industrial and Engineering Chemistry, 2018. **66**: p. 343-355.
16. Parmar, P.A., et al., *Harnessing the Versatility of Bacterial Collagen to Improve the Chondrogenic Potential of Porous Collagen Scaffolds.* Advanced healthcare materials, 2016. **5**(13): p. 1656-1666.
17. Gaspar, A., et al., *Collagen-based scaffolds for skin tissue engineering.* Journal of medicine and life, 2011. **4**(2): p. 172-177.
18. Marques, C.F., et al., *Collagen-based bioinks for hard tissue engineering applications: a comprehensive review.* Journal of Materials Science: Materials in Medicine, 2019. **30**(3): p. 32.
19. Bailey, A. and M. Shimokomaki, *Age related changes in the reducible cross-links of collagen.* FEBS letters, 1971. **16**(2): p. 86-88.
20. Brinckmann, J., et al., *Different Pattern of Collagen Cross-Links in Two Sclerotic Skin Diseases: Lipodermatosclerosis and Circumscribed Scleroderma.* Journal of Investigative Dermatology, 2001. **117**(2): p. 269-273.
21. Byers , P.H., et al., *X-Linked Cutis Laxa.* New England Journal of Medicine, 1980. **303**(2): p. 61-65.
22. Dyer, D.G., et al., *Accumulation of Maillard reaction products in skin collagen in diabetes and aging.* Journal of Clinical Investigation, 1993. **91**(6): p. 2463-2469.
23. Herbert, C.M., et al., *Biosynthesis and maturation of skin collagen in scleroderma, and effect of D-penicillamine.* The Lancet, 1974. **303**(7850): p. 187-192.
24. Ihme, A., et al., *Ehlers-Danlos syndrome type VI: collagen type specificity of defective lysyl hydroxylation in various tissues.* Journal of investigative dermatology, 1984. **83**(3): p. 161-165.
25. van der Slot-Verhoeven, A.J., et al., *The type of collagen cross-link determines the reversibility of experimental skin fibrosis.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2005. **1740**(1): p. 60-67.

26. Yamauchi, M., et al., *Collagen Cross-Linking in Sun-Exposed and Unexposed Sites of Aged Human Skin.* Journal of Investigative Dermatology, 1991. **97**(5): p. 937-940.
27. Eyre, D.R., *The specificity of collagen cross-links as markers of bone and connective tissue degradation.* Acta Orthopaedica Scandinavica, 1995. **66**(sup266): p. 166-170.
28. Savolainen, E.-R., et al., *Deficiency of galactosylhydroxylysyl glucosyltransferase, an enzyme of collagen synthesis, in a family with dominant epidermolysis bullosa simplex.* New England Journal of Medicine, 1981. **304**(4): p. 197-204.
29. Fujii, K., M.L. Tanzer, and P.H. Cooke, *Collagen fibrogenesis and the formation of complex crosslinks.* Journal of Molecular Biology, 1976. **106**(1): p. 223-227.
30. Hanset, R. and M. Ansay, *Dermatosparaxie (peau déchirée) chez le veau: un défaut général du tissu conjonctif, de nature héréditaire.* Ann Med Vet, 1967. **7**: p. 451-470.
31. Covington, A.D., *Modern tanning chemistry.* Chemical Society Reviews, 1997. **26**(2): p. 111-126.
32. Schaefer, L. and R.M. Schaefer, *Proteoglycans: from structural compounds to signaling molecules.* Cell and tissue research, 2010. **339**(1): p. 237.
33. Rao, H., et al., *Proteomic identification of adhesive on a bone sculpture-inlaid wooden artifact from the Xiaohe Cemetery, Xinjiang, China.* Journal of Archaeological Science, 2015. **53**: p. 148-155.
34. Lucas, A. and J. Harris, *Ancient Egyptian materials and industries*. 2012: Courier Corporation.
35. Bleicher, N., et al., *Molecular evidence of use of hide glue in 4th millennium BC Europe.* Journal of Archaeological Science, 2015. **63**: p. 65-71.
36. Solazzo, C., et al., *Identification of the earliest collagen- and plant-based coatings from Neolithic artefacts (Nahal Hemar cave, Israel).* Scientific Reports, 2016. **6**: p. 31053.
37. Adams, R.D., *Adhesive bonding: science, technology and applications*. 2005: Elsevier.

38. Pinhasi, R., et al., *First direct evidence of chalcolithic footwear from the near eastern highlands.* 2010. **5**(6): p. e10984.

39. O'Sullivan, N.J., et al., *A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing.* Scientific Reports, 2016. **6**: p. 31279.

40. Hulmes, D.J.S., *Collagen Diversity, Synthesis and Assembly*, in *Collagen: Structure and Mechanics*, P. Fratzl, Editor. 2008, Springer US: Boston, MA. p. 15-47.

41. Benson, B., et al., *Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino acid sequence.* Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(19): p. 6379.

42. Hoppe, H.-J. and K.B.M. Reid, *Collectins — soluble proteins containing collagenous regions and lectin domains — and their roles in innate immunity.* Protein Science, 1994. **3**(8): p. 1143-1158.

43. Krejci, E., et al., *Primary structure of a collagenic tail peptide of Torpedo acetylcholinesterase: co-expression with catalytic subunit induces the production of collagen-tailed forms in transfected cells.* The EMBO Journal, 1991. **10**(5): p. 1285-1293.

44. Kodama, T., et al., *Type I macrophage scavenger receptor contains α-helical and collagen-like coiled coils.* Nature, 1990. **343**(6258): p. 531-535.

45. Luo, Y. and M. Karsdal, *Type XI collagen*, in *Biochemistry of collagens, laminins and elastin*. 2016, Elsevier. p. 77-80.

46. Kadler, K.E., A. Hill, and E.G. Canty-Laird, *Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators.* Current Opinion in Cell Biology, 2008. **20**(5): p. 495-501.

47. Aberoumand, A., *Comparative study between different methods of collagen extraction from fish and its properties.* World Applied Sciences Journal, 2012. **16**(3): p. 316-319.

48. Ramachandran, G., M. Bansal, and R. Bhatnagar, *A hypothesis on the role of hydroxyproline in stabilizing collagen structure.* Biochimica et Biophysica Acta (BBA)-Protein Structure, 1973. **322**(1): p. 166-171.

49. Reich, G., *From collagen to leather-the theoretical background*. 2007: BASF Service center.

50. Eckhard, U., E. Schönauer, and H. Brandstetter, *Structural basis for activity regulation and substrate preference of clostridial collagenases G, H, and T.* Journal of Biological Chemistry, 2013. **288**(28): p. 20184-20194.

51. Sricholpech, M., et al., *Lysyl Hydroxylase 3 Glucosylates Galactosylhydroxylysine Residues in Type I Collagen in Osteoblast Culture.* Journal of Biological Chemistry, 2011. **286**(11): p. 8846-8856.

52. Yamauchi, M. and M. Sricholpech, *Lysine post-translational modifications of collagen.* Essays in biochemistry, 2012. **52**: p. 113-133.

53. Raghunath, M., P. Bruckner, and B. Steinmann, *Delayed Triple Helix Formation of Mutant Collagen from Patient with Osteogenesis Imperfecta.* Journal of Molecular Biology, 1994. **236**(3): p. 940-949.

54. Koide, T. and K. Nagata, *Collagen Biosynthesis*, in *Collagen: Primer in Structure, Processing and Assembly*, J. Brinckmann, H. Notbohm, and P.K. Müller, Editors. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 85-114.

55. Makareeva, E. and S. Leikin, *Procollagen Triple Helix Assembly: An Unconventional Chaperone-Assisted Folding Paradigm.* PLOS ONE, 2007. **2**(10): p. e1029.

56. Martinek, N., et al., *Is SPARC an Evolutionarily Conserved Collagen Chaperone?* Journal of Dental Research, 2007. **86**(4): p. 296-305.

57. Pace, J.M., et al., *Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the proα1(I) chain of type I procollagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta.* Journal of Medical Genetics, 2001. **38**(7): p. 443-449.

58. Chessler, S.D., G.A. Wallis, and P.H. Byers, *Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of type I collagen result in defective chain association and produce lethal osteogenesis imperfecta.* Journal of Biological Chemistry, 1993. **268**(24): p. 18218-18225.

59. Schofield, J.D., J. Uitto, and D.J. Prockop, *Formation of interchain disulfide bonds and helical structure during biosynthesis of procollagen by embryonic tendon cells.* Biochemistry, 1974. **13**(9): p. 1801-1806.

60. Olsen, B.R., H.-P. Hoffmann, and D.J. Prockop, *Interchain disulfide bonds at the COOH-terminal end of procollagen synthesized by matrix-free cells from chick embryonic tendon and cartilage.* Archives of Biochemistry and Biophysics, 1976. **175**(1): p. 341-350.

61. Bächinger, H.P., et al., *Folding Mechanism of the Triple Helix in Type-III Collagen and Type-III pN–Collagen: Role of Disulfide Bridges and Peptide Bond Isomerization.* European journal of biochemistry, 1980. **106**(2): p. 619-632.

62. Doege, K.J. and J.H. Fessler, *Folding of carboxyl domain and assembly of procollagen I.* Journal of Biological Chemistry, 1986. **261**(19): p. 8924-35.

63. Dion, A.S. and J.C. Myers, *COOH-terminal propeptides of the major human procollagens: Structural, functional and genetic comparisons.* Journal of Molecular Biology, 1987. **193**(1): p. 127-143.

64. Koivu, J., *Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen.* FEBS Letters, 1987. **212**(2): p. 229-232.

65. Zafarullah, K., et al., *Synthesis and conformational properties of a recombinant C-propeptide of human type III procollagen.* Matrix Biology, 1997. **16**(4): p. 201-209.

66. Bernocco, S., et al., *Biophysical Characterization of the C-propeptide Trimer from Human Procollagen III Reveals a Tri-lobed Structure.* Journal of Biological Chemistry, 2001. **276**(52): p. 48930-48936.

67. Hulmes, D.J.S., *Building Collagen Molecules, Fibrils, and Suprafibrillar Structures.* Journal of Structural Biology, 2002. **137**(1): p. 2-10.

68. Lees, J.F., M. Tasab, and N.J. Bulleid, *Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen.* The EMBO Journal, 1997. **16**(5): p. 908-916.

69. Bulleid, N.J., R. Wilson, and J.F. Lees, *Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation.* Biochemical Journal, 1996. **317**(1): p. 195-202.

70. Buevich, A.V., et al., *Site-Specific NMR Monitoring of cis−trans Isomerization in the Folding of the Proline-Rich Collagen Triple Helix.* Biochemistry, 2000. **39**(15): p. 4299-4308.

71. Bächinger, H.P., et al., *The Role of Cis-Trans Isomerization of Peptide Bonds in the Coil⇄ Triple Helix Conversion of Collagen.* European journal of biochemistry, 1978. **90**(3): p. 605-613.

72. Stephens, D.J. and R. Pepperkok, *Imaging of procollagen transport reveals COPI-dependent cargo sorting during ER-to-Golgi transport in mammalian cells.* Journal of Cell Science, 2002. **115**(6): p. 1149-1160.

73. Bonfanti, L., et al., *Procollagen Traverses the Golgi Stack without Leaving the Lumen of Cisternae: Evidence for Cisternal Maturation.* Cell, 1998. **95**(7): p. 993-1003.

74. Leblond, C., *Synthesis and secretion of collagen by cells of connective tissue, bone, and dentin.* The Anatomical Record, 1989. **224**(2): p. 123-138.

75. Trelstad, R.L. and K. Hayashi, *Tendon collagen fibrillogenesis: Intracellular subassemblies and cell surface changes associated with fibril growth.* Developmental Biology, 1979. **71**(2): p. 228-242.

76. Leikina, E., et al., *Type I collagen is thermally unstable at body temperature.* Proceedings of the National Academy of Sciences, 2002. **99**(3): p. 1314-1318.

77. Trelstad, R.L., *Vacuoles in the embryonic chick corneal epithelium, an epithelium which produces collagen.* The Journal of cell biology, 1971. **48**(3): p. 689.

78. Sorushanova, A., et al., *The collagen suprafamily: from biosynthesis to advanced biomaterial development.* Advanced Materials, 2019. **31**(1): p. 1801651.

79. Wu, C.H., C.B. Donovan, and G.Y. Wu, *Evidence for pretranslational regulation of collagen synthesis by procollagen propeptides.* Journal of Biological Chemistry, 1986. **261**(23): p. 10482-4.

80. Bella, J., B. Brodsky, and H.M. Berman, *Hydration structure of a collagen peptide.* Structure, 1995. **3**(9): p. 893-906.

81. Bella, J. and H.M. Berman, *Crystallographic evidence for Cα–H··· O= C hydrogen bonds in a collagen triple helix.* Journal of molecular biology, 1996. **264**(4): p. 734-742.

82. Berisio, R., et al., *Imino acids and collagen triple helix stability: characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats.* Journal of the American Chemical Society, 2004. **126**(37): p. 11402-11403.

83. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation.* Biopolymers: Original Research on Biomolecules, 2001. **58**(5): p. 459-464.

84. Mizuno, K., et al., *Vascular Ehlers-Danlos Syndrome Mutations in Type III Collagen Differently Stall the Triple Helical Folding.* Journal of Biological Chemistry, 2013. **288**(26): p. 19166-19176.

85. Ishikawa, Y., et al., *Biochemical Characterization of the Prolyl 3-Hydroxylase 1·Cartilage-associated Protein·Cyclophilin B Complex.* Journal of Biological Chemistry, 2009. **284**(26): p. 17641-17647.

86. Vranka, J.A., et al., *Prolyl 3-Hydroxylase 1 Null Mice Display Abnormalities in Fibrillar Collagen-rich Tissues Such as Tendons, Skin, and Bones.* Journal of Biological Chemistry, 2010. **285**(22): p. 17253-17262.

87. Canty, E.G. and K.E. Kadler, *Procollagen trafficking, processing and fibrillogenesis.* Journal of Cell Science, 2005. **118**(7): p. 1341-1353.

88. Neuman, R.E. and M.A. Logan, *The determination of hydroxyproline.* J Biol Chem, 1950. **184**(1): p. 299-306.

89. Privalov, P.L., E.I. Tiktopulo, and V.M. Tischenko, *Stability and mobility of the collagen structure.* Journal of Molecular Biology, 1979. **127**(2): p. 203-216.

90. Passoja, K., et al., *Cloning and characterization of a third human lysyl hydroxylase isoform.* Proceedings of the National Academy of Sciences, 1998. **95**(18): p. 10482-10486.

91. Walker, L.C., M.A. Overstreet, and H.N. Yeowell, *Tissue-specific expression and regulation of the alternatively-spliced forms of lysyl hydroxylase 2 (LH2) in human kidney cells and skin fibroblasts.* Matrix Biology, 2005. **23**(8): p. 515-523.

92. Salo, A.M., et al., *The lysyl hydroxylase isoforms are widely expressed during mouse embryogenesis, but obtain tissue-and cell-specific patterns in the adult.* Matrix biology, 2006. **25**(8): p. 475-483.

93. Açil, Y., et al., *Ehlers-Danlos syndrome type VI: cross-link pattern in tissue and urine sample as a diagnostic marker.* Journal of the American Academy of Dermatology, 1995. **33**(3): p. 522-524.

94. Steinmann, B., D.R. Eyre, and P. Shao, *Urinary pyridinoline cross-links in Ehlers-Danlos syndrome type VI.* American journal of human genetics, 1995. **57**(6): p. 1505.

95. Bank, R.A., et al., *Defective collagen crosslinking in bone, but not in ligament or cartilage, in Bruck syndrome: indications for a bone-specific telopeptide lysyl hydroxylase on chromosome 17.* Proceedings of the National Academy of Sciences, 1999. **96**(3): p. 1054-1058.

96. Hyland, J., et al., *A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers–Danlos syndrome type VI.* Nature genetics, 1992. **2**(3): p. 228.

97. Yeowell, H.N. and L.C. Walker, *Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers–Danlos syndrome type VI.* Molecular genetics and metabolism, 2000. **71**(1-2): p. 212-224.

98. Ha-Vinh, R., et al., *Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2.* American journal of medical genetics Part A, 2004. **131**(2): p. 115-120.

99. Puig-Hervás, M.T., et al., *Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome–Osteogenesis imperfecta phenotypic spectrum.* Human mutation, 2012. **33**(10): p. 1444-1449.

100. van der Slot, A.J., et al., *Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis.* Journal of Biological Chemistry, 2003. **278**(42): p. 40967-40972.

101. Zhou, P., et al., *Novel mutations in FKBP10 and PLOD2 cause rare Bruck syndrome in Chinese patients.* PloS one, 2014. **9**(9): p. e107594.

102. Uzawa, K., et al., *Differential expression of human lysyl hydroxylase genes, lysine hydroxylation, and cross-linking of type I collagen during osteoblastic differentiation in vitro.* Journal of Bone and Mineral Research, 1999. **14**(8): p. 1272-1280.

103. van der Slot, A.J., et al., *Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon.* Matrix biology, 2004. **23**(4): p. 251-257.

104. Pornprasertsuk, S., et al., *Lysyl hydroxylase-2b directs collagen cross-linking pathways in MC3T3-E1 cells.* Journal of Bone and Mineral Research, 2004. **19**(8): p. 1349-1355.

105. Brinckmann, J., et al., *Interleukin 4 and prolonged hypoxia induce a higher gene expression of lysyl hydroxylase 2 and an altered cross-link pattern: important pathogenetic steps in early and late stage of systemic scleroderma?* Matrix biology, 2005. **24**(7): p. 459-468.

106. Schwarze, U., et al., *Mutations in FKBP10, which result in Bruck syndrome and recessive forms of osteogenesis imperfecta, inhibit the hydroxylation of telopeptide lysines in bone collagen.* Human molecular genetics, 2012. **22**(1): p. 1-17.

107. Piez, K., *The separation of the diastereoisomers of isoleucine and hydroxylysine by ion exchange chromatography.* Journal of Biological Chemistry, 1954. **207**(1): p. 77-80.

108. Hamilton, P.B. and R.A. Anderson, *Hydroxylysine: isolation from gelatin and resolution of its diastereoisomers by ion exchange chromatography.* Journal of Biological Chemistry, 1955. **213**(1): p. 249-258.

109. Brownell, A.G. and A. Veis, *The intracellular location of the glycosylation of hydroxylysine of collagen.* Biochemical and biophysical research communications, 1975. **63**(2): p. 371-377.

110. Kimura, S., *Determination of the glycosylated hydroxylysines in several invertebrate collagens.* The Journal of Biochemistry, 1972. **71**(2): p. 367-370.

111. Oikarinen, A., H. Anttinen, and K.I. Kivirikko, *Hydroxylation of lysine and glycosylation of hydroxylysine during collagen biosynthesis in isolated chick-embryo cartilage cells.* Biochemical Journal, 1976. **156**(3): p. 545-551.

112. Witkop, B., *The application of Hudson's lactone rule toγ-andδ-hydroxyamino acids and the question of the configuration ofδ-hydroxy-l-lysine from collagen.* Experientia, 1956. **12**(10): p. 372-374.

113. Gjaltema, R.A.F. and R.A. Bank, *Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease.* Critical Reviews in Biochemistry and Molecular Biology, 2017. **52**(1): p. 74-95.

114. Heikkinen, J., et al., *Lysyl Hydroxylase 3 Is a Multifunctional Protein Possessing Collagen Glucosyltransferase Activity.* Journal of Biological Chemistry, 2000. **275**(46): p. 36158-36163.

115. Sricholpech, M., et al., *Lysyl hydroxylase 3-mediated glucosylation in type I collagen: Molecular loci and biological significance.* Journal of Biological Chemistry, 2012. **287**(27): p. 22998-23009.

116. Brinckmann, J., et al., *Overhydroxylation of lysyl residues is the initial step for altered collagen cross-links and fibril architecture in fibrotic skin.* Journal of investigative dermatology, 1999. **113**(4): p. 617-621.

117. Notbohm, H., et al., *Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro.* Journal of Biological Chemistry, 1999. **274**(13): p. 8988-8992.

118. Fratzl, P., *Collagen: structure and mechanics*. 2008: Springer Science & Business Media.

119. Hulmes, D.J., et al., *Analysis of the primary structure of collagen for the origins of molecular packing.* Journal of molecular biology, 1973. **79**(1): p. 137-148.

120. Ricard-Blum, S. and G. Ville, *Collagen cross-linking.* International Journal of Biochemistry, 1989. **21**(11): p. 1185-1189.

121. Bailey, A., N. Light, and E. Atkins, *Chemical cross-linking restrictions on models for the molecular organization of the collagen fibre.* Nature, 1980. **288**(5789): p. 408.

122. Mechanic, G.L., *Collagen crosslinks: Direct evidence of a reducible stable form of the Schiff base Δ6dehydro-5, 5'-dihydroxylysinonorleucine as 5-keto-5'-hydroxylysinonorleucine in bone collagen.* Biochemical and biophysical research communications, 1974. **56**(4): p. 923-927.

123. Tanzer, M.L. and G. Mechanic, *Collagen reduction by sodium borohydride: Effects of reconstitution, maturation and lathyrism.* Biochemical and biophysical research communications, 1968. **32**(5): p. 885-892.

124. Ribel-Madsen, S., et al., *Urinary markers of altered collagen metabolism in fibromyalgia patients.* Scandinavian journal of rheumatology, 2007. **36**(6): p. 470-477.

125. Chan, C.C., et al., *Bone turnover in young hypoestrogenic women on hormonal therapy.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 2006. **124**(2): p. 204-206.

126. Tanzer, M.L., et al., *Structure of two histidine-containing cross-links from collagen.* Journal of Biological Chemistry, 1973. **248**(2): p. 393-402.

127. Yamauchi, M., et al., *Structure and formation of a stable histidine-based trifunctional cross-link in skin collagen.* Journal of Biological Chemistry, 1987. **262**(24): p. 11428-11434.

128. Naylor, E.C., R.E. Watson, and M.J. Sherratt, *Molecular aspects of skin ageing.* Maturitas, 2011. **69**(3): p. 249-256.

129. Bailey, A.J. and R.G. Paul. *The mechanisms and consequences of the maturation and ageing of collagen.* in *Proceedings of the Indian Academy of Sciences-Chemical Sciences*. 1999. Springer.

130. Eyre, D.R., M.A. Weis, and J.-J. Wu, *Advances in collagen cross-link analysis.* Methods, 2008. **45**(1): p. 65-74.

131. Siegel, R.C. and J.B. Lian, *Lysyl oxidase dependent synthesis of a collagen cross-link containing histidine.* Biochemical and biophysical research communications, 1975. **67**(4): p. 1353-1359.

132. Tanzer, M.L., R. Fairweather, and P.M. Gallop, *Isolation of the crosslink, hydroxymerodesmosine, from borohydride-reduced collagen.* Biochimica et Biophysica Acta (BBA)-Protein Structure, 1973. **310**(1): p. 130-136.

133. Davis, N.R., *Stable crosslinks of collagen.* Biochemical and biophysical research communications, 1973. **54**(3): p. 914-922.

134. Reiser, K.M., S.M. Hennessy, and J.A. Last, *Analysis of age-associated changes in collagen crosslinking in the skin and lung in monkeys and rats.* Biochimica et Biophysica Acta (BBA)-General Subjects, 1987. **926**(3): p. 339-348.

135. Rojkind, M., O.O. Blumenfeld, and P.M. Gallop, *Localization and partial characterization of an aldehydic component in tropocollagen.* Journal of Biological Chemistry, 1966. **241**(7): p. 1530-1536.

136. Bailey, A. and C. Peach, *Identification of a labile intermolecular cross-link in collagen.* Biochemical Journal, 1969. **111**(3): p. 12P.

137. Bailey, A., C.M. Peach, and L. Fowler, *Chemistry of the collagen cross-links. Isolation and characterization of two intermediate intermolecular cross-links in collagen.* Biochemical Journal, 1970. **117**(5): p. 819-831.

138. Robins, S.P. and A.J. Bailey, *The chemistry of the collagen cross-links. The characterization of fraction C, a possible artifact produced during the reduction of collagen fibres with borohydride.* Biochemical Journal, 1973. **135**(4): p. 657-665.

139. Kang, A.H. and J. Gross, *Relationship between the intra and intermolecular cross-links of collagen.* Proceedings of the National Academy of Sciences, 1970. **67**(3): p. 1307-1314.

140. Bailey, A.J. and T.J. Sims, *Chemistry of the collagen cross-links. Nature of the cross-links in the polymorphic forms of dermal collagen during development.* Biochemical Journal, 1976. **153**(2): p. 211-215.

141. Bernstein, P.H. and G. Mechanic, *A natural histidine-based imminium cross-link in collagen and its location.* Journal of Biological Chemistry, 1980. **255**(21): p. 10414-10422.

142. Shoulders, M.D. and R.T. Raines, *Collagen structure and stability.* Annual review of biochemistry, 2009. **78**: p. 929-958.

143. Buehler, M.J., *Nature designs tough collagen: explaining the nanostructure of collagen fibrils.* Proceedings of the National Academy of Sciences, 2006. **103**(33): p. 12285-12290.

144. Naffa, R.M., *Understanding the Molecular Basis of the Strength Differences in Skins Used in Leather Manufacture*. 2017, Massey University: Palmerston North, New Zealand.

145. Au - Ernst, O. and T. Au - Zor, *Linearization of the Bradford Protein Assay.* JoVE, 2010(38): p. e1918.

146. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical Biochemistry, 1976. **72**(1): p. 248-254.

147. Yamauchi, M. and M. Shiiba, *Lysine hydroxylation and cross-linking of collagen*, in *Post-translational Modifications of Proteins*. 2008, Springer. p. 95-108.

148. Wang, L., et al., *Extraction of pepsin-soluble collagen from grass carp (Ctenopharyngodon idella) skin using an artificial neural network.* Food Chemistry, 2008. **111**(3): p. 683-686.

149. Wu, J., et al., *Extraction and isolation of type I, III and V collagens and their SDS-PAGE analyses.* Transactions of Tianjin University, 2011. **17**(2): p. 111.

150. Nalinanon, S., et al., *Use of pepsin for collagen extraction from the skin of bigeye snapper (Priacanthus tayenus).* Food Chemistry, 2007. **104**(2): p. 593-601.

151. Kumazawa, Y., et al., *A rapid and simple LC-MS method using collagen marker peptides for identification of the animal source of leather.* Journal of agricultural and food chemistry, 2016. **64**(30): p. 6051-6057.

152. Cliche, S., et al., *Extraction and characterization of collagen with or without telopeptides from chicken skin.* Poultry science, 2003. **82**(3): p. 503-509.

153. Zhou, P. and J.M. Regenstein, *Effects of alkaline and acid pretreatments on Alaska pollock skin gelatin extraction.* Journal of Food Science, 2005. **70**(6): p. c392-c396.

154. Jongjareonrak, A., et al., *Isolation and characterization of collagen from bigeye snapper (Priacanthus macracanthus) skin.* Journal of the Science of Food and Agriculture, 2005. **85**(7): p. 1203-1210.

155. Piper, D. and B.H. Fenton, *pH stability and activity curves of pepsin with special reference to their clinical importance.* Gut, 1965. **6**(5): p. 506.

156. Zeugolis, D.I., R.G. Paul, and G. Attenburrow, *Factors influencing the properties of reconstituted collagen fibers prior to self-assembly: Animal species and collagen extraction method.* Journal of Biomedical Materials Research Part A, 2008. **86A**(4): p. 892-904.

157. Yamauchi, M. and M. Shiiba, *Lysine Hydroxylation and Cross-linking of Collagen*, in *Post-translational Modifications of Proteins: Tools for Functional Proteomics*, C. Kannicht, Editor. 2008, Humana Press: Totowa, NJ. p. 95-108.

158. Duan, R., et al., *Properties of collagen from skin, scale and bone of carp (Cyprinus carpio).* Food chemistry, 2009. **112**(3): p. 702-706.

159. Bank, R.A., et al., *Sensitive fluorimetric quantitation of pyridinium and pentosidine crosslinks in biological samples in a single high-performance liquid chromatographic run.* Journal of Chromatography B: Biomedical Sciences and Applications, 1997. **703**(1): p. 37-44.

160. Neuman, R.E. and M.A. Logan, *The determination of collagen and elastin in tissues.* Journal of Biological Chemistry, 1950. **186**(2): p. 549-556.

161. Cohen, S.A., *Amino acid analysis using precolumn derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate.* Methods Mol Biol, 2000. **159**: p. 39-47.

162. Brückner, H. and M. Wachsmann, *Evaluation of fluorescent monochloro-s-triazine reagent for amino acid analysis.* Chromatographia, 2003. **57**(1): p. S143-S146.

163. Cohen, S.A. and D.P. Michaud, *Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids via high-performance liquid chromatography.* Analytical biochemistry, 1993. **211**(2): p. 279-287.

164. Horie, H. and K. Kohata, *Analysis of tea components by high-performance liquid chromatography and high-performance capillary electrophoresis.* Journal of Chromatography A, 2000. **881**(1): p. 425-438.

165. Heinrikson, R.L. and S.C. Meredith, *Amino acid analysis by reverse-phase high-performance liquid chromatography: Precolumn derivatization with phenylisothiocyanate.* Analytical Biochemistry, 1984. **136**(1): p. 65-74.

166. Fiechter, G. and H.K. Mayer, *Characterization of amino acid profiles of culture media via pre-column 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate derivatization and ultra performance liquid chromatography.* Journal of Chromatography B, 2011. **879**(17): p. 1353-1360.

167. van Wandelen, C. and S.A. Cohen, *Using quaternary high-performance liquid chromatography eluent systems for separating 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate-derivatized amino acid mixtures.* Journal of Chromatography A, 1997. **763**(1): p. 11-22.

168. Deb Choudhury, S. and G. Norris, *Improved binary high performance liquid chromatography for amino acid analysis of collagens*. 2005.

169. Eastoe, J., *The amino acid composition of mammalian collagen and gelatin.* Biochemical Journal, 1955. **61**(4): p. 589.

170. Gallop, P.M. and S. Seifter, *Preparation and properties of soluble collagens*, in *Methods in Enzymology*. 1963, Academic Press. p. 635-641.

171. Cohen, S.A. and C. van Wandelen, *Amino acid analysis of unusual and complex samples based on 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate derivatization*, in *Techniques in Protein Chemistry*. 1997, Elsevier. p. 185-196.

172. Hill, R.L., in *Advances in Protein Chemistry*, C.B. Anfinsen, et al., Editors. 1965, Academic Press. p. 37-107.

173. Gordon, A., A. Martin, and R. Synge, *A study of the partial acid hydrolysis of some proteins, with special reference to the mode of linkage of the basic amino-acids.* Biochemical journal, 1941. **35**(12): p. 1369.

174. Schnider, S.L. and R. Kohn, *Effects of age and diabetes mellitus on the solubility and nonenzymatic glucosylation of human skin collagen.* The Journal of clinical investigation, 1981. **67**(6): p. 1630-1635.

175. Naffa, R., et al., *Liquid chromatography-electrospray ionization mass spectrometry for the simultaneous quantitation of collagen and elastin crosslinks.* Journal of Chromatography A, 2016. **1478**: p. 60-67.

176. Robins, S.P., et al., *Increased skin collagen extractability and proportions of collagen type III are not normalized after 6 months healing of human excisional wounds.* Journal of investigative dermatology, 2003. **121**(2): p. 267-272.

177. Eyre, D.R., M. Weis, and J. Rai, *Analyses of lysine aldehyde cross-linking in collagen reveal that the mature cross-link histidinohydroxylysinonorleucine is an artifact.* Journal of Biological Chemistry, 2019.

178. Torre-Blanco, A., et al., *Copolymerization of normal type I collagen with three mutated type I collagens containing substitutions of cysteine at different glycine positions in the alpha 1 (I) chain.* Journal of Biological Chemistry, 1992. **267**(7): p. 4968-4973.

179. Yang, C.L., et al., *Collagen II from articular cartilage and annulus fibrosus: Structural land functional implication of tissue specific posttranslational modifications of collagen molecules.* European journal of biochemistry, 1993. **213**(3): p. 1297-1302.

180. Torre-Blanco, A., et al., *Temperature-induced post-translational over-modification of type I procollagen. Effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self-assembly of collagen into fibrils.* Journal of Biological Chemistry, 1992. **267**(4): p. 2650-2655.

181. Eyre, D.R. and M.J. Glimcher, *Analysis of a crosslinked peptide from calf bone collagen: evidence that hydroxylysyl glycoside participates in the crosslink.* Biochemical and biophysical research communications, 1973. **52**(2): p. 663-671.

182. Amudeswari, S., J. Liang, and B. Chakrabarti, *Polar-apolar characteristics and fibrillogenesis of glycosylated collagen.* Collagen and related research, 1987. **7**(3): p. 215-223.

183. Bätge, B., et al., *Glycosylation of human bone collagen I in relation to lysylhydroxylation and fibril diameter.* The Journal of Biochemistry, 1997. **122**(1): p. 109-115.

8. Appendices

Table A. 1. Response factors for each amino acid peak were calculated by dividing the concentration in pmol by the peak area averaged over 8 injections.

		AA St	Inj Vol (uL)	1	RATIO
	M .Wt.	RT	AREA	Conc (pmol)	Response
Hyp	131.13	26.78	968404.69	500	0.00051631
Asx	133.11	30.12	170938.02	100	0.00058501
Ser?	126.11	34.65	425646.99	100	0.00023494
Gly/Glu	75.07	39.21	281895.62	100	0.00035474
His	155.15	42.65	345582.46	100	0.00028937
NH3	17.03	47.16	826449.43	100	0.00012100
Ala	119.12	48.43	360252.21	100	0.00027758
Pro	174.2	51.08	395152.99	100	0.00025307
Thr	89.09	51.84	346975.71	100	0.00028820
Arg	115.13	54.36	829716.90	100	0.00012052
Cys	121.16	62.59	47541.16	50	0.00105172
Tyr	181.19	63.80	324861.53	100	0.00030782
Valine	117.15	64.72	545566.48	100	0.00018330
Hyl1	162.19	66.21	100053.74	100	0.00099946
Met	149.21	67.24	358644.29	100	0.00027883
Hyl2	162.19	69.20	169153.18	100	0.00059118
Ile	146.19	76.13	644319.33	100	0.00015520
Lys	131.17	77.73	972916.43	100	0.00010278
Nor	131.17	80.18	2113827.89	200	0.00000000
Phe/Leu	165.19	82.51	820720.54	100	0.00012184

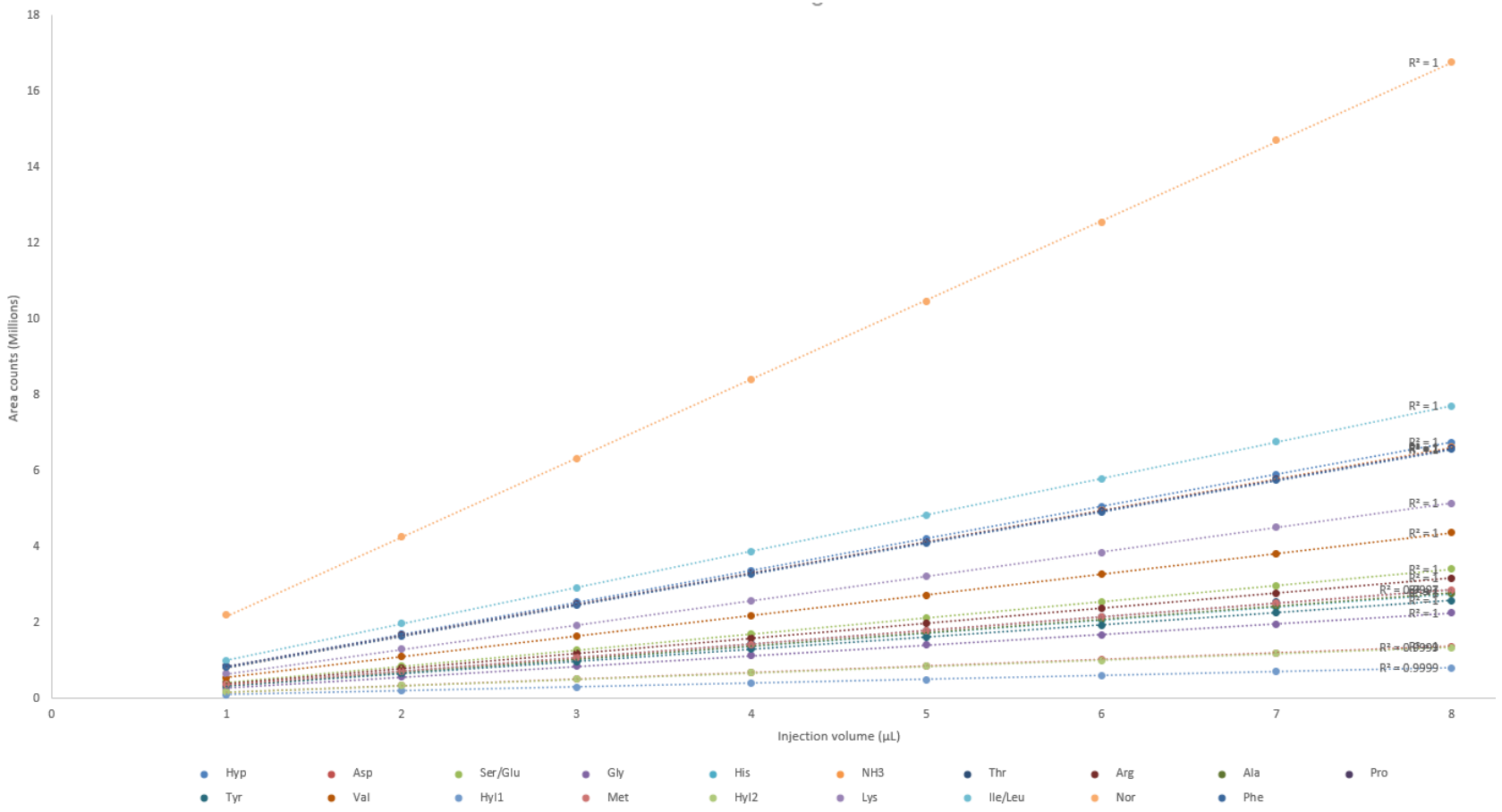
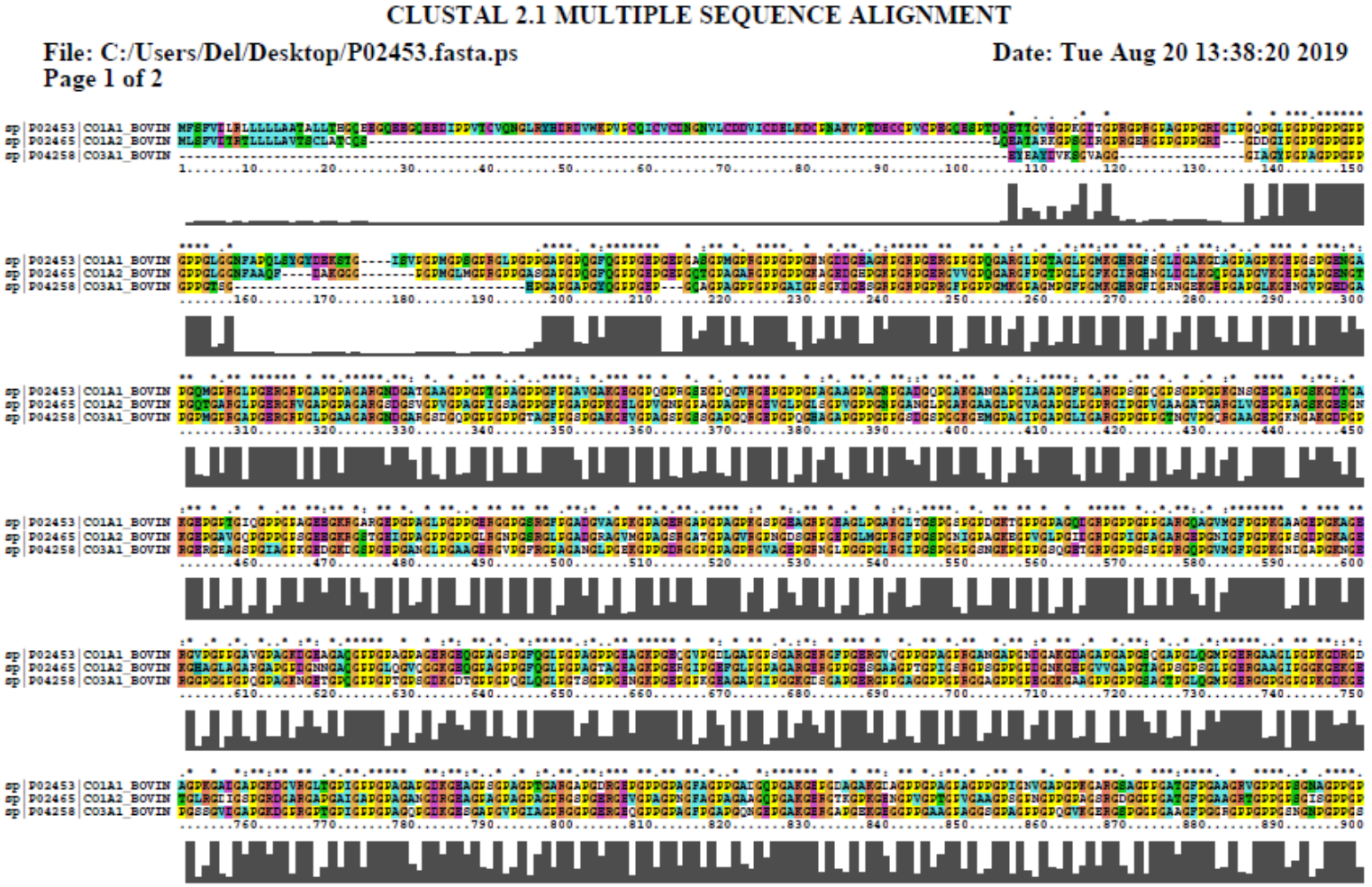


Figure A. 1. Calibration curve of HPLC of AQC-derivatised amino acid standards using a 130-minute run time as described in section 4.3.

Table A. 2. Coefficient of variations (CV%) of cross-links determined by mass spectrometry.

	Coefficient of Variation (CV)															
Xlinks	*PR-COW*	*PN-COW*	*AN-COW*	*GR-COW*	*PR-SHEEP*	*PN-SHEEP*	*AN-SHEEP*	*GR-SHEEP*	*PR-DEER*	*PN-DEER*	*AN-DEER*	*GR-DEER*	*PR-GOAT*	*PN-GOAT*	*AN-GOAT*	*GR-GOAT*
DHLNL	21.44	6.64	21.92	26.77	65.16	4.13	79.45	12.68	22.71	50.24	22.81	28.07	129.53	34.92	20.77	16.24
HLNL	8.42	50.78	60.44	26.48	64.93	4.66	73.86	12.63	15.94	27.71	75.61	12.52	94.27	50.28	19.33	12.79
HHL	4.86	92.19	7.31	28.44	140.20	0.88	3.25	5.38	8.41	1.65	2.23	16.87	81.35	141.99	7.34	16.64
HHMD	0.98	59.25	91.50	8.99	73.54	22.18	73.30	0.91	0.20	57.04	43.62	0.48	71.74	41.91	88.36	6.67
LNL	29.49	34.28	68.23	36.83	42.69	50.96	37.30	34.69	18.96	32.08	120.19	25.85	66.18	58.29	26.50	21.06
DPYR	67.59	123.47	61.58	37.08	134.66	47.39	96.27	92.66	48.74	44.50	149.26	58.33	73.07	117.37	23.38	23.50
PYR	16.46	113.85	12.72	35.12	131.54	30.66	42.66	48.73	69.27	24.83	103.06	13.40	79.38	126.03	20.79	18.72
Avg CV	21.32	68.64	46.24	22.67	85.96	7.96	57.47	7.90	11.82	34.16	36.07	14.49	94.22	67.28	33.95	13.08

Figure A. 2. ClustalX sequence alignment of collagen α1 (I), α2 (I), and α1 (III).



CLUSTAL 2.1 MULTIPLE SEQUENCE ALIGNMENT

File: C:/Users/Del/Desktop/P02453.fasta.ps **Date: Tue Aug 20 13:38:20 2019**

Page 2 of 2