An investigation into effects of *in vitro* glycation on type I collagen fibrillar structure and related biochemistry

Thesis submitted for the degree of Doctor of Philosophy

by

Simon John Goodson

Biophysics Group

School of Optometry and Vision Sciences

Cardiff University

2013

Abstract

Overall, this project aimed to evaluate and advance the use of two techniques in the study of structural changes brought about by *in vitro* glycation of mouse tail tendon type I collagen.

Firstly, using transmission electron microscopy with positive staining, study of collagen glycation by glucose, galactose, fructose and ribose was conducted. Site specificity of glycation along the fibril was seen and it was demonstrated for the first time using this method that there was variation in the progression of the process between the monosaccharides, suggesting different reaction mechanisms and determinants of site specificity for each sugar. From the early changes after 1 week galactose treatment, early-preferred glycation sites for this sugar were indicated. Using sorbitol, broad non-glycation-related sugar binding along the fibril was identified and this was also seen following galactose, fructose and ribose treatment.

Secondly, using X-ray diffraction with isomorphous replacement, a 1D electron density map for collagen fibrils in mouse tail was constructed for the first time as a baseline for this and future intended glycation studies in mice. Use of this technique to study tendon incubated for 2 weeks in ribose and galactose showed electron density changes that appeared consistent with the corresponding electron microscopy glycation data.

Using HPLC with fluorescence detection, it was shown that the four sugars produced different “fingerprint” profiles of products of glycation. Galactose produced a similar profile to glucose as hypothesised. However despite its chemical similarity to glucose, galactose also produced glycation products that were absent from the glucose profile. The biochemical information produced was invaluable for interpretation of the data from the structural techniques, inasmuch as providing representations of the relative extents of glycation occurring during each monosaccharide and control treatment.

In conclusion, this thesis has demonstrated for the first time that transmission electron microscopy with positive staining, and X-ray diffraction with isomorphous replacement techniques together are useful complementary methods for studying glycation-related changes in type I collagen structure. It is anticipated that these techniques together will be of value in further *in vitro* and *in vivo* studies of type I collagen glycation, contributing to a better understanding of the structural bases of the deleterious effects of this process during diabetes and ageing.

Acknowledgements

Firstly, I would like to express due gratitude to my supervisors Dr Justyn Regini and Prof Keith Meek. I am grateful for them giving me the opportunity to embark on this PhD voyage of discovery and for the guidance they have offered me during my studies. I am grateful to Dr Joseph Orgel for giving me the chance to spend over a year doing some of this research within his group at Illinois Institute of Technology in Chicago, USA, for a generous allocation of APS beam-time and for giving me the tools and support to allow me to accomplish my “mission” out there. Thanks also to my advisor, Dr Tom Margrain, for his encouragement and for lending me a listening ear when I needed it.

I would like to thank Dr Rob Young for all the help and support he has offered me in terms of useful scientific and methodological discussions, listening to my rants and providing moral support and encouragement. I am grateful to Prof Andrew Quantock for his cheery support and not least, his tangible enthusiasm for the products of glycation component of this thesis presented in Chapter 3. I am also grateful to Dr Christian Pinali for his help, support and wry wit. Thanks to Dr Gangjin Kang for being such an amiable, charming and pleasant office mate during my first year, for listening to my moans and groans and for being such a top-notch English student. Thank you to Dr Carlo Knupp for his immensely useful tutorials on phase theory and for his kind encouragement, especially after my presentations to the group. In addition, thank you to the other members of the Biophysics group who have supported me in their own, many and varied ways and those who have missed me when I’m not there. A special mention goes to Siân Morgan in that regard.

Sincere gratitude is also due to Steve Morgan, especially for oiling the wheels of my stay in Chicago so adeptly. Also thanks to John Speck for efficiently dealing with the many financial issues and questions I put his way.

Thank you to Mike Underwood in the Joint Facilities Unit, School of Biosciences, for being so obliging in providing subjects for my experiments.

I would like to express my gratitude also, to Nick Avery at Bristol University, for agreeing to collaborate in the part of this study presented in Chapter 3 and for kind support he has shown.

I also offer thanks to my colleagues in Chicago. To Dr Olga Antipova for her useful insights. To Professor Tom Irving for allowing me onto his beam-line at APS and helping me look for lost data. A special mention goes to my other IIT colleagues, most

especially Joanna Poprawski, Rama Sashank Madhurapantula and Vesela Chifchieva, for their friendship, support and helping me fully appreciate and enjoy being in such a fine city.

Finally, thank you to my family and friends outside of the “Land of Dreams” for believing in me and caring.

This work was made possible by funding from the BBSRC.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AGE	Advanced glycation end-product
α_l	Relative phase for meridional order l
APS	Advanced Photon Source
BioCAT	Biophysics Collaborative Access Team
CML	Carboxymethyl lysine
CV	Coefficient of variation (also, %CV)
deH-HLNL	Dehydro-hydroxylysinonorleucine
deH-LNL	Dehydro-lysinonorleucine
dH_2O	Filtered, deionised water
DM	Diabetes mellitus
DOGDIC	3-deoxyglucosone-derived imidazolium cross-link
$\|F(l)\|$	Amplitude component of the structure factor for meridional order l
F_p, F_h, F_{ph}	Structure factors (combined amplitude and phase data) for protein, heavy atom h and protein + heavy atom, respectively
GAG	Glycosaminoglycan
GODIC	Glyoxal-derived imidazolium cross-link
h	Heavy atom (isomorphous replacement)
l	Axial lattice coordinate in reciprocal space/X-ray diffraction meridional order number
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulphonic acid
HHL	Histidino-hydroxylysinonorleucine
HK	Hydroxylysine (diagrams)
HLKNL	Hydroxylysino-ketonorleucine
HL-Pyr	Hydroxylysyl-pyridinoline
Hyl	Hydroxylysine (text)
Hyp	Hydroxyproline
ID	Insertion device
IIT	Illinois Institute of Technology
LKNL	Lysino-ketonorleucine
Lys	Lysine
LysPyr	Lysyl-pyridinoline
MMP	Matrix metalloproteinase
MODIC	Methylglyoxal-derived imidazolium cross-link
MOPS	3-(N-Morpholino)propanesulphonic acid
MTT	Mouse tail tendon

OHPyr	Hydroxylysyl-pyridinoline
PBS	Phosphate buffered saline
PG	Proteoglycan
PIC	Protease inhibitor cocktail
PTA	Phosphotungstic acid
$P(z)$	Axial one-dimensional Patterson function
RTT	Rat tail tendon
$\rho(z)$	Electron density along the fibrillar axis, as a function of z
SD	Standard deviation
TEM	Transmission electron microscope/y
TFA	Trifluoroacetic acid
T_m	Melting temperature (of collagen)
UA	Uranyl acetate
z	Axial (along collagen fibril) Cartesian coordinate in real space

Chapter 1: Introduction

1.1 Collagen

Collagen is a protein found in every multicellular organism within the animal kingdom. Both quantitatively and functionally, it is one of the most important proteins present in animals: it is the most abundant protein in mammals, constituting over a quarter of the total amount of protein present (Kavitha and Thampan, 2008).

The name *collagen* is derived from the Greek words meaning *to make or give birth to glue* - i.e. *kolla* + *gen* - (Stryer, 1981; Van der Rest and Garrone, 1991; Harper, 2008), and thus the name provides some insight into the protein's function. Characteristically, collagen is an insoluble, fibrous protein found in most tissues. In the extracellular matrix (ECM) it provides strength, fulfilling connective, structural, force transmission and other specialist roles throughout the body, including cell migration, angiogenesis and tissue repair (Kadler et al., 2007). It is found in such diverse tissues as tendons and ligaments, bone, skin, basement membranes, cartilage and blood vessel walls (Canty and Kadler, 2005). In addition, it is found in more specialist tissues, such as cornea (Meek and Holmes, 1983; Regini, Elliott and Hodson, 2004) and heart valve (Jimenez and Bashey, 1978; Balguid et al., 2007). Collagen is vital for the form and correct functioning of most tissues.

All collagen molecules consist of three polypeptide α-chains arranged about one another, at least in part forming a triple helical structure (Figure 1.1). The primary structure of each triple helical region of each of these polypeptides consists of repeating triplet amino acid sequences of Gly-*X*-*Y*, where *X* and *Y* are any amino acid other than glycine. The polypeptides' collagen sequences are rich in imino acid residues: "*X*" is commonly proline and "*Y*" is often hydroxyproline (Ramshaw, Shah and Brodsky, 1998), a modified version of proline found in few other proteins. It is produced by extensive posttranslational hydroxylation of "*Y*" prolines by the prolyl 4-hydroxylase enzymes (For a review of these, the reader is referred to Myllyharju (2003)), which has a specific requirement for ascorbic acid (vitamin C) (Myllyla, Majamaa and Gunzler, 1984). The resulting molecule is analogous to a rope of remarkable strength and yet, at the same time, potential for significant elasticity and flexibility. This is essential, for example, in the role of collagen in tendon. In tendon, energy produced by muscular force is stored efficiently as elastic strain energy, and transmitted to bone as kinetic

energy to effect locomotion and activity (Alexander and Bennet-Clark, 1977; Roberts et al., 1997; Silver, Freeman and Bradica, 2006).

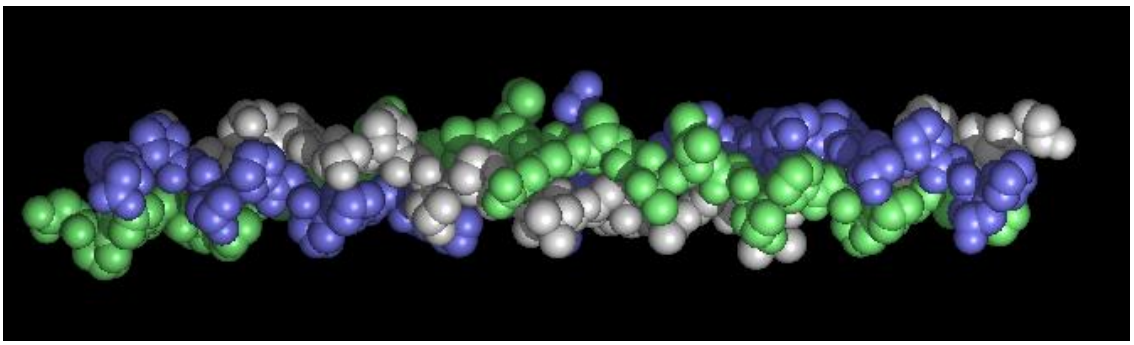


Figure 1.1 Space filling representation of the arrangement of the three constituent polypeptide α-chains in the triple helical ("collagenous") region of a collagen molecule. Each separate α-chain is represented in a different colour for clarity.
Image © 2003-2011 ZhuoYun "Esther" Zhuang and Dr. Kalju Kahn, Department of Chemistry and Biochemistry, UC Santa Barbara, taken from http://web.chem.ucsb.edu/~molvisual/prot_funct.html, accessed August 2013.

1.1.1 Types of collagen

It is misleading to represent collagen as a single protein: it is, in fact, a family of proteins. There are now known to be 28 different members of this protein group, which have been reviewed comprehensively (Gordon and Hahn, 2010). The most recently discovered member of the family was first reported and characterised by Veit et al. (2006). Each collagen type, by convention, is denoted using Roman numerals. Therefore they are types I to XXVIII.

The distribution, form and function of the collagen family of proteins is diverse and useful reviews illustrating this have been written previously (Myllyharju and Kivirikko, 2004; Gordon and Hahn, 2010). The collagens are classified according to form or function, generally fibril forming, fibril-associated collagens with interrupted triple helices (FACITs), network-forming, beaded filament-forming, transmembranous and endostatin collagens. The classifications of the collagens discovered to date together with an indication of their distributions and functions are presented in Table 1.1.

Collagen type	Classification	Distribution/remarks
I	Fibril-forming	Non-cartilaginous connective tissues – e.g. tendon, ligament, cornea, bone, skin
II	Fibril-forming	Cartilage, vitreous humour and nucleus pulposus
III	Fibril-forming	Co-distributes with collagen I, especially in embryonic skin and hollow organs
IV	Network-forming	Basement membranes
V	Fibril-forming	Co-distributes with collagen I, especially in embryonic tissues and in cornea
VI	Beaded filament-forming	Widespread, especially muscle
VII	Anchoring fibrils	Dermal-epidermal junction
VIII	Network-forming	Descement's membrane
IX	FACIT	Co-distributes with collagen II, especially in cartilage and vitreous humour
X	Network-forming	Hypertrophic cartilage
XI	Fibril-forming	Co-distributes with collagen II
XII	FACIT	Found with collagen I
XIII	Transmembrane	Neuromuscular junctions, skin
XIV	FACIT	Found with collagen I
XV	Endostatins	Located between collagen fibrils that are close to basement membranes; found in the eye, muscle and microvessels; a close structural homologue of collagen XVIII
XVI	FACIT	Integrated into collagen fibrils and fibrillin-1 microfibrils
XVII	Transmembrane	Also known as the bullous pemphigoid antigen 2/BP180; localised to epithelia; an epithelial adhesion molecule; ectodomain cleaved by ADAM proteinases
XVIII	Endostatins	Associated with basement membranes; endostatin is proteolytically released from the C-terminus of collagen XVIII; important for retinal vasculogenesis
XIX	FACIT	Rare; localised to basement membrane zones; contributes to muscle physiology and differentiation
XX	FACIT	Widespread distribution, most prevelant in corneal epithelium
XXI	FACIT	Widespread distribution
XXII	FACIT	Localised at tissue junctions – e.g. myotendinous junction, cartilage-synovial fluid, hair follicle-dermis
XXIII	Transmembrane	Limited tissue distribution; exists as a transmembrane and shed form
XXIV	Fibril-forming	Shares sequence homology with the fibril-forming collagens; has minor interruptions in the triple helix; selective expression in developing cornea and bone
XXV	Transmembrane	CLAC-P – precursor protein for CLAC (collagenous Alzheimer amyloid plaque component)
XXVI	Beaded-filament-forming	Also known as EMI domain-containing protein 2, protein Emu2, Emilin and multimerin domain-containing protein 2
XXVII	Fibril-forming	Shares sequence homology with the fibril-forming collagens; has minor interruptions in the triple helix; found in embryonic cartilage, developing dermis, cornea, inner limiting membrane of the retina and major arteries of the heart; restricted to cartilage in adults; found in fibrillar-like assemblies
XXVIII	Beaded-filament-forming	A component of the basement membrane around Schwann cells; a von Willebrand factor A domain-containing protein with numerous interruptions in the triple helical domain

Table 1.1 All the collagens identified so far (types I→XXVIII), with a summary of type and function for each. Republished with permission of The Company of Biologists Ltd., from Kadler et al. (2007), © 2007; permission conveyed through Copyright Clearance Center, Inc.

1.1.2 Fibril-forming collagens

These are types I, II, III, V, XI, XXIV and XXVII collagens (Table 1.1). Of these, types I, II and III collagens make up 80 – 90% of the collagen in the body (Kavitha and Thampan, 2008). The fibril forming collagens form the bulk of fibrous connective tissue and in essence, provide both the structural framework of tissues and act in an important mechanical role in the musculoskeletal system via formation of tendons (Silver et al., 2006) and most of the protein matrix of bone (Burger et al., 2008; Jiang et al., 2008), as well as articular cartilage (Pirttiniemi et al., 1996). The molecules of fibril-forming collagen have a long and uninterrupted major triple helical domain, which is typically more than 1000 amino acid residues in length (UniProt, 2012). Type I collagen molecules have been shown to readily associate spontaneously to form microscopic fibrils *in vitro* (Gross and Kirk, 1958; Li et al., 2009), which are similar to those present *in vivo* (Gobeaux et al., 2008). However, fibrillogenesis *in vivo* does not occur spontaneously, being regulated and controlled by a number of other proteins (Humphries et al., 2008) (for reviews of current knowledge of the process, see Kadler, Hill and Canty-Laird (2008); Banos, Thomas and Kuo (2008)). Interestingly, alongside other proteins and non-fibrillar collagens, type V collagen is important in regulation of formation of fibrils comprised primarily of type I collagen (Wenstrup et al., 2004; Wenstrup et al., 2006) and type XI collagen has the equivalent role in type II collagen fibrillogenesis (Fernandes et al., 2007).

Type I collagen is the most prevalent form of collagen in the body, being the predominant form in tendon, bone, skin (Stryer, 1981; Canty and Kadler, 2005) and cornea (Meek and Fullwood, 2001). Type II collagen is found in cartilage, along with type I collagen (Miller and Matukas, 1969; Pirttiniemi et al., 1996). Type III collagen is found closely associated with type I collagen (Cameron et al., 2002) to varying degrees, in different tissues and at different stages in development. It is found in its highest concentrations in tissues exhibiting more elastic properties, such as skin (Silver and Birk, 1984; Liu et al., 1997) blood vessel walls (Farquharson and Robins, 1989; Gong et al., 2008) and intestinal tissue (Liu et al., 1997). Type III collagen is the second most abundant collagen type in the body after type I (Cameron et al., 2002).

1.1.3 Structure of type I collagen molecules

The defining attribute that unites all “collagens” is the presence of one or more collagenous or triple helical regions. In most collagens, these have non-collagenous

interruptions and the nature and distribution of these across the family have been reviewed recently (Gordon and Hahn, 2010).

Type I collagen is the "purest" form of collagen, inasmuch as almost the entire molecule consists of one long triple helical region without interruptions (Stryer, 1981; Silver et al., 2006; Shoulders and Raines, 2009) of the conformation shown in Figure 1.1. This constitutes a rope-like rod of over 1000 amino acid residues in length. It is ~300 nm long and 1.5 nm in diameter (Miller and Wray, 1971; Silver et al., 2006; Shoulders and Raines, 2009). Only the N- and C-terminals or telopeptides do not adopt the triple helical conformation, but these constitute just 3% of the molecule's amino acid content (UniProt, 2012).

The repeating Gly-Pro-Hyp (Hyp is hydroxyproline) is the commonest Gly-*X*-*Y* triplet, having a relative frequency of 10.5% when analysed across the triple helical sequences in the α-chains of all collagens. In total, 28% of amino acid residues in the "*X*" positions of these triplets are proline and 38% of those in the "*Y*" positions, are hydroxyproline (Ramshaw et al., 1998). Of the remaining 399 possible triplet compositions, only 24 are observed across the collagens with a frequency of >1% and only 49% are seen at all in the fibrillar collagens. The reason for the level of conservation observed is the stability of the collagen helix. Ramshaw et al. (1998) established that there was a clear relationship between frequency of Gly-*X*-*Y* triplets in triple helical regions and the collagen melting temperature (T_m - the temperature at which half the triple helical structure is lost) associated with them. Type I collagen is assembled from two different polypeptide chains, i.e. an $\alpha 1$–chain and an $\alpha 2$–chain ($\alpha 1$(I) and $\alpha 2$(I), respectively, the Roman numeral in parentheses denoting the collagen type), the composition being $[\alpha 1(I)]_2\alpha 2(I)$ in healthy individuals (Stryer, 1981). There are subtle but important differences between the $\alpha 1$(I) and $\alpha 2$(I) chains (Hofmann, Fietzek and Kuehn, 1980) and it has been shown that the respective 2:1 ratio in the stoichiometry is important for correct form and function of type I collagen fibrils (McBride Jr et al., 1997).

The collagen triple helix consists of its three constituent α–chains each rotated into a left-handed helix form and these in turn are wound around one another into the characteristic right-handed triple helix. Within this arrangement the left-handed rotation of the each α–chain undergoes one turn every three amino acid residues. This is the significance of a glycine being present every third amino acid. The triple helices are associated such that the glycine's small side chain - a hydrogen atom – is pointing in towards the centre of the triple helix therefore not sterically disrupting the close

association that exists between the three α–chains in forming this superhelix (Figure 1.2) (Van der Rest and Garrone, 1991; Myllyharju and Kivirikko, 2004).

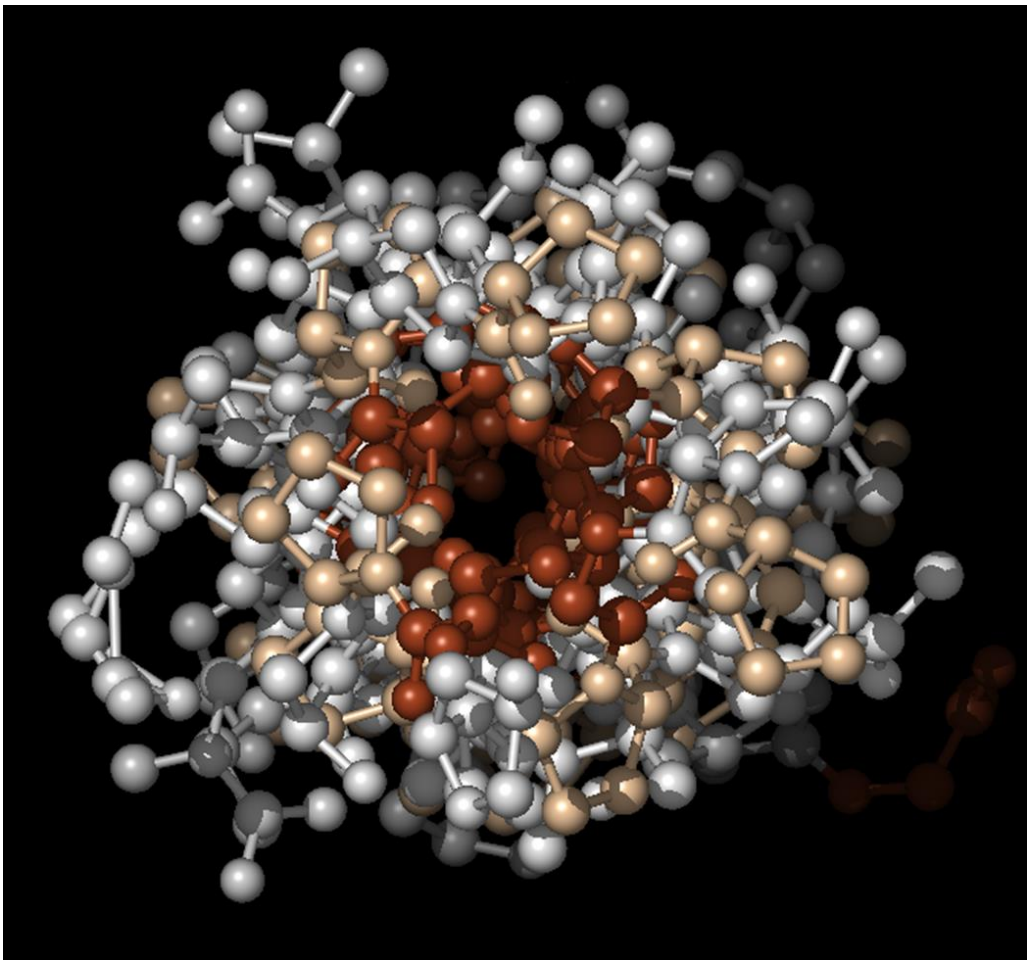


Figure 1.2 View of a space-filling molecular model of a synthetic collagen triple helix end-on, with glycines on the inside (brown). The *X* & *Y* residues – i.e. prolines (wheat coloured), and hydroxyprolines and isoleucines (white/gray) – are on the outside. Image © 2003-2011 ZhuoYun "Esther" Zhuang and Dr. Kalju Kahn, Department of Chemistry and Biochemistry, UC Santa Barbara, taken from http://web.chem.ucsb.edu/~molvisual/prot_funct.html, accessed August 2013.

The side chains of residues in the *X* and *Y* positions point outwards, and so will not sterically impair the formation of the triple helical structure. The single strands' helical structure is largely dictated and stabilized by their high imino acid content, which produce a conformational and steric-repulsion effect (Stryer, 1981). This in turn ensures the glycines' correct orientation to facilitate the assembly of the triple helix. This entity is held together by transverse hydrogen bonding between the peptide NH groups of glycines and peptide CO groups on the other two strands, and Van der

Waal's interactions between amino acid residues on adjacent strands (Stryer, 1981; Shoulders and Raines, 2009).

Using T_m values, it has been ascertained that the presence of hydroxyproline is even more effective in maintaining the collagen triple helix's structure at higher temperatures than proline, hence the modification (Ramshaw et al., 1998). The importance of this hydroxylation, especially in warm-blooded animals' tissues, can therefore be appreciated.

1.1.4 Collagen fibril structure

1.1.4.1 Arrangement of collagen molecules within fibrils

Fibrillogenesis *in vivo* is a complex and involved process that has been reviewed comprehensively elsewhere (Canty and Kadler, 2005; Banos et al., 2008). The resulting collagen fibrils can be hugely variable in both length and diameter, depending on both the tissue-source of the fibrils and the stage of development. They can vary between a few microns in length during early *in utero* development (Birk et al., 1990; Starborg et al., 2009) to a matter of millimetres (Canty and Kadler, 2005). Similarly, they can vary in diameter between a few nanometres to ~ 500 nm in mammals, depending on tissue and stage of development (Meek and Leonard, 1993; Canty and Kadler, 2005; Humphries et al., 2008; Li et al., 2013). The *in vivo* process of arranging collagen molecules into fibrils appropriate to the tissue involved appears to require interaction with a variety of other molecules, such as non-fibrillar collagens, proteoglycans, integrins and fibronectin, for the purposes of nucleation, organisation and regulation of the process. This has been reviewed in the literature (Canty and Kadler, 2005; Kadler et al., 2008), though there is still clearly much to be elucidated about the complex mechanisms involved.

The ordered arrangement of individual collagen molecules along fibrils was first deduced by Hodge and Petruska (1963). Collagen molecules are arranged along the fibril with the same orientation and virtually end-to-end, but with a gap between adjacent ends. Neighbouring axial rows of monomers are offset by approximately a quarter length relative to one another giving rise to the so-called "quarter stagger" arrangement. Therefore across the fibril, there are five different possible axially offset positions that can be occupied by these end-to-end rows of molecules. The structure is illustrated in Figure 1.3A. This gives rise to a regular, distinctive banding pattern as

seen by transmission electron microscopy (TEM) when collagen fibrils are “negatively stained” (Valentine and Horne, 1962), for example with sodium phosphotungstate at neutral pH (Figure 1.3B). The alternating dark and light bands, dark where heavy-metal stain has infiltrated regions of low packing density of collagen in the “gap” regions and light where the stain is excluded by higher packing density in the “overlap” regions, show a remarkably constant length or “*D*-period”. One dark+light band shows a *D*-period of ~67 nm (Bear, 1944; Hodge and Petruska, 1963) although this distance can vary in collagen from different tissues and under different conditions (Bear, 1944; Meek and Holmes, 1983). This is equivalent to a translation distance of 234 amino acids (Meek, Chapman and Hardcastle, 1979; Hofmann et al., 1980).

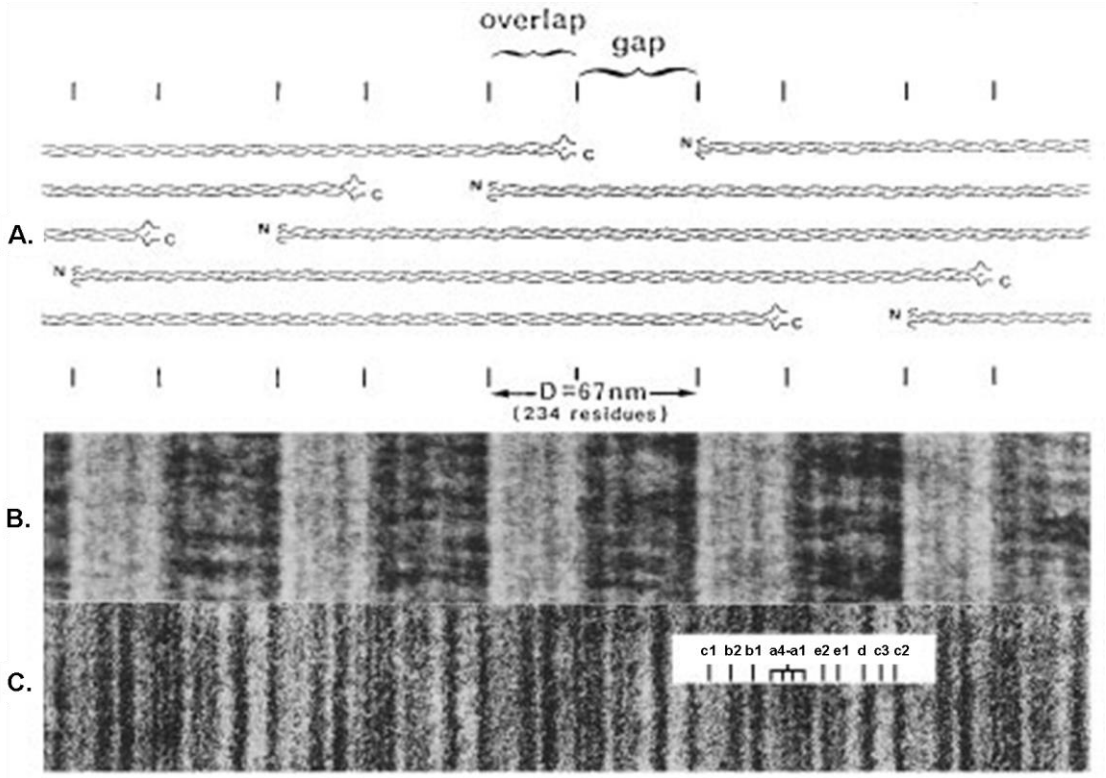


Figure 1.3 A. A diagrammatic illustration of the regular offset arrangement of type I collagen monomers in a fibril. The gap-and-overlap regions together, cover a distance (*D*-period) of approximately 67nm. **B.** A reconstituted bovine collagen fibril, negatively stained with 1% sodium phosphotungstate, pH 7.0, and viewed under the TEM. **C.** Similar fibril to **B**, positively stained, first with 1% phospotungstic acid at pH 3.4, then 1% uranyl acetate, pH 4.2. **A**, **B** and **C** are aligned. With both staining methods, regular banding patterns can be seen clearly. The lines in the overlaid panel in **C** shows 12 commonly-resolved positive staining bands, they being (from left to right) *c1, b2, b1, a4, a3, a2, a1, e2, e1, d, c3, c2*. Reproduced with permission, from Kadler K E, Holmes D F, Trotter J A and Chapman J A, (1996) *Biochem J* **316(1)** 1-11. © 1996 The Biochemical Society.

The same regularity and periodicity of banding can also be seen with “positive staining” (Figure 1.3C). In this case, the pH employed is lower and the principle of staining more complex. Whereas with negative staining, regions of high stain uptake exist where there is lower molecular density, i.e. it is based on physical accessibility to the stain, positive staining has an electrostatic basis. Staining ions interact with chemical groups of the opposite charge and so the banding pattern relates to distribution of the side-chain(s) of interest. Use of phosphotungstic acid (PTA) results in positive staining of positively-charged lysine and arginine side-chains along fibrils by the negatively charged phosphotungstate anion (Tzaphlidou, Chapman and Meek, 1982b; Hadley, Meek and Malik, 1998). Conversely use of uranyl acetate (UA) for positive staining results in visualisation under the TEM, of regions along the fibril containing comparatively high concentrations of acidic amino acids (aspartate and glutamate) due to binding of the positively charged uranyl ion (Tzaphlidou, Chapman and Alsamman, 1982a). The phosphotungstate cannot penetrate far into the fibril due to its comparatively large size. The uranyl ion is smaller therefore tends to be able to move further into the fibril. Therefore staining by uranyl ions alone tends to show the better definition (Meek and Holmes, 1983). Although the relative intensities of the bands are different for the two stains, the peak positions are the same using both (Hodge and Schmitt, 1960; Meek and Holmes, 1983) thereby demonstrating co-localisation of the basic and acidic amino acids along the fibril. This discovery and the realisation that these patterns were an effective summation of the positions of the lysines and arginine, and aspartates and glutamates in the individual molecules across the fibril, were important in Hodge and Petruska being able to construct their model of collagen molecular arrangement along fibrils (Hodge and Petruska, 1963). This is indicative of the presence of different regions along the collagen *D*-period, which vary in flexibility. The proline/hydroxyproline-containing regions tend to be more rigid, whereas other regions are more amenable to bending and stretching. There are up to twelve positive staining bands visible under the TEM and these broadly correspond to these flexible regions (Silver et al., 2001; Silver, Horvarth and Foran, 2002). For a more detailed review of positive and negative staining and factors affecting stain uptake by collagen, the reader is referred to Chapman et al. (1990).

The packing of collagen molecules across the fibril is less regular than along it. It is often referred to as pseudohexagonal, being characterised by each collagen molecule being surrounded by six others (Miller and Tocchetti, 1981; Wess et al., 1998b). Naturally there is a hole in the gap region and this is therefore less ordered across the fibril (Wess et al., 1998a; Orgel et al., 2001).

1.1.4.2 Enzymic cross-linking

Certain lysine residues of nascent collagen α-chains are post-translationally hydroxylated by a family of enzymes called lysyl hydroxylases, of which three isoforms have thus far been identified (Myllyla et al., 2007). As in the case of the prolyl 4-hydroxylases, these enzymes are ascorbate-dependant (Myllyla et al., 1984). The resulting hydroxylysine (Hyl) residues are important for the later formation of specifically-located extracellular collagen cross-links.

Following the assembly of the collagen molecules into fibrils, intermolecular cross-linking takes place. This process serves to further stabilise the structure, increasing fibrillar strength and preventing excessive slippage between collagen molecules (Sasaki and Odajima, 1996b; Avery and Bailey, 2005). This is a very specific mechanism of cross-linking and is initiated by lysyl oxidase enzymes, which oxidatively deaminates the ε-amino group of a single telopeptide lysine or hydroxylysine residue to form the corresponding aldehyde (Bailey, Paul and Knott, 1998). Lysyl oxidase acts on lysines in the telopeptide regions only when it is bound to a conserved sequence, i.e. -Hyl-Gly-His-Arg- in the triple-helical region of an adjacent collagen molecule (Reiser, McCormick and Rucker, 1992a; Bailey et al., 1998). Following this, cross-link formation is a spontaneous condensation reaction brought about by the reaction of telopeptide lysyl or hydroxylysyl aldehydes with a lysine or hydroxylysine in the triple helix of a neighbouring molecule to form a dimeric Schiff base double bond. Four such species can be formed (Figure 1.4): telopeptide lysyl aldehydes react with triple-helix lysine or hydroxylysine residues to form the aldimines dehydro-lysinonorleucine (deH-LNL) and dehydro-hydroxylysinonorleucine (deH-HLNL); hydroxylysine aldehydes can react with triple-helix lysine or hydroxylysine residues to form the ketoimines lysino-ketonorleucine (LKNL) and hydroxylysino-ketonorleucine (HLKNL) after a rearrangement – the Amadori rearrangement – which gives rise to these stable cross-links (Bailey et al., 1998; Bailey, 2001).

With maturation, these divalent cross-links spontaneously react further with species on other molecules, to give more reduced, stable trivalent cross-links (Figure 1.5). The other species involved are histidine residues, and other lysyl- and hydroxylysyl-aldehydes. Histidine reacts with deH-HLNL to form histidino-hydroxylysinonorleucine (HHL). The reaction of HLKNL or LKNL with a further telopeptide hydroxylysyl-aldehyde residue results in the formation of a hydroxylysyl-pyridinoline (HL-Pyr) or lysyl-pyridinoline (LysPyr) cross-links, respectively (Willett et al., 2007). HL-Pyr and

LysPyr (sometimes referred to as pyridinium cross-links) are more stable and non-reducible than their immature, divalent predecessors (Fujimoto and Moriguchi, 1978; Avery, Sims and Bailey, 2009).

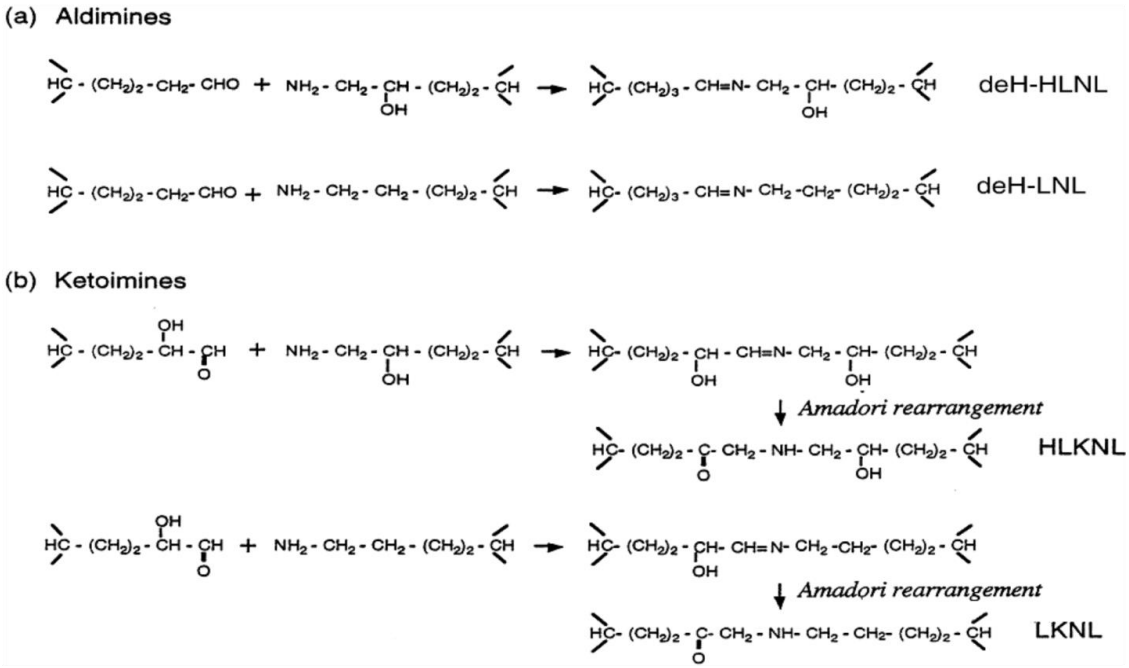


Figure 1.4 The telopeptide lysine/hydroxylysine aldehyde reactions with lysine/hydroxylysines in the collagen triple helix of a neighbouring molecule in a fibril, to form divalent immature cross-links which are stable under physiological conditions. Reprinted from Bailey et al. (1998), © 1998, used with permission from Elsevier.

These immature and pyridinium cross-links occur in specific locations, linking molecular ends (non-helical telopeptides) to helical loci. These are highly conserved sites between species. They form between the lysyl/hydroxylysyl aldehydes in the α1- and α2-chain N-telopeptides and lysines/hydroxylysines at helix positions 930 in the α1-chain or 933 in the α2-chain, and between the lysyl/hydroxylysyl aldehydes in the α1-chain C-telopeptides and lysines/hydroxylysines in and lysines/hydroxylysines in position 87 in triple helix on the α1- or α2-chain (Reiser et al., 1992a; Hanson and Eyre, 1996). The HHL forms between the deH-HLNL formed between a C-terminal telopeptide lysyl aldehyde and a hydroxylysine at position 87 in the triple helix region of an α1-chain, and histidine 92 in the α2-chain (Mechanic et al., 1987; Yamauchi et al., 1987; Reiser et al., 1992a).

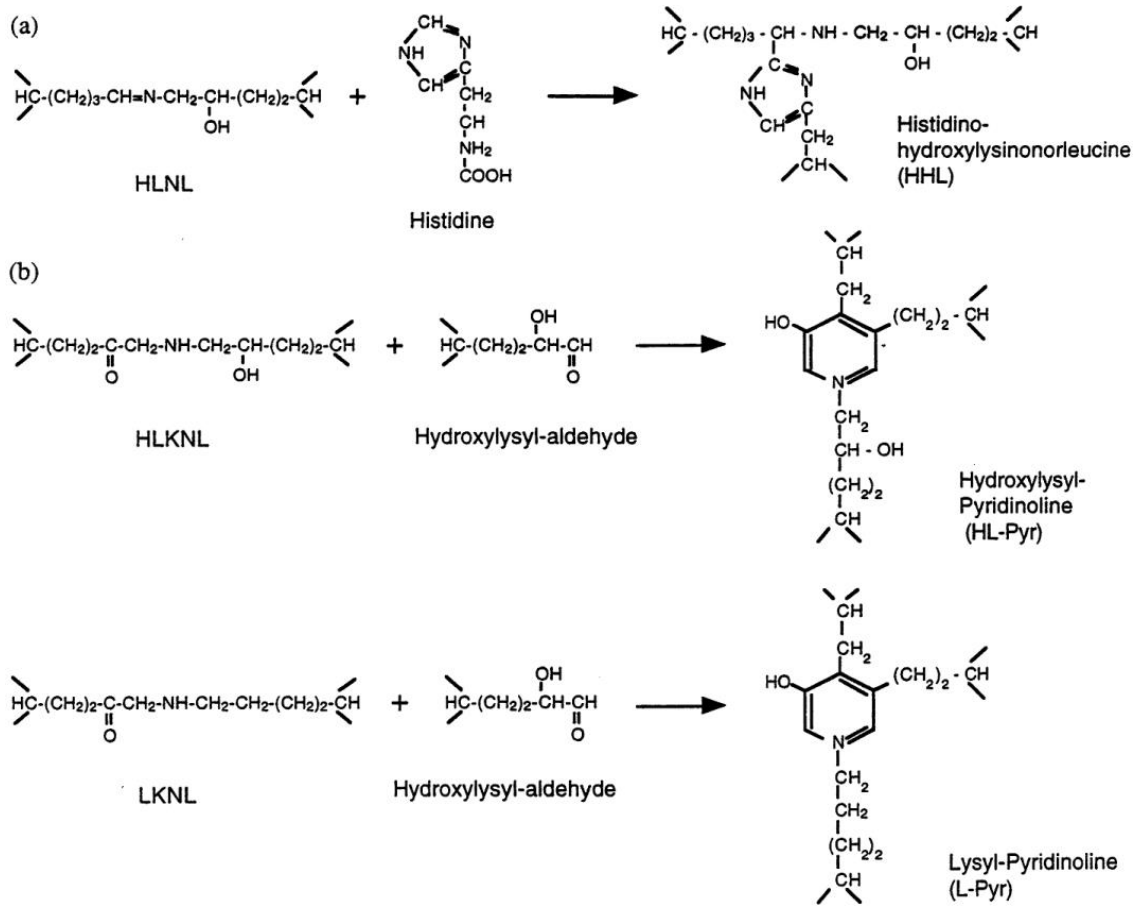


Figure 1.5 Formation of the characterised mature cross-links: (a) HHL derived from the divalent Schiff base HLNL and triple helical histidine and (b) the pyridinolines derived from the divalent keto-imines and the telopeptide hydroxylysyl aldehyde. Reprinted from Bailey et al. (1998), © 1998, used with permission from Elsevier.

The types of cross-links found are more reliant on tissue type, location and maturity than on species. Before birth and in the early months or years of life, the immature cross-links predominate before reacting to form the trivalent, mature cross-links. The nature of the cross-links formed and their rate of formation is heavily influenced by lysyl hydroxylase and lysyl oxidase activities in different tissues (Bailey, 2001; Avery and Bailey, 2005).

In tendon there is variability in hydroxylation depending on the particular tendon being considered. Varying proportions of deH-HLNL and HLKNL are formed, leading to both HHL and pyridinoline cross-links predominating with increasing age. Therefore tendon's mechanical strength and thermal stability tends to increase with age (Avery and Bailey, 2005; Willett et al., 2010). Curiously, in rat tail tendon (RTT) tissue hydroxylation is very low and so deH-HLNL is initially formed. However, though this decreases with age and the tendon continues to strengthen, it does not react with

histidine to form HHL. Therefore it is possible that a cross-link that has yet to be identified is formed (Bailey, 2001; Avery and Bailey, 2005). In comparison with a load-bearing tissue such as human patellar tendon, tail tendon from young rats has different mechanical properties showing it to be less mechanically resilient to applied stress and more prone to mechanical rupture. This is associated with a markedly lower level of mature cross-links and higher level of immature cross-links in the RTT as compared with the human tissue (Svensson et al., 2013). Therefore clearly, it can be considered that the pattern and profile of enzymic cross-linking of tissue is attuned to the requirements of that tissue.

1.1.5 Tendon: a model type I collagen-containing tissue

1.1.5.1 The role of type I collagen in the function of tendon

Tendon tissue of various types is frequently used as a source of collagen for studies and it is used in this project. In the fibres, there is a high content of relatively pure type I collagen fibrils (Duance et al., 1977; Chanut-Delalande et al., 2001; Silver et al., 2001; Silver, Freeman and Seehra, 2003) and therefore there is scope for study of aspects of the biology and functionality of this model of collagen including biochemistry, mechanical properties and structure, either by use of entire tendon (Wess et al., 1993; McBride Jr et al., 1997; Willett et al., 2007; Svensson et al., 2013) or collagen extracted from this tissue, but with a requirement for minimal purification (Reiser, Amigable and Last, 1992b; Ait-Belkacem et al., 2012).

The hierarchy of tendon structure is shown in Figure 1.6. Within the tendon, the type I collagen performs an energy transmission role and possesses viscoeleastic properties enabling it to do this. The tendon is sufficiently elastic to enable it to recover from a strain of up to 5% elongation and be able to return to its original length (Screen et al., 2004). The bases for collagen's elasticity in tendon are **1**. tendon crimps (Screen et al., 2004; Franchi et al., 2007a), **2**. molecular elongation (Mosler, Folkhard and Knorzer, 1985; Sasaki and Odajima, 1996ab) and **3**. molecular slippage within fibrils (Mosler et al., 1985; Sasaki and Odajima, 1996a).

Tendon crimps are clearly visible in fibre bundles using light microscopy and appear to have their basis in fibrillar crimps, visible by electron microscopy, in which fibrils exhibit angular turns along their length (Franchi et al., 2007a). Most of the elongation, ~4%, during loading of tendon occurs during the initial crimp straightening and concomitant

collagen fibre straightening and realignment in the direction of the applied tensile force (Screen et al., 2004). The tendon crimps are therefore considered to be analogous to shock absorbers (Franchi et al., 2007a). Following this phase, the fibres are most probably under direct tension and any further elongation is as a result of slippage of hierarchical elements of the tendon structure or extension of the collagen components (Screen et al., 2004). Regarding crimps in the structure of collagen, it should also be noted that from X-ray diffraction studies of collagen fibrillar structure, that the path of molecules is not straight through the microfibril (the smallest, pentameric crystalline unit of fibril structure that can be considered) (Fraser, MacRae and Miller, 1987), but rather slightly meandering and inclined away from being parallel to the fibril by up to 10^{o} (Wess et al., 1995; Orgel et al., 2006). Orgel et al. (2006) considered that this molecular crimp may itself contribute to the properties of the fibril in absorbing and transmitting mechanical force.

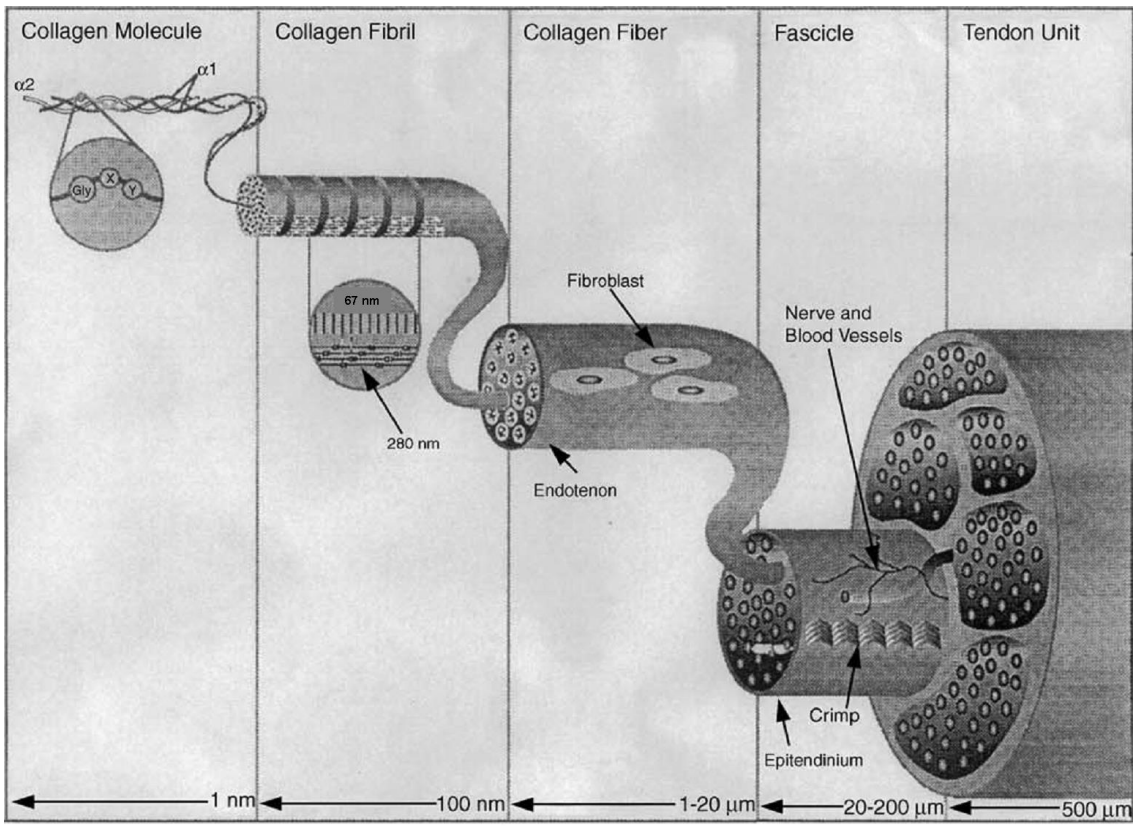


Figure 1.6 Heirarchy of tendon showing the relationships of structure and scale between collagen molecules, fibrils, fibers, fascicles and tendon units. Reprinted with minor adaptation, from Silver et al. (2003), © 2003, used with permission from Elsevier.

Molecular elongation is likely to have a basis, both in extension of the triple helix via the flexible regions (Silver et al., 2001; Silver et al., 2002) and the stretching of the telopeptides (Mosler et al., 1985). As described in section 1.1.4.2, the telopeptide regions are cross-linked to the triple helices in a very specific way, and so this molecular elongation would be inextricably linked with molecular slippage. The greater the degree of cross-linking, the greater the tensile strength of the fibril and the extent to which a tensile force would be shared out between the molecules within the fibril (Sasaki and Odajima, 1996b; Avery and Bailey, 2005).

Overall, the stress-strain curve for collagen molecules and fibrils exhibit Hookean properties. Sasaki et al. (1996a) found, for example, that the Young's modulus of 20 month old bovine tendon collagen molecules is three times that of the fibrils. The stress-strain curve is more complex for tendon, with an initial "toe" (or flatter, low-stress elongation) region corresponding to the initial ~4% extension resulting from the crimp-straightening phase (Gathercole and Keller, 1991; Screen et al., 2004). Beyond this, a linear stress-strain relationship exists with a stress-strain gradient approximately equal to that of collagen fibrils (Sasaki and Odajima, 1996a).

Although dependent on location and function, an example of the importance of tendon's mechanical and elastic properties in transmission of motive force between muscle and bone was illustrated by Roberts et al. (1997) in running turkeys. The ability of the tendon to store elastic energy allows the muscle to produce the necessary force whilst doing the minimum amount of work.

1.1.5.2 Other components of tendon

Informative reviews have been published elsewhere regarding the structure and composition of tendon tissue (Kannus, 2000; Franchi et al., 2007b).

Water is the quantitatively-most important constituent of tendon tissue, contributing approximately 60% of its entire weight (Cameron, Lanctot and Fullerton, 2012), with collagen constituting approximately 80% of the dry weight (Willett et al., 2007) (Kannus et al. (2000) cite between 65% and 80%, and this is likely to vary with tendon source and developmental stage). It is nonetheless important to appreciate that there are other constituents of tendon tissue, which are important for its function. These are:-

1. Tendon cells or "tenocytes" – 90-95% of these are fibroblasts (Kannus, 2000), which are the cells that lay down collagen fibrils in tendon (Canty et al., 2004). Other cells

present include chondrocytes (5-10% of the total) at pressure points along the tendon plus other cells associated with, for example, the tendon vasculature (Kannus, 2000).

2. Proteoglycans (PGs) – These make up <1% of the dry weight of most tendons (Vogel, 2004). PGs consist of a protein core component, with one or more glycosaminoglycan (GAG) chains attached. They fulfil several important and distinct functions within tendon, most especially regulation of fibrillogenesis (Scott, 1996; Banos et al., 2008) and maintenance of the viscoelastic properties of the collagen components of tendon (Pins et al., 1997) and more detailed reviews of their metabolism and functions have been published elsewhere (Yoon and Halper, 2005; Rees, Dent and Caterson, 2009). By their chemical nature, PGs are very hydrophilic and therefore have an important role in keeping the collagenous tissue adequately hydrated (Kannus, 2000). Some of them are non-covalently attached to the surface of collagen fibrils (Scott, Orford and Hughes, 1981) and have a direct involvement in the arrangement and spacing of the fibrils (Scott, 1996). The most prevalent PGs in tendon are decorin and biglycan, which have the GAGs dermatan sulphate and chondroitin sulphate attached (Derwin et al., 2001). The collagen binding PGs tend to bind to specific sites along the fibril (Scott, 1996). Decorin's binding sites have been mapped to the *d* and *e1-e2* positive staining TEM bands along the *D*-period (Di Lullo et al., 2002; Sweeney et al., 2008). Given the homology between decorin and biglycan (Iozzo and Murdoch, 1996; Scott, 1996), biglycan is likely to bind to the same sites (Schonherr et al., 1995).

3. Other small organic molecules – there is wide biological acceptance that the concentration of glucose in the extracellular fluid is comparable with that in the blood (plasma glucose is typically around 5 mmol/l in healthy human subjects). Cells have a requirement for glucose as a primary source of energy.

4. Inorganic components – various ions are known to be present in significant quantities in tendon. Sodium, potassium and chloride ions are present though at least between the collagen molecules there is selective exclusion of these, most probably on the basis of size and associated hydration. Therefore the concentrations in tendon are lower than in the extracellular fluid (Cameron et al., 2012).

The various hierarchical structural tiers of tendon tend to be surrounded by connective tissue sheaths, which vary but are broadly named (with some variation) the paratenon, epitenon and endotenon. These help reduce friction between components of the tendon and between tendon and other structures. They also carry the nerves and

vasculature within the tendon (for reviews giving more information about these, see Kannus (2000) and Franchi et al. (2007b)).

The structure of type I collagen goes beyond being directly related to its structural and mechanical purpose. Different regions of the collagen molecule contain specific sites for interactions with various binding partners that can be found within tendon, over fifty of which have thus far been identified (Di Lullo et al., 2002; Sweeney et al., 2008). These include mediators of collagen-cell interactions (Goldberg and Burgeson, 1982; Reigle et al., 2008), proteolgycans (Reigle et al., 2008) and other proteins such as cell-bound integrins and secreted fibronectins that are involved in organising fibrillogenesis (Kadler et al., 2008).

1.2 Glycation of proteins

Glycation is the non-enzymatic reaction between monosaccharide sugars and the amino groups in proteins, nucleic acids and lipids (Singh et al., 2001). It is distinct from enzymic glycosylation inasmuch as it tends to be relatively non-specific and harmful. As long ago as 1912 Louis Camille Maillard first noted and studied the browning reaction, which occurs when sugar molecules react with protein molecules (Maillard, 1912). Hence, the term “Maillard reaction” has become synonymous with protein glycation. There are numerous review articles on the subject (Brownlee, 1995; Singh et al., 2001; Cho et al., 2007; Capote and Sanchez, 2009). The process of protein glycation from a chemical-conceptual point of view is relatively simple, and can be divided into three stages (Cho et al., 2007; Capote and Sanchez, 2009) (Figure 1.7):-

1. The initial stage, where the carbonyl group of the sugar reacts with an amino acid amine group, resulting in an intact sugar group attachment to the protein.

2. The intermediate stage, in which further fragmentation of the sugar-derived group occurs giving rise to further, smaller carbonyl group-containing reactive intermediates.

3. The late stage, in which various further reactions take place forming advanced glycation end-products (AGEs).

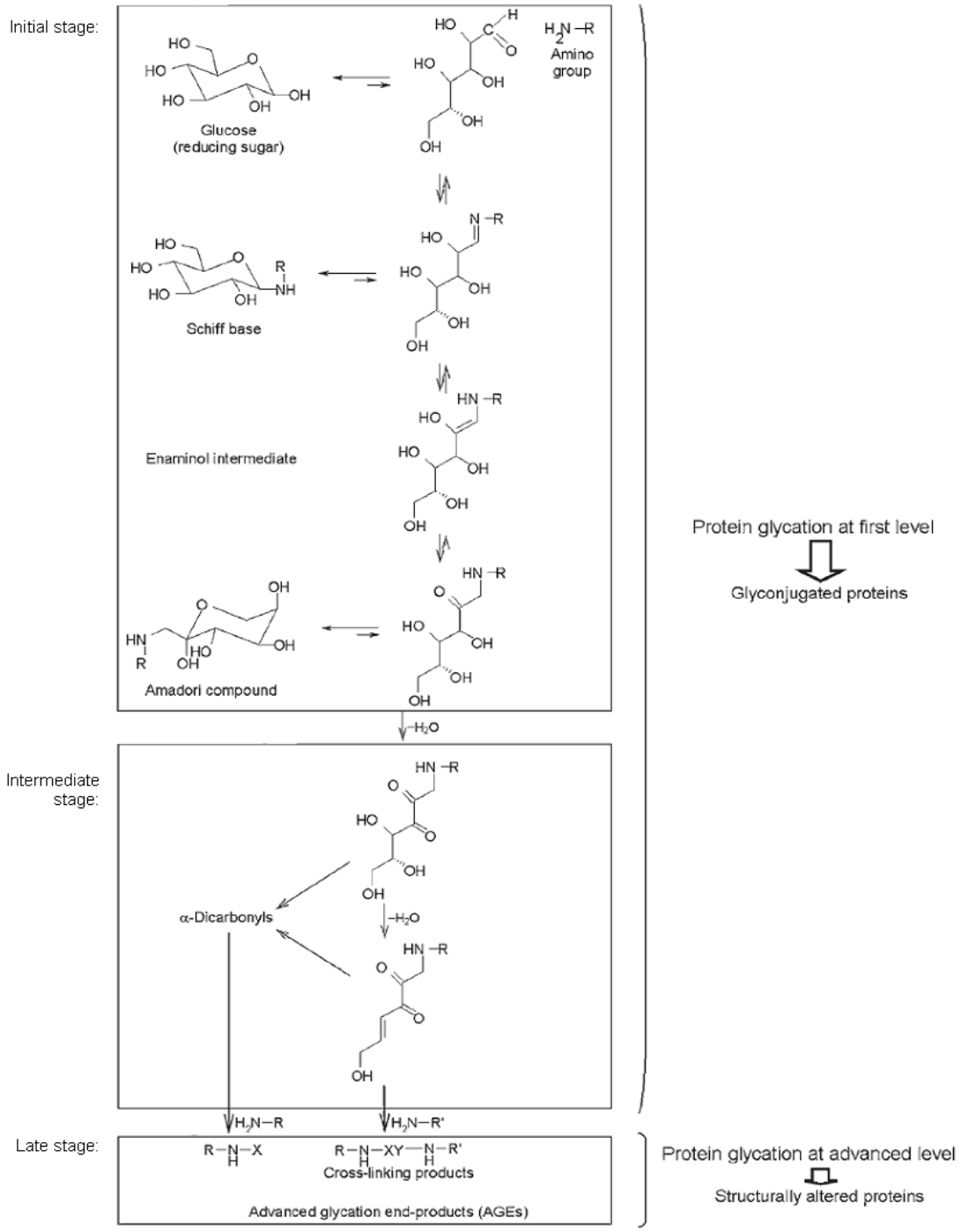


Figure 1.7 A schematic diagram of the Maillard reaction, the non-enzymatic glycation of a protein (-R) by a glucose (monosaccharide sugar). Note that it is the acyclic form of glucose (or other sugar) that reacts. "X" represents the products of the reactions of α-dicarbonyls formed during the intermediate fragmentation stage, with other amino groups. "XY" represents cross-links formed with other amino acids (modified or not) during later glycation reactions. Reprinted with minor adaptations, from Capote and Sanchez (2009), © 2009, used with permission from John Wiley and Sons.

1.2.1 Initial stage

The first stage of glycation occurs as a result of the nucleophilic attack by the N-terminal amine group of a polypeptide or protein or the ε amino groups in lysines on the carbonyl group of the sugar, to form a Schiff base. This is entirely reversible (Cerami, Vlassara and Brownlee, 1985). It is important to note that although pentose and hexose sugars – including glucose – exist in a predominantly cyclic form in solution, a small proportion of each sugar also exists in a straight chain or “acyclic” form. This is typically < or <<1% of the total. It is this acyclic form that is reactive, as this possesses the reactive carbonyl group (Bunn and Higgins, 1981). Protein glycation reaction appears to be first order with respect to sugar concentration under physiological conditions (Eble, Thorpe and Baynes, 1983; Brownlee, 1995; Vrdoljak et al., 2004).

Following Schiff base formation, progression occurs via rearrangement of the initial adduct to form the Amadori compound or product (see Figure 1.7), which is more stable (Cho et al., 2007; Capote and Sanchez, 2009).

1.2.2 Intermediate stage

Although relatively stable, over time, the Amadori product reacts further. The Amadori rearrangement is entirely reversible, >75% reverting back to the sugar and free amine, but fragmentation of the Amadori product can also occur in an oxidative or non-oxidative environment, giving rise to a number of compounds (Zyzak et al., 1995). These include the α-dicarbonyl compounds glyoxal, methylglyoxal and 3-deoxyglucosone (3-DG) (Thornalley, Langborg and Minhas, 1999) and various smaller sugars (pentoses and tetroses) (Zyzak et al., 1995).

1.2.3 Late stage

The α-dicarbonyl species formed are very reactive and go on to react with other N-terminal and basic amino acid groups, and through a series of further reactions and rearrangements, a potentially large number of stable, heterogeneous AGEs are formed (Cho et al., 2007; Capote and Sanchez, 2009). Examples of these are carboxymethyllysine (CML) (Ahmed, Thorpe and Baynes, 1986), carboxymethylarginine (CMA) (Iijima et al., 2000), pentosidine (Sell and Monnier, 1989), glucosepane (Biemel, Alexander Friedl and Lederer, 2002; Sell et al., 2005), and others such as MOLD (methylglyoxal-lysine dimer) (Odani et al., 1998), glyoxal-

lysine dimer (GOLD) (Kitamura et al., 2011), 3-deoxyglucosone-derived imidazolium cross-link (DOGDIC), methylglyoxal-derived imidazolium cross-link (MODIC) and glyoxal-derived imidazolium cross-link (GODIC) (Biemel et al., 2002; Sell et al., 2005). Except for CML and CMA these are all cross-links between two amino acid residue side-chains within proteins. GOLD and MOLD are lysine-lysine cross-links, whereas pentosidine, glucosepane DOGDIC, MODIC and GODIC are lysine-arginine cross-links (Sell et al., 2005; Cho et al., 2007) (See Figure 1.8).

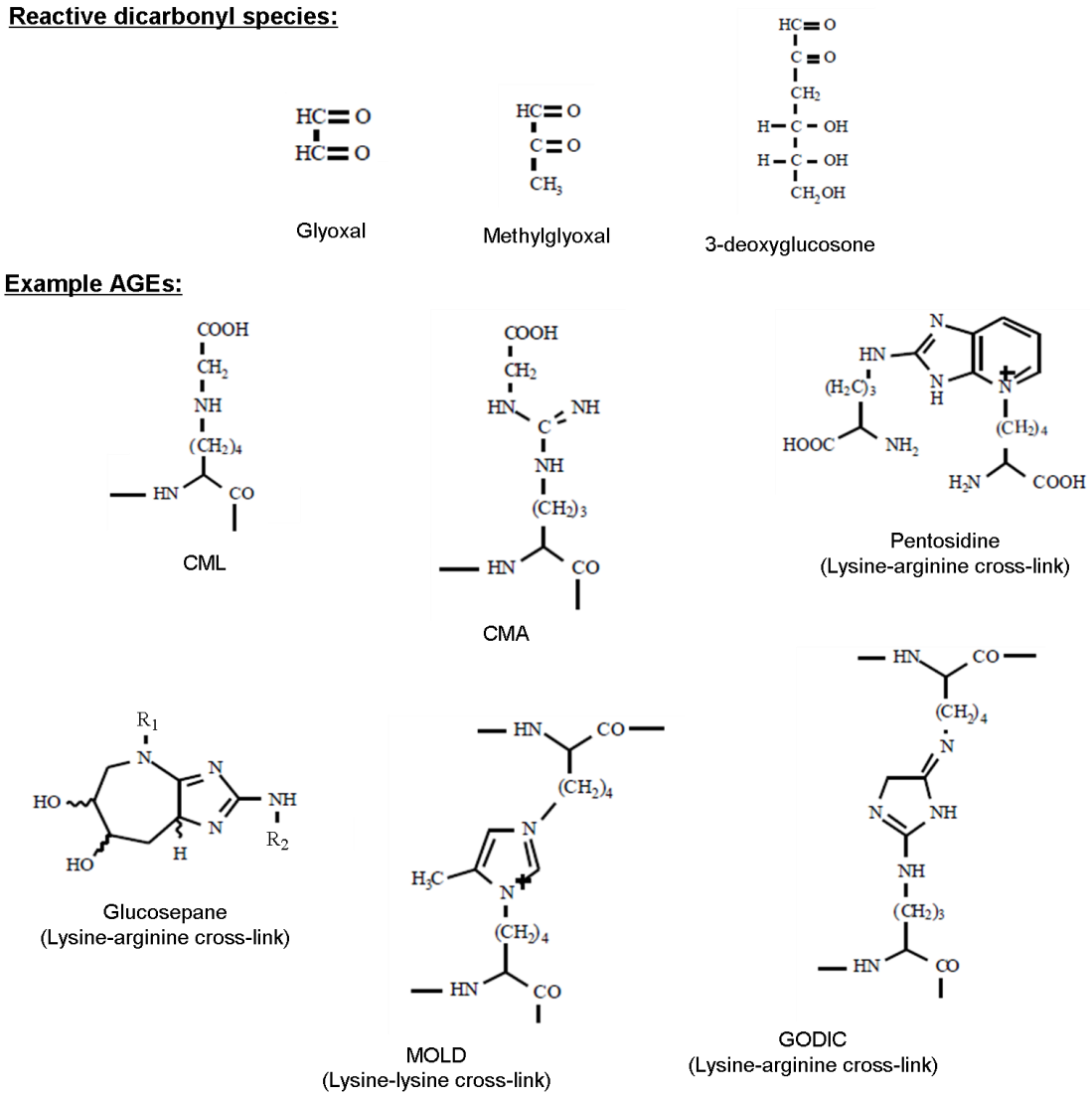


Figure 1.8 Examples of some of the more studied products of glycation in proteins. Shown are reactive dicarbonyl species that result from fragmentation of the rearranged attached post-Amadori sugar adduct, and various AGEs formed as a result either of further attack of these dicarbonyls or the rearranged sugar adducts on other basic residues. R_1 = $-(CH_2)_4-CH(NH_4^+)-COO^-$, R_2 = $-(CH_2)_3-CH(NH_4^+)-COO^-$. Structures reproduced from Cho et al. (2007), © 2007, used with permission from Bentham Science Publishers.

Many products of glycation have been proposed and identified and their structures and mechanisms of formation reviewed extensively by Cho et al. (2007). However, studies trying to elucidate the nature and role of AGEs in tissues have been hampered to some extent in the past, by their relative instability to procedures used to isolate them, such as acid and alkaline hydrolysis (Eble et al., 1983; Iijima et al., 2000). This led to a failure to identify some important cross-links, whilst exaggerating the quantitative importance of other more stable AGEs. For example, pentosidine, which is stable to acid hydrolysis, was thought to be a more important cross-link than it is (Sell and Monnier, 1989; Sell, Nemet and Monnier, 2010). Conversely glucosepane, DOGDIC, GODIC and MODIC – now known to be more quantitatively important cross-links *in vivo* - are known to be unstable during acid hydrolysis (Biemel et al., 2002). Therefore use of "gentler" procedures such as enzymic hydrolysis, have facilitated further study and identification of cross-links formed (Bensusan, Dixit and McKnight, 1971; Iijima et al., 2000; Biemel et al., 2002).

1.2.4 Effects of glycation on protein structure and function

Avery and Bailey (2005) considered that there are two main implications of glycation of a protein on its function:-

1. The modification of lysine and arginine side-chains by glycation is likely to affect the interaction of the protein molecule with other molecules, be they cell receptors, ligands or substrates.

2. During formation of AGEs and hence cross-links between amino acid residues within a protein the shape, freedom of movement and mechanical properties of the protein may all be affected.

These cover a wide range of potential effects on a large number of proteins *in vivo*. It is not surprising that many studies and much of the literature regarding glycation are concerned with the condition diabetes mellitus (DM) (Monnier et al., 1999; Sell et al., 2005; Goh and Cooper, 2008; Kumar et al., 2009). Patients with DM have chronic hyperglycaemia so as the rate of glycation is directly dependent on glucose concentration, the degree of glycation of their proteins will be significantly greater than in non-diabetic patients. Therefore protein glycation has an important involvement in the various multi-systemic complications associated with DM (Goh and Cooper, 2008) such as diabetic nephropathy (Vlassara et al., 1994), cataract formation (Huby and

Harding, 1988; Zilva, Pannall and Mayne, 1991) and cardiovascular disease (Chang et al., 2009).

In the case of longer-lived proteins such as collagen (Verzijl et al., 2000; Avery and Bailey, 2005; Heinemeier et al., 2013) and lens crystallins (Sell and Monnier, 2004), the age of an individual can be important. AGEs accumulate with ageing due to long-term exposure of these proteins, even to normal levels of glucose. In fact the mechanisms of glycation in diabetes and ageing are similar such that diabetes has been described as “accelerated ageing” (Bailey, 2001). Therefore in the literature, glycation-related changes associated with diabetes and ageing are often considered together (Schnider and Kohn, 1980; Brownlee, 1995; Paul and Bailey, 1996; Monnier, 2006).

As glycation is relatively non-specific in the scope of the damage it causes to proteins, its effects on an individual tissue may well as the result of a number of processes occurring. For example, diabetic nephropathy is associated with various apparently-AGE-mediated changes such as glomerular basement membrane thickening, glomeruloscerosis and up-regulation of collagen synthesis (Vlassara et al., 1994; 2000; Goh and Cooper, 2008).

Less complex examples occur where the changes are more directly attributable to glycation of a single protein. For example, Lapolla et al. (2000) found an increase in glycation of the immunoglobulin G (IgG) molecule in diabetic patients as opposed to non-diabetic controls and showed the preferential glycation of the antigen-binding (Fab) portion of the molecule as compared with the constant (Fc) domain. They suggested that this helped explain the immune deficiency reported in diabetic patients, as the binding between IgG molecule and its antigen might be affected (Lapolla et al., 2000).

Crystallins, the major proteins in the lens of the eye, are long-lived and remain *in situ* throughout life. They cannot be replaced and therefore must last a lifetime for normal functioning to be retained. Because of its optical function, the lens must be transparent and have a high refractive index. The proteins within the lens must therefore be soluble-yet-concentrated and provide short-range order (Bloemendal et al., 2004). The α-crystallins are chaperone or heat shock proteins (Hsps) which, as well as having a structural function, serve to bind other partially denatured proteins preventing their aggregation and thus damage to the lens tissue (MacRae, 2000; Bloemendal et al., 2004). Glycation can have a profound effect on this protective mechanism. Oimomi et al. (1988) demonstrated the association between both early- and late-stage glycation

products and cataracts in both elderly and age-matched diabetic patients whilst more recently, Kumar et al. (2009) showed an association between *in vivo* inhibition of glycation in diabetes helping to preserve α-crystallin chaperone activity, with this in turn apparently ameliorating both onset and maturation of cataract (Oimomi et al., 1988; Kumar et al., 2009). Another study found glycation of the other crystallins (β- and γ-crystallins) also to be significant in diabetes (Kawasaki et al., 1998). Therefore, potentially complex interrelationships appear to exist in simple non-compartmentalised soluble protein mixtures.

Effects of diabetes on the cardiovascular system are largely a result of stiffening of the large arteries such as the aorta. Accumulation of AGEs on long-lived proteins – most notably, collagen – within the arterial wall is thought to cause intermolecular cross-links, resulting in loss of elasticity of the tissue and altered vascular dynamics (Brownlee, 1995). The altered properties of the artery appear to increase the vascular load on the heart directly via changed pulsatile and wave reflection properties, resulting in cardiac hypertrophy (Chang et al., 2006; Chang et al., 2009).

In addition, it has been suggested that luminal narrowing in the vasculature as a result of diabetic complications and, indeed, with age is as a result of AGE-modified collagen reacting with and covalently trapping other plasma proteins, such as albumin and immunoglobulins (Brownlee, Pongor and Cerami, 1983; Brownlee, 1995). This could also occur via formation of immune complexes, as a result of attack by the immune system on glycated proteins that are antigenically “alien”. Such an process involving a lipoprotein, especially LDL, could be important in atherogenesis (Ulrich and Cerami, 2001).

1.3 Glycation of collagen

A small number of proteins are sufficiently long-lived that the accumulation of AGEs that can occur with ageing or in diabetes is significant. Alongside lens crystallins, collagen is often considered in the literature as being such a protein (Cerami, Vlassara and Brownlee, 1986; Sell and Monnier, 2004). The turnover of collagen is slow though this varies with the source tissue. Articular collagen has been found to have a half-life in humans at least comparable with the average human lifespan, whereas skin collagen has a half-life of around fifteen years (Verzijl et al., 2000). The accumulated levels of AGEs such as pentosidine and CML have been shown to be in direct relation to age and to be significantly greater in the longer lived articular collagen than that

taken from skin (Verzijl et al., 2000). In comparison, the turnover of human Achilles tendon collagen has been found to be relatively fast in the first seventeen years of life during the major growth phase, but negligible thereafter (Heinemeier et al., 2013).

Collagen's importance as the most widespread protein in the body and its presence in many different tissue types means that glycation of and AGE formation in collagen are likely to contribute to a variety of effects. More obviously, these include a loss of elasticity in, and wrinkling of skin and stiffening of tendons and joints, although other less immediately-visible effects on other systems, such as the kidneys and cardiovascular systems (considered in section 1.2.4) are also present and likely to result in morbidity and mortality in many subjects (Brownlee, 1995; Bailey, 2001).

Studies have taken place, demonstrating the association between glycation of collagen and age or diabetic complications. For example, studies of the glycation of skin tissue *in vitro* have demonstrated that collagen glycation therein can bring about some of the morphologic and phenotypic changes associated with ageing skin (Pageon and Asselineau, 2005; Pageon et al., 2007). They have also shown clinically, that the extent of glycation of skin collagen does correlate with the existence of other complications of diabetes in patients (Monnier et al., 1999).

Various effects of glycation on collagen have been studied in isolation, giving insight into the effects observed *in vivo*. These will be considered below.

1.3.1 Physical properties

Unsurprisingly the increased, non-specific cross-linking of collagen that occurs with increased glycation can have a marked effect on its physical behaviour. With increased AGE formation and cross-linking, collagen fibrils and fibres are effectively covalently stabilised beyond normal. Therefore changes in the mechanical and other physical properties of collagen are to be expected (Avery and Bailey, 2005).

The work of Chang et al. (2006, 2009) alluded to in section 1.2.4, showed evidence of a direct effect of the changed viscoelastic properties of collagen during diabetes on vascular dynamics *in vivo.* Other studies used isolated tendon *in vitro*, either from diabetic (induced) animals (Andreassen, Seyer-Hansen and Bailey, 1981) or used in the non-glycated state, to be glycated and tested after removal (Bai et al., 1992; Reddy, Stehno-Bittel and Enwemeka, 2002; Reddy, 2004). It is consistently reported from such studies that glycation causes increased stiffness and strength, and altered

stress-strain curves showing an increase in Young's modulus (Andreassen et al., 1981; Bai et al., 1992; Reddy et al., 2002; Reddy, 2004).

More recently Francis-Sedlak et al. (2009) measured the viscoelastic changes brought about by varying degrees of glycation of collagen gels. Using rheometry, they found significant increases in the storage and loss moduli of collagen which were related to the level of glycation.

1.3.2 Chemical properties

Due to the presence of various AGEs, often containing aromatic ring systems (Cho et al., 2007), glycated collagen shows an increased level of fluorescence which increases with the level of glycation (Tanaka et al., 1988b; Bai et al., 1992; Francis-Sedlak et al., 2009). Correlations have been shown between skin collagen fluorescence and age and duration of diabetes (Monnier et al., 1999; Sell et al., 2010).

In addition, in increasing the extent of covalent interlinking of collagen molecules AGE cross-links make collagen less susceptible to solubilisation in acid and impart increasing stability to enzymic (e.g. pepsin) or cyanogen bromide degradation. This has also been used in assessing the degree of glycation (Tanaka et al., 1988b; Bai et al., 1992; Berzhitskaya et al., 2002; Reddy, 2004), although it can clearly be problematic in reducing extraction of AGE cross-links for biochemical and chemical study.

1.3.3 Interaction with other matrix components

Much work has already been done to identify sites on the type I collagen molecule and along the collagen fibril concerned with interactions with other matrix components (Di Lullo et al., 2002; Sweeney et al., 2008). The primary sequences of the collagen peptides are clearly not just concerned with maintaining the triple helical structure. Within the collagenous Gly-*X*-*Y* repeat, other sequences are concerned with binding many different biological molecules as well as having importance in interactions with cells. Many binding sites and regions have been mapped, including those concerned with binding integrins and fibronectin (Kadler et al., 2008), other collagens such as collagen V, various enzymes such as the matrix metalloproteinases (MMPs) (Matsumoto et al., 1997) and PGs (Di Lullo et al., 2002; Sweeney et al., 2008). From this work, it is evident that the sites of glycation indicated (based on Reiser et al.'s

(1992b) work, which studied preferential sites of glycation along just 25% of the collagen *D*-period) might interfere with some of these binding functions.

Studies performed recently confirm this. For example, Reigle et al. (2008) demonstrated that glycation weakens binding of some PGs to type I collagen and also reduces *in vitro* endothelial cell migration over the surface of layers of glycated collagen as compared to non-glycated collagen. Using fibroblasts on collagen gels, Francis-Sedlak (2009) found the same effect of glycation of collagen on cell migration although interestingly, found that cell proliferation increased.

MMPs are enzymes found in the ECM, which are responsible for the homeostasis of collagen-containing tissues (Matsumoto et al., 1997; Perumal, Antipova and Orgel, 2008). As suggested in section 1.3.2, glycation and AGE formation in the collagen in these tissues might render it less susceptible to proeteolytic degradation by such enzymes. Bartling et al. (2009) found that AGE-modified collagen was significantly less susceptible to MMP digestion and that, alongside decreased ability of lung cancer cell lines to migrate through glycation-aged collagen preparations, this was a factor in the lower invasiveness of tumours in more elderly cancer patients.

Given the number of binding sites on collagen, there are multifactorial considerations in investigating AGE effects on the ECM. Elucidation of these will enable a greater understanding of the complex interrelationships between AGE formation and diabetes and age-related pathologies.

1.3.4 Changes in the structure and organisation of collagen due to glycation

Various techniques have been employed in the study of structural changes and effects brought about by glycation of collagen, at the molecular, fibrillar and super-fibrillar level.

At the molecular level, using circular dichroism Reigle et al. (2008) found that following *in vitro* glycation of RTT, glycated collagen molecules solubilised into acid showed no significant change in triple helical content as compared with non-glycated collagen. However binding experiments with heparin – a structural analogue to the heparin sulphate GAG chains found on some PGs, which have a specific binding site along the collagen molecule (Di Lullo et al., 2002; Sweeney et al., 2008) – resulted in a significantly lower triple helical content in the glycated collagen than the non-glycated

protein (Reigle et al., 2008). Therefore glycation appears to destabilise the collagenous molecular structure.

In the same study, using spectrophotometry to monitor progress, Reigle et al. (2008) also demonstrated that the kinetics of *in vitro* fibrillogenesis were altered by glycation, the process taking place more slowly for the glycated collagen than for the unglycated protein. However using negative staining and TEM, they found that the control and glycated collagens both formed fibrils that retained a 67 nm *D*-periodic length and showed no significant differences in the ranges of diameters of the fibrils formed (Reigle et al., 2008).

Also employing TEM work in this area Bai et al. (1992) found that on incubation of RTT with 0.2M ribose in phosphate buffered saline (PBS), the cross-sectional appearance of the fibrils changed markedly. Even after a few days, they became larger and deviated from their initial (and control) circular appearance, becoming much more irregularly-shaped. After two weeks, all the fibrils showed a distorted cross-section, and many of them were in contact or fused with others. Interestingly, despite these marked cross-sectional changes, no differences were discernable in the axial structure and the 67 nm *D*-periodicity remained the same as in the fibrillogenesis experiments carried out by Reigle et al. (2008). Using X-ray diffraction to study RTT glycated *in vitro* Tanaka et al. (1988a) produced data, which also showed an unchanged *D*-period length following the glycation treatments they studied and that the molecular packing expanded across the fibril and became less ordered. They found that the arrangement of the molecules along the fibril remained tightly ordered, but produced evidence that indicated a change in the conformation or electron densities of the molecular terminal telopeptide regions (Tanaka et al., 1988a). This latter finding was also reported for human tendon tissue in diabetic patients as compared with control (James et al., 1991). Work using neutron diffraction suggested that hydroxylysine residues associated with the N- and C-telopeptides for cross-linking were also particularly susceptible to glycation during *in vivo* studies of diabetes in rats (Wess et al., 1993).

Previous TEM work has concentrated on the positive staining banding pattern described in section 1.1.4.1 (see also, Figure 1.3). This staining method relies on binding of PTA stain to the basic amino acid residues lysine and arginine (Tzaphlidou et al., 1982b; Chapman et al., 1990; Hadley et al., 1998). As glycation and AGE formation involve these positively charged residues, it is reasonable to presume that post-glycation reductions in positive staining banding pattern intensities can yield information about glycation along fibrils. Hadley et al. (1998) used this to investigate

the glycation of human scleral collagen by fructose and found the glycation not to be a uniform event along the fibril, but rather to occur preferentially in four of the positive staining bands of the *D*-period (Hadley et al., 1998). Previously, using biochemical techniques, Reiser et al. (1992b) had also identified four specific lysine residues along type I collagen molecules, within fibrils, that are preferentially glycated *in vitro* and suggested that this was due in part to primary amino acid sequence. Reiser et al. (1992b) also suggested that the proximity of other groups, especially acidic amino acids might catalyse the Amadori reaction, and other work has provided evidence for this in other proteins (Shapiro et al., 1980; Venkatraman, Aggarwal and Balaram, 2001; Johansen, Kiemer and Brunak, 2006). However this did not explain all of the preferential sites of glycation found and they also suggested that higher order structure might be an important factor and that with increasing age and so more disordered tissue resultant from glycation-related damage, site-specificity of glycation might diminish (Reiser et al., 1992b). However, Reiser et al.'s (1992b) data did not support this as being a major determining factor of site specificity.

Other studies were in agreement that there are preferentially glycated sites. Wess et al. (1990; 1993) using neutron diffraction to study tendon AGE cross-linking sites *in vivo* and Mikulíková et al. (2007) using electrophoretic and HPLC separation techniques coupled to mass spectrometry to study *in vitro*-glycated tendon, both identified preferentially-glycated lysines. However, there is not complete consensus between these studies. There is consistency between some of the identified sites, i.e. Wess et al. (1990) and Mikulíková et al. (2007) agreed on one site of preferential glycation (lysine 855), and Reiser et al. (1992b) agreed with one of Wess et al.'s (1990) suggested sites (lysine 434) but not with the other (lysine 855). Reiser et al. (1992b) and Hadley et al. (1998) also agreed on another specific locus (lysine 479). However the other sites identified by Reiser et al. (1992b) and Mikulíková et al. (2007) appear to be uncorroborated by other studies.

Clearly different approaches can generate different specific data regarding preferred sites of glycation. Even within the same study, Le Pape et al.'s (1984) investigation of preferred glycation sites has yielded discrepant results between an *in vivo* diabetes model and an *in vitro* glycation approach. In addition, not all studies have concurred that there *are* preferred sites (see Brennan (1989b)). Therefore although there is general agreement inasmuch as the breadth of effects that glycation has on structure and other aspects of normal collagen functioning, study of more specific aspects of the origins and mechanisms of these effects might be difficult to accomplish until the

nature and locations of the glycation reactions can be determined with greater consensus.

1.4 Design and purpose of this study

Vincent Monnier's group at Case Western Reserve University in Cleveland, Ohio, USA, have already produced a transgenic mouse accelerated ageing model, to study *in vivo* ageing of lens crystallins brought about by glycation-like reactions as a result of high levels of ascorbate in lenticular tissue. The group found the increased formation of AGEs over 12 months in these mice, to be similar to AGE accumulation in human lenses over several decades (Fan and Monnier, 2008). This work continues, both with respect to study of the accelerated "glycation" process *in vivo* and potential rationales for its amelioration or prevention (Fan and Monnier, 2008; Fan et al., 2011).

These collaborators are currently working to produce an accelerated ageing collagen model in genetically modified mice, which it is anticipated will enable further study of glycation of this protein *in vivo* (Private communications, Vincent Monnier). Therefore with this future work in mind, type I collagen from mice has been selected for use in this project.

The overall objectives of this study are to further investigate the structural effects of *in vitro* glycation on type I collagen fibrils from mouse tail tendon (MTT). The methods used will be TEM to study changes in the relative band intensities along collagen positively stained with PTA, and X-ray diffraction with isomorphous replacement techniques to establish for the first time, a baseline one-dimensional MTT collagen electron density map and then study changes in the electron density along the *D*-period as a result of glycation.

1.4.1 Methods to be used in determinating structural changes

1.4.1.1 TEM study of positive staining band intensity changes along collagen fibrils resulting from glycation

Prior to this work, the studies by Hadley et al. (1998; 2001) have been the only uses of changes in PTA positive staining band intensity along collagen fibrils in the study of sites of glycation. Positive staining was described in section 1.1.4.1 (see also Figure 1.3). Removal of the positive charge on the basic amino acids lysine and arginine by

glycation and AGE formation processes can be visualised using TEM to capture electron micrographs of the positively stained fibrils. Changes in *relative* intensities of the bands are assumed to be indicative of sites of positive charge removal as a result of glycation and AGE formation. The practicability and success of this approach were demonstrated by Hadley et al. (1998). Whereas their initial study looked at glycation with one sugar (fructose, 0.5M) over 5 and 11 days (Hadley et al., 1998), it is intended to expand the scope of this study to look at the changes over a greater number of time periods and how they compare using different sugars.

1.4.1.2 X-ray diffraction with isomorphous replacement to establish a model one-dimensional electron density map along the MTT type I collagen fibril *D*-period and study the changes resulting from glycation

X-ray diffraction is a means of measuring the interrelationships between centres of electron density down to the molecular and sub-molecular level in well ordered structures and the reader is referred to other texts for information on its theory and use (Cantor and Schimmel, 1980; Glusker, Lewis and Rossi, 1994). Relevant theory is also described in the appropriate chapters in this thesis (Chapter 5 and Chapter 6).

As the wavelength of X-rays is <10 nm and in diffraction experiments, commonly around 0.1 nm (1 Å), they are a means of studying regular (crystalline) arrangements three orders of magnitude smaller than that which could be achieved with visible light (λ of visible light is between 400 and 700 nm). However as a suitable means of focussing X-rays has yet to be devised for such experiments, the data are recorded as diffraction patterns rather than images directly representing structure. A simplified X-ray diffraction experiment as applied to collagen fibril-containing tissue is represented in Figure 1.9.

One dimensional axial electron density maps have been constructed for type I collagen previously (Bradshaw, Miller and Wess, 1989; Orgel, Wess and Miller, 2000) using RTT as the source. These made use of the intensities (and therefore amplitudes) of the meridional reflections, which are easily measurable from the pattern. However, relative phase information is also required and this cannot be measured from a single pattern. A technique called isomorphous replacement can be used to determine the phases of each reflection, by using heavy atoms which bind to known, specific sites along the collagen molecule (see Bradshaw et al. (1989), Orgel et al. (2000), Antipova and Orgel

(2010)). The amplitude and phase information (structure factors) for all the meridional reflections can then be added to produce an axial electron density map.

X-ray diffraction has been employed previously in the study of aspects of collagen structural changes as a result of glycation (Tanaka et al., 1988a; James et al., 1991; Hadley et al., 2001) and has yielded valuable information. Wess et al. (1990; 1993) employed neutron diffraction techniques to study preferred sites of glycation as discussed above (section 1.3.4) and suggested the possibility that X-ray diffraction might be a way to detect more sites of glycation without need for reduction (or deuteration) or other modification. However to date nobody has used X-ray diffraction data to construct electron density maps of collagen in an attempt to locate sites of glycation along the collagen *D*-period. That is therefore a major objective of this work.

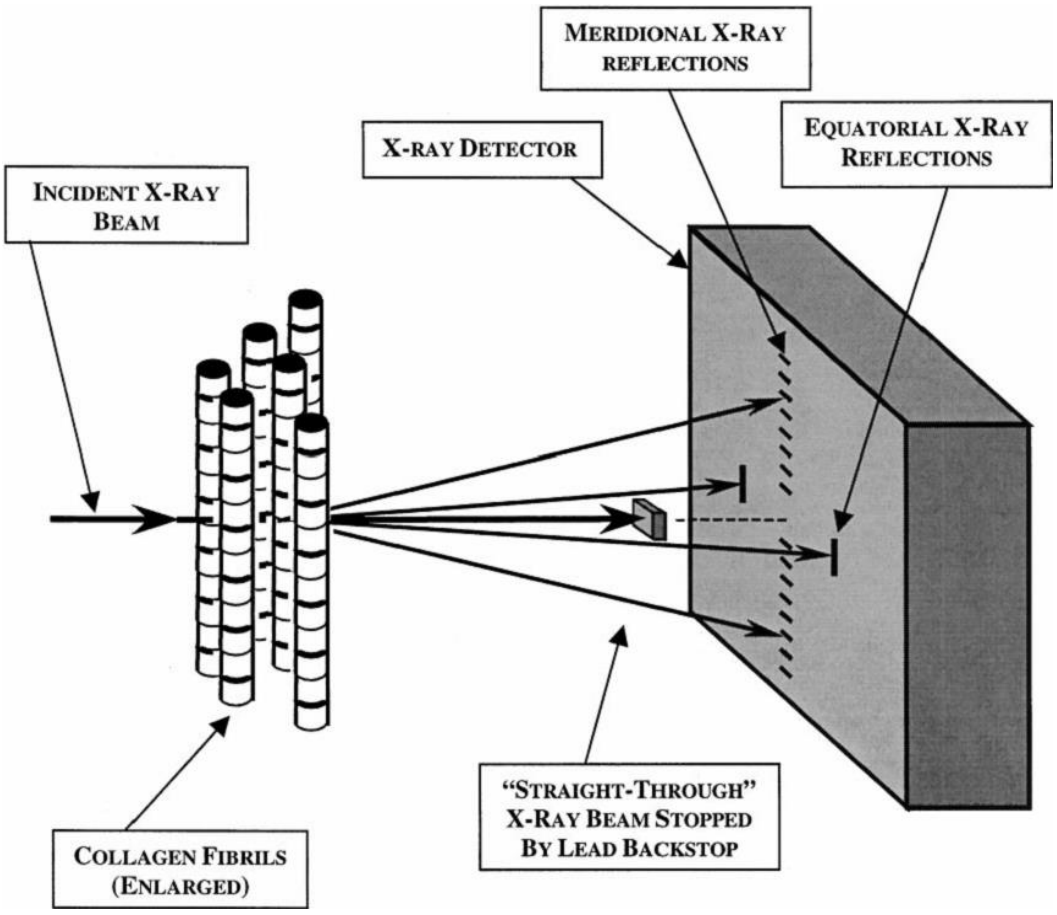


Figure 1.9 Simplified representation of an X-ray diffraction experiment studying collagen fibrils. The data recorded on the X-ray detector are reflections relating to the scattering of the incident X-ray beam by the regular arrangement of centres of electron density within the specimen. Therefore each reflection's intensity gives information about the centres of electron density it represents. The meridional reflections tend to be sharp as they represent the strong one-dimensional crystalline order along the fibril. The equatorial reflections represent the molecular packing across the fibril. These reflections, when seen, are more diffuse. Reprinted from Meek and Quantock (2001), © 2000, used with permission from Elsevier.

1.4.2 Study design and objectives

This project has studied glycation using four different monosaccharide sugars, i.e. glucose, galactose, fructose and ribose (Figure 1.10). Glucose is considered physiologically most relevant, as it is the major sugar present in the blood and extracellular fluid. However as a relatively low proportion of it is in the acyclic aldehyde form in solution (~0.002%) (Hayward and Angyal, 1977), it is a slow glycator (Bunn and Higgins, 1981) and therefore not always the most convenient and pragmatic choice for *in vitro* glycation experiments. Like glucose, galactose is an aldohexose and chemically and structurally similar but it has a greater proportion of the open acyclic form (~0.02%) (Hayward and Angyal, 1977) and would be expected to be a significantly faster glycator (Bunn and Higgins, 1981). Fructose is a ketohexose and ribose is an aldopentose, both being present in solution with a greater proportion in the acyclic form (~0.7% and 0.05%, respectively) (Hayward and Angyal, 1977) than either glucose or galactose. Fructose and ribose are therefore relatively fast glycators (Bunn and Higgins, 1981) and so are commonly used in *in vitro* glycation experiments (Tanaka et al., 1988a; Bai et al., 1992; Hadley et al., 1998; Mikulíková et al., 2007).

The effects of these four sugars on the positive charge on the surface of fibrils were studied using the PTA positive staining-TEM method (section 1.4.1.1) over 1, 2, 3 and 4 weeks. It was anticipated that the method will be sufficiently sensitive and specific to test the hypothesis that, even allowing for different rates of glycation, the preferred sites of glycation for each sugar along the *D*-period differ. It was also hypothesised that with sufficicient consensus within these data, it would be possible to suggest preferred amino acid glycation sites, as previously (Hadley et al., 1998).

By use of X-ray diffraction with isomorphous replacement techniques, it was hoped that a one-dimensional axial electron density map of MTT could be constructed and anticipated that this would be similar to RTT map constructed previously (Bradshaw et al., 1989; Orgel et al., 2000). Subsequently, it was hoped that by constructing similar maps of glycated MTT, it would be possible to demonstrate this technique to be sufficiently sensitive to detect any glycation-related changes in electron density in a location-specific manner along the MTT fibril collagen *D*-period. The hypothesis was that the electron density changes elucidated during that component of the study would match and bear relation to the site specific glycation loci identified by the PTA positive staining TEM glycation work and would show differences between sugars. Where there were discrepancies, it was anticipated that these might be capable of giving information

regarding the kinetics and timescale between initial sugar attachment and AGE formation.

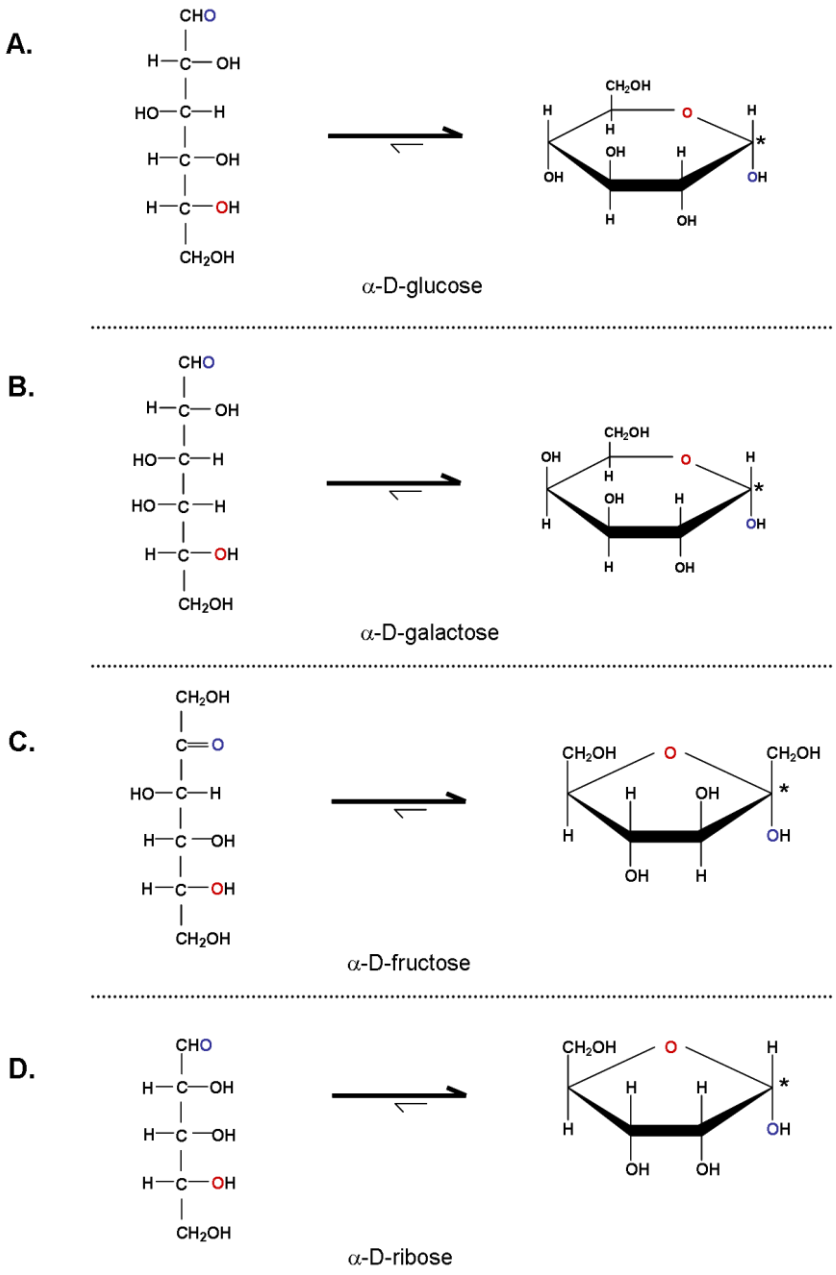


Figure 1.10 Structures of the four sugars used in this study: **A.** glucose, **B.** galactose, **C.** fructose, **D.** ribose. All of the sugars exist predominantly in the cyclic form (shown on the right) in solution, but it is the carbonyl group (-CHO in **A**, **B** and **D** and >C=O in **C**) in the acyclic forms (left) which initially react with amine groups during the first step of glycation. The carbon atoms within the ring structures have been omitted for clarity. An indication of the mechanism of cyclization is given by colouring of the two participant oxygen atoms. The carbonyl carbons in the cyclic forms are also indicated by asterisks. By switching the groups attached to this carbon above and below the plane of the ring for each sugar, the β-form of each sugar would be shown (also widely present in solution).

A further non-structural component of the study was included. This comprised a non-structural investigation of different profiles of products of glycation produced during all the treatments throughout the study, using HPLC separation and fluorescence detection techniques in collaboration with Nick C. Avery at the Matrix Biology Research Group in the School of Clinical Veterinary Medicine, University of Bristol (see Chapter 3). It was hypothesised that this work would establish the similarity of glucose and galactose biochemically, in terms of their profiles of glycation products, and therefore that galactose might be a relevant, faster-glycating, pragmatic experimental alternative to glucose in *in vitro* glycation investigation experiments. Conversely, it was hypothesised that fructose and ribose would show very different profiles of glycation products, indicative of their differing chemistries. It was anticipated that this component of the study might be useful in providing an additional dimension of information, which would be useful in the interpretation of the other, structural data.

Overall, it was hoped to establish that 1. TEM to measure changes in surface positive charge distribution along fibrils, and 2. X-ray diffraction with isomorphous replacement techniques to study changes in electron density within collagen fibrils, are two complementary methods, which are of value in investigating structural aspects of the glycation of fibrillar type I collagen.

Chapter 2: Materials and methods

2.1 Materials

Hydrochloric acid (sg 1.18), sodium chloride and D-(-)-fructose were obtained from Fisher Scientific UK Ltd (Loughborough, Leicestershire, UK). Sodium hydroxide pellets used to make NaOH solution, were from BDH Laboratory Supplies (Poole, Dorset). Diethylenetriaminepentaacetic acid (DTPA), 3-(N-morpholino)propanesulphonic acid (MOPS), D-(+)-galactose, D-(-)-ribose, D-(+)-glucose, D-sorbitol and protease inhibitor cocktail (PIC) in DMSO solution (P8340-1ML) were obtained from Sigma-Aldrich Company Ltd (Gillingham, Dorset, UK).

Software packages used throughout this work were: Microsoft Office 2003 (Excel, Powerpoint, Word) for data storage, manipulation and presentation; OriginPro v.8.5 graphing software (OriginLab Corportation, Northampton, MA, USA) for data and statistical analysis, plus production of graphs and plots.

2.2 Composition of glycation buffer and solutions

It was intended to use the same solutions for the glycation incubations throughout this project and careful consideration was given to the buffering system to be used, as described previously (Rodda, 2000). It was intended that the *in vitro* glycation experiments should be carried out at extracellular physiological pH, therefore ~pH 7.4. Although commonly used, it is known that phosphate buffer can catalyse and affect the mechanism of glycation reactions (Watkins et al., 1987; Gil, Salcedo and Romero, 2005). It has also been suggested that Good's buffers (Good et al., 1966) have some inhibitory effect on glycation (Watkins et al., 1987; Nemet, Strauch and Monnier, 2011), though the evidence for this is less compelling as it has been generated in comparison with systems using phosphate buffer. On the basis of advice (Private communication from Vincent Monnier, Case Western Reserve University, Cleveland, Ohio, USA), MOPS was selected as the buffering agent. It has a pKa of 7.2 at 25^{o}C (www.sigmaaldrich.com, 2013a) and is therefore suitable for buffering at physiological pH.

It is also known that the presence of transition metal ions can affect the progression and mechanisms of formation of AGEs via oxidative pathways (Culbertson et al.,

2003). The metal ions present would be present in small amounts as impurities from the chemicals, disposables, tissue, etc and difficult to standardise, especially with transfer between different experimental sites. Therefore for these experiments, it was also considered appropriate to use a chelating agent to remove these ions. Following advice (Private communication from Vincent Monnier, Case Western Reserve University, Cleveland, Ohio, USA), it was decided that DTPA was a good choice and would be included. It was anticipated that this metal ion removal would also help prevent activity of proteases such as MMPs (Matsumoto et al., 1997) during the incubations, therefore complementing the protease inhibitory activity of the PIC that was another constituent of the glycation mixtures.

For glycation incubations, a 0.2M concentration of all sugars was used in line with other *in vitro* glycation studies (Tanaka et al., 1988a; Bai et al., 1992; Meli et al., 2003).

2.2.1 Solution manufacture

All incubations took place in buffer of the composition 50 mM MOPS, 0.9% NaCl (physiological saline), 5 mM DTPA, with 0.04% sodium azide to retard bacterial growth. The solid components were all weighed out and dissolved in 800 ml dH_2O in a beaker with a magnetic stirrer, initially to make up 1 litre of a 2x stock solution with double the above concentrations. During this, it was necessary to add 1M NaOH solution drop-wise with a pastette to facilitate dissolution of the free acid DTPA. The stock solution was made up to volume in a volumetric flask and stored in a 1 litre screw top bottle at room temperature.

The buffer solution was made from the stock solution, 100 ml at a time. A 50 ml volume of the 2x stock solution and approximately 40 ml dH_2O were mixed in a conical flask and with continuous mixing using a magnetic stirrer at room temperature, adjusted to pH 7.40 as measured by an electronic pH meter, by drop-wise addition of 1M NaOH solution. The solution was then made up to 100 ml in a volumetric flask and stored in a 100 ml screw topped plastic bottle at room temperature.

Separate solutions of four glycating sugars – glucose, galactose, fructose and ribose – plus sorbitol, a non-glycating sugar alcohol control (See Tanaka et al. (1988a)), were made up individually in buffer to a concentration of 0.2 mol/l, as follows. The requisite amounts of each sugar/sorbitol to make up 100 ml of 0.2M solution were weighed accurately. In turn, each was dissolved in separate 50 ml aliquots of 2x stock buffer

solution and ~40ml dH_2O added. The solutions were adjusted to pH 7.40, made up to 100 ml and stored as described for the buffer solution.

2.3 Preparation of MTTs

For some of the experimental set-up stages of this work, tendon tissue dissected from mouse tails stored at -20^{o}C for an indeterminate period of time in the School of Optometry and Vision Sciences was used, as stated in the appropriate experimental sections. All of the data produced for the fulfilment of the aims of this study were obtained using MTT dissected from tails obtained from CD-1 mice (*Mus musculus*). The ideal age was considered to be 3 months, though age and gender were variable as the mice used were culled according to a schedule independent of this study. It was considered acceptable to store the tails and dissected MTT below 0^{o}C before use, as the evidence is that the collagen fibrillar and molecular structure post-thawing and pre-freezing are not significantly different (Chanzy, Franc and Herbage, 1976; Fulwood and Meek, 1994). However it was determined to avoid freezing MTT tissue after glycation incubations during the structural components of this study in light of the possibility that this might not apply after treatment.

2.3.1 Obtaining mouse tails

All mice were obtained from the Joint Facilities Unit in the School of Biosciences, Cardiff University. They were killed using CO_2 overdose and their tails removed immediately *post mortem*, as proximally to the body as possibly. The tails were sealed in plastic bags and transported on ice to the School of Optometry and Vision Sciences, Cardiff University. They were stored at -20^{o}C until required.

2.3.2 Dissection of tendons from mouse tails

The appropriate number of mouse tails from section 2.3.1 were removed from the freezer and allowed to thaw. Each tail contains four evenly-spaced tendons running the length of the tail, each affixed to the bones at each joint. These were dissected in a manner similar to that described previously (Goh et al., 2008). With a scalpel, the skin was cut along the length of the tail between tendons and being careful not to damage the tendons, such that four tail-length sections of skin remained affixed to the tail. Using fine-tipped tweezers, starting at the proximal end of the tail, the first skin section

was peeled back to the tip and discarded. It was found that the best way of sliding the tendon out was by use of an unstitching method using the fine-tipped tweezers, with a pair of conventional tweezers to hold the tail steady. Care was taken to avoid contamination of the tendon with muscle tissue. In this way, the required length of tendon was removed from the bone and could be cut free using the scalpel. It was immediately placed in buffer or physiological saline (0.9% NaCl) solution with or without 1% PIC, as deemed appropriate. Throughout the dissection procedure, physiological saline was sprinkled liberally over the tail using a pastette, to prevent the tendon tissue from drying out.

The same procedure was used to remove the other three tendons, peeling back the appropriate strip of protective skin and dissecting the underlying tendon as described. The tendons were then either used in the appropriate experiments or stored in sealed plastic bags at -20^{o}C until required.

2.4 Glycation incubations

A 0.2M concentration of all sugars was selected (see section 2.2). Previous *in vitro* glycation work has employed a range of incubation times. Using 0.2M of ribose to glycate RTT, Tanaka et al. (1988a) demonstrated significant changes in the packing arrangement of collagen across the fibril within 2 weeks of incubation and Bai et al. (1992) showed significant changes in collagen extractability after pepsin digestion, fibril packing density and collagen extractability after 1 week. Using 0.1M glucose and ribose to glycate bovine tendon collagen, Mikulíková et al. (2007) were able to detect preferred glycation sites for both sugars by employing mass spectrometry coupled to HPLC and electrophoretic separation techniques after 7 days. Hadley et al. (1998) used 0.5M fructose to glycate human scleral collagen and found, by TEM, consequent removal of positive charge along the fibril within 11 days. It was therefore decided that a range of glycation incubation times between 1 and 4 weeks would be used throughout this study.

2.4.1 General glycation procedure

All glycation (including buffer and sorbitol control) incubations were performed in the solutions prepared in section 2.2.1. In all cases, a volume of 1% of the glycation or control solution of PIC was added as per the manufacturers directions (www.sigmaaldrich.com, 2013b) to prevent protease activity during the incubations.

There was variability in the amounts of tendon and solution required for these incubations, depending on the experiment. In all cases tendon was added, ensuring that all of it was fully immersed in the solution. All incubation tubes were sealed as stated and maintained at 37°C using a heating block, incubator or water bath as appropriate. The tubes were regularly checked and gently agitated and the temperature checked. Incubation times used were 1, 2, 3 and 4 weeks.

Incubations were always terminated by removal of glycation solution and washing steps.

2.5 Simulated fibril construction

The simulated fibril was constructed on a Microsoft Excel spreadsheet and was based on, and almost identical to, one kindly donated by Dr Christian Pinali in the School of Optometry and Vision Sciences during the early stages of this project. It was constructed from the *Mus musculus* α1(I) and α2(I) chain amino acid sequences obtained from Uniprot (UniProt, 2012) with the $[\alpha 1(I)]_2\alpha 2(I)$ stoichiometry. Standard one letter notation was used to represent amino acids, with "O" representing hydroxyproline in place of coded prolines in the *Y*-position of Gly-*X*-*Y* within the triple helical regions of the α–chains (Ramshaw et al., 1998). The α–chains for the molecule were easily aligned using the ends of the repeating Gly-*X*-*Y* triple helical regions at the N- and C-terminal ends, and such that every amino acid position along the molecule within the triple helical region would contain a single glycine from one of the α–chains.

Molecules were placed over several columns and aligned across the simulated fibril, with a progressive 234 amino acid stagger (Meek et al., 1979; Hofmann et al., 1980). This could be accomplished by placing the entire molecule in a column with 1170 (i.e. 234 x 5) rows. The molecule-with-gap was offset in the next set of columns, using wrap-around, by 234 amino acid positions (rows). This process was repeated until five molecule-widths – the minimum necessary to represent all the molecular alignments across a microfibrillar crystallite - had been incorporated into the simulation. Using spare inserted columns alongside the main primary sequence, N- and C-telopeptides were folded to bring their lys/hyl residues into register with their cross-linking hyl/lys partners in the triple helix (George, Malone and Veis, 1999; Brady and Robins, 2001; Malone, George and Veis, 2004; Malone and Veis, 2004). It was recognised that this method of simulation is simplified inasmuch as assignment of triple helical and telopeptide amino acid positions to one spreadsheet cell each could not take into

account differences in the translation distances between the two, which are known to exist (Hulmes et al., 1977; Hulmes, Miller and White, 1980; Vitagliano et al., 1995; Orgel et al., 2000).

This simulated fibril was available for manipulation in Excel and OriginPro to predict various characteristics of the fibril for appropriate comparison to experimental data and is represented in its entirety in Appendix 1.

Chapter 3: Fluorescent advanced glycation end-product formation in mouse tail tendon following incubations in solutions of different sugars

3.1 Introduction

An introduction to the route of the formation of AGEs from the initial glycation or Maillard reaction, with appropriate references, was given in Chapter 1, section 1.2. With respect to quantitation of AGE formation, it has been shown that this is problematic as attempts to do so using a range of methodologies can yield varying and inconsistent results after glycation with different sugars (Syrový, 1994), which can belie their relative glycation reactivities (Bunn and Higgins, 1981). This can be explained in part by the differing routes and mechanisms of AGE formation. These can depend upon the various fragmentation pathways of the attached sugar to produce a number of reactive intermediates such as glyoxal, methylglyoxal and 3-deoxyglucosone, which can then go on to mediate formation of different AGEs (Thornalley et al., 1999). They are also directly dependent on the differing chemistries of the sugars concerned (Sell and Monnier, 1989; Suarez et al., 1989; Bailey et al., 1995; Culbertson et al., 2003; Jakas et al., 2008) and the local structural environment surrounding the glycation site within or on the protein (Venkatraman et al., 2001; Slatter, Avery and Bailey, 2008).

Much of the work carried out to characterise and quantitate glycation and AGEs formed in different biological contexts has used HPLC separation in concert with mass- and NMR spectrometry (Kawasaki et al., 1998; Biemel et al., 2002; Slatter et al., 2008; Nemet et al., 2011). These have yielded valuable information in protein glycation research including elucidation of the structure of the pentosidine AGE cross-link from human dura mater collagen (Sell and Monnier, 1989), characterisation and structure determination of α- and βNFC-1 AGEs after *in vitro* ribose glycation of RTT collagen (Paul et al., 1998) and identification of N^{ω}-carboxymethylarginine as an AGE product of *in vitro* glycation of bovine skin collagen with glucose (Iijima et al., 2000). More recently, such methodologies confirmed ornithine as an AGE-mediated breakdown product of arginine in human skin collagen and lens crystallins (Sell and Monnier,

2004), and enabled partial characterisation of the novel, complex AGE cross-link LW-1 in human skin collagen (Sell et al., 2010).

It has been established that chromatographic techniques are valuable research tools in their own right for the detection and purification of both novel and documented AGEs (Bailey et al., 1995; Iijima et al., 2000) and for their assay (Slatter et al., 2008; Avery et al., 2009; Willett et al., 2010). Although this entire study is predominantly structural, information giving a comparative biochemical perspective to these results is considered useful and potentially valuable to the analysis and interpretation of the data found in the following chapters. Therefore here, in collaboration with Nick C. Avery at the Matrix Biology Research Group in the School of Clinical Veterinary Medicine, University of Bristol, work was carried out to study profiles of products of MTT collagen glycation by the four different sugars stated in Chapter 1, section 1.4.2 using an established HPLC separation with fluorescence detection technique (Avery et al., 2009; Willett et al., 2010). It is intended to test the hypothesis that the different sugars used in this study will give rise to significantly different HPLC fluorescence profiles of products of glycation and that these will be traceable and progressive with each successive time period. It is also hypothesised that due to their chemical similarities, glucose and galactose will give rise to chromatograms which, allowing for differences in their rates of glycation of MTT collagen, will bear the closest resemblance to each other.

3.2 Materials and methods

3.2.1 Materials

For general laboratory chemicals and consumables used throughout the experiments in this entire work, refer to Chapter 2, section 2.1. The sample hydrolysis preparative step, plus the hydroxyproline and HPLC cross-links analysis portions of this work were carried out by Nick C. Avery at the Matrix Biology Research Group in the School of Clinical Veterinary Medicine, University of Bristol using established protocols within the group's laboratories. Data were acquired and analysed using Dionex software (Thermo Fisher, Hemel Hempsted, Hertfordshire, UK), prior to extraction onto OpenOffice v.3.3 spreadsheets and then further manipulation using Microsoft Excel 2003 and OriginPro v.8.5.

3.2.2 Preparation of MTT

The obtaining of mouse tails and MTT dissection process were described in Chapter 2, section 2.3. The MTT tissue used in these glycation experiments were from four recently killed individual female mice of indeterminate age, their tails having been stored in the freezer in the interim as previously described. Tendon tissue had been dissected into 0.9% NaCl solution containing 1% v/v PIC. Following complete tendon extraction from each tail, all tendon tissue had been transferred into small plastic bags which were sealed, labelled and stored at -20^{o}C until use in these experiments.

3.2.3 Glycation incubations

Incubations were carried out for each of the sugars glucose, galactose, fructose and ribose, each over 1, 2, 3 and 4 weeks. For each of these 16 groups, approximately 0.75 mg (dry weight) portions of MTT that had been stored frozen from the four individual mice described in section 3.2.2, were weighed out individually and placed into separate 10ml glass test tubes. Additionally, MTT from two of the mice were weighed out in the same way for 1 week buffer and sorbitol controls such that both of these incubations were run in duplicate. Similarly, MTT from the other two mice were weighed out for 4 week buffer and sorbitol controls and these too were set up in duplicate. To each tube was added 1.5 ml of the appropriate buffer, buffered sorbitol or buffered sugar solution described in Chapter 2, section 2.2 and in addition, 15 μl of PIC. After thorough vortex mixing, each tube was inspected to ensure all the collagen fragments in its tube were fully immersed in the incubation solution. All the tubes were then sealed and made airtight using Parafilm and placed in an incubator kept at 37^{o}C. Throughout the incubation time on a daily or bi-daily basis, all tubes were inspected and gently agitated by hand and the incubator temperature monitored.

At the end of the incubation time, the incubation solution was carefully removed from each tube to waste using a plastic pastette, taking care to ensure all of the MTT fragments remained in the tube. The MTT in each was then washed three times, with fresh 1.5 ml aliquots of 0.9% saline solution. Each wash was for several minutes, with thorough mixing. After the third wash, a further 1.0 ml of saline was added, vortex mixed briefly so that all the collagen fragments were in suspension and then quickly decanted into plastic 1.5 ml Eppendorf tubes. Any collagen fragments that remained stuck to the wall of the test tube were transferred to the receptacle Eppendorf tube with

fine tipped tweezers. The Eppendorfs were then capped and stored at -20°C in the freezer.

3.2.4 Sample analyses

The samples were transported from the School of Optometry and Vision Sciences in Cardiff to the Matrix Biology Research Group in the School of Clinical Veterinary Medicine, University of Bristol in a polystyrene box containing freezer packs which had been equilibrated to -20°C. They were stored there at -86°C until preparation and analysis.

3.2.4.1 Hydrolysis of samples

This was carried out as described by Avery et al. (2009), as follows. The samples were thawed and the collagen fragments were removed from the saline in each Eppendorf tube and added to ~1 ml of 6N hydrochloric acid in borosilicate tubes, one for each individual MTT sample. The tubes were sealed and heated to 115°C for 24 hours. Following the acid hydrolysis, the samples were allowed to cool, the tube seals broken and the tubes chilled to -80°C. They were then lyophilised to remove all residual hydrochloric acid. After drying, the samples were each rehydrated with approximately 250 μl water, such that the final concentration of collagen hydrolysis derivatives in the solution was ≥50 μg/25 μl.

3.2.4.2 Hydroxyproline assay

The hydroxyproline content of each hydrolysate was determined using a continuous-flow autoanalyser (Burkhard Scientific, Uxbridge, UK) utilising Bannister and Burns' (1970) adaptation of the Bergman and Loxley (1963) technique. Briefly, 100 μl of a 20x aqueous dilution of sample hydrolysate from Section 3.2.4.1 was mixed with 182 μl of diluent (1 part propan-2-ol to 2 parts dH_2O) and 139 μl of oxidant (a freshly prepared 7% w/v solution of chloramine-T in dH_2O mixed with 0.419M sodium acetate/0.157M sodium citrate buffer in 38.5% v/v propan-2-ol, pH 6.0, in the ratio 1:5). This was incubated at room temperature for 4 minutes, well mixed with 696 μl of Ehrlich's reagent solution (3 parts of freshly prepared *p*-dimethylamino-benzaldehyde solution in 60% perchloric acid to a concentration of 2 g in 3 ml acid, mixed with 13 parts of propan-2-ol) and the reaction mixture was incubated at 70°C for 25 minutes. The mixture was then cooled to room temperature by passage through tubing immersed in

running water, diluted by a factor of three in diluent and the absorbance at 550 nm was recorded using colorimetric detection. Data acquisition and analysis was performed using Dionex AI 450 software.

The assay was calibrated using external pure hydroxyproline standards. The hydroxyproline concentration was measured using these and so the collagen content could be calculated on the basis of collagen consisting of 14% hydroxyproline by weight (Reddy et al., 2002; Avery et al., 2009).

3.2.4.3 HPLC cross-links analysis

This was carried out in a manner described by Avery et al. (2009) and Willett et al. (2010). A portion of each rehydrated hydrolysis sample from section 3.2.4.1 was diluted in water and trifluoroacetic acid (TFA) solution in water to produce a final concentration of approximately 1 μg collagen equivalents per μl of 1% aqueous TFA solution. The samples were analysed using the Dionex Ultimate 3000 HPLC system using a Thermo Fisher (Hemel Hempsted, Hertfordshire, UK) Hypercarb S, 100 x 4.6 mm graphitic carbon column kept at 40^{o}C. A volume, normally 25 μl, of each diluted rehydrated hydrolysis sample was introduced onto the column by the Dionex Ultimate 3000 auto sampler, with a 0.5% TFA in H_2O mobile phase flowing through the column at a rate of 1 ml/min. This flow rate remained the same throughout the analytical cycle. The mobile phase composition was kept constant for 5 minutes and then a linear gradient of 0 – 15% tetrahydrofuran (THF) in 0.5% aqueous TFA was introduced to run over the next 30 minutes. The column eluate was continuously monitored using a Dionex RF2000 fluorescence detector and the data acquired and saved by the Dionex Chromeleon software. Up until 17 minutes after sample injection (12 minutes after THF mobile phase gradient commencement), excitation λ = 295 nm, emission λ = 405 nm. After that the excitation and emission wavelengths were switched to 335 nm and 385 nm, respectively. At the end of the mobile phase THF gradient, the THF concentration in the 0.5% aqueous TFA was linearly increased to 25% over 1 minute, held at that level for 1 minute and then linearly decreased to 0% THF over the remaining 3 minutes of the 40 minute cycle. All hydrolysed samples were analysed in this way. In addition, standards (prepared in-house and donated by Vincent Monnier, Dept. of Biochemistry, Case Western Reserve University, Cleveland, OH, USA) containing various enzymic and AGE cross-links were run periodically in between samples. These contained pentosidine of known concentration and so enabled the pentosidine within each sample to be quantified. The pentosidine was expressed per nmol of collagen, based

on the hydroxyproline assay results from section 3.2.4.2. The standards also contained other enzymic and non-enzymic (AGE) cross-links, on the basis of which retention times could be established to enable identification of other peaks.

3.3 Results

3.3.1 AGE cross-links

Individual example HPLC chromatograms from the hydrolysed MTT collagen samples from these experiments are presented, for each control and sugar treatment group, in Figure 3.1, Figure 3.2, Figure 3.3, Figure 3.4 and Figure 3.5. On all chromatograms, peaks identified by the Dionex Chromeleon software on the basis of standard peak positions are indicated. The constraints were set within the software such that a peak was identified as being a given cross-link/AGE by least squares analysis if it fell within a range of the assigned retention time ±0.1 minute. Some HPLC peaks were not identified due to falling slightly outside their set positions and ranges, despite an apparent identity being obvious from the positions of peaks identified on other chromatograms. Enzymic pyridinoline cross-links are resistant to acid hydrolysis (Fujimoto and Moriguchi, 1978) and capable of being detected and measured using HPLC with fluorescence detection (Eyre et al., 1987; Hanson and Eyre, 1996; Avery et al., 2009), and so their peaks are identified in these data as well. In the cases of the buffer and sorbitol controls (Figure 3.1), glucose (Figure 3.2), galactose (Figure 3.3) and fructose-treated (Figure 3.4) MTT collagen, 25 μl of the appropriately-reconstituted MTT hydrolysate (see sections 3.2.4.1 and 3.2.4.3) were injected onto the HPLC column. All of these are therefore represented on the same *y*-axis detector signal scale (in mV), so the peak heights on these chromatograms are directly comparable. It was found that a smaller (5 μl) volume of the ribose-treated MTT hydrolysate had to be used due to the strength of the signal. Therefore during a peak height comparison with the other sugar incubation chromatograms, the ribose trace peak heights should be multiplied by five.

Figure 3.1 shows buffer and sorbitol control MTT traces plotted on the same axes for their 1 week (Figure 3.1A) and 4 week (Figure 3.1B) incubations. The changes observed are small, as expected, and any changes seen between the two time periods could be attributable to the use of MTT samples drawn from tissue from two of the four mice used in these experiments for the 1 week incubations and the other two mice for

the 4 week incubations. Therefore inter-individual variation is likely to be a factor. This could explain the drop in the hydroxylysyl-pyridinoline (OHPyr), which is the apparent identity of the peak(s) at 13.4 minutes on the 4 week chromatograms, and conversely, a possible modest increase in the unassigned peak with an average retention time of 18.5 minutes. Although not assigned, the buffer and sorbitol control peaks running with an average retention time of 21.9 minutes after 1 week of incubation is likely to represent pre-existing quantities of the K1 AGE (Bailey et al., 1995) in the MTT tissue. Unsurprisingly, neither K1 nor pentosidine appear to increase appreciably in either control. The latter is borne out by the specific pentosidine results presented in Table 3.1 (Page 55). Although these increased slightly between 1 and 4 weeks incubation for both controls and all the individual specific pentosidine levels after 4 weeks in both controls were higher than all of those after 1 week, these could not be stated to be statistically significant (i.e. $p<0.05$) on the basis of the small sample sizes (duplicates) used. Again, inter-individual variation could also be a factor. Comparison of both 1 and 4 week controls (buffer and sorbitol) shows the traces to be similar for both after the two incubation times, and no novel peaks have appeared after 4 weeks as compared with 1 week.

Study of Figure 3.2 shows the progression of fluorescent peaks during incubation with glucose. The reduction in the OHPyr peak (identified on the 1 week chromatogram) between 1 and 2 weeks incubation and then its resurgence between 3 and 4 weeks is interesting. These are traces from MTT samples obtained from the same individual mouse and a mechanism for this variation in the level of a stable enzymic cross-link is not known, though it could be a mis-assignment of another more labile hydrolysis product. Similarly, the unassigned peak at 18.4 minutes (average) appears to lose height between 1 and 2 weeks and then steadily gain in height over the next two time periods (3 and 4 weeks). Pentosidine shows a small-but-steady increase over the four time periods, but with only the changes in specific level between 2 and 3 weeks, and overall between 1 and 4 weeks being shown to be statistically significant (See Table 3.1, Page 55). The K1 AGE peak – assigned on the 3 and 4 week chromatograms but clearly present on the 1 and 2 week traces – appears relatively small and static for the first two time periods, increasing slightly between 2 and 3 weeks and showing a more notable increase between 3 and 4 weeks, overtaking pentosidine in height.

Apart from the changes in relative peak heights of pre-existing peaks, the 1 and 2 week glucose HPLC fluorescence traces (Figure 3.2A & B) are difficult to distinguish from the controls (Figure 3.1). After 3 weeks (Figure 3.2C), a number of small novel glycation-

related peaks become evident, most notably a series of three small peaks evenly spaced eluting after the pentosidine peak and a further peak forming a shoulder on the broad ~39 minute peak associated with the accelerated THF ramp at the end of the mobile phase cycle. After 4 weeks, these all become much better defined, as do the small shoulder eluting immediately before the 18.4 minute peak, a small peak midway between K1 and pentosidine and a cluster of small unidentified peaks immediately preceding OHPyr. Though the changes are quite obvious, glycation of MTT with glucose produced modest changes as determined by this analysis.

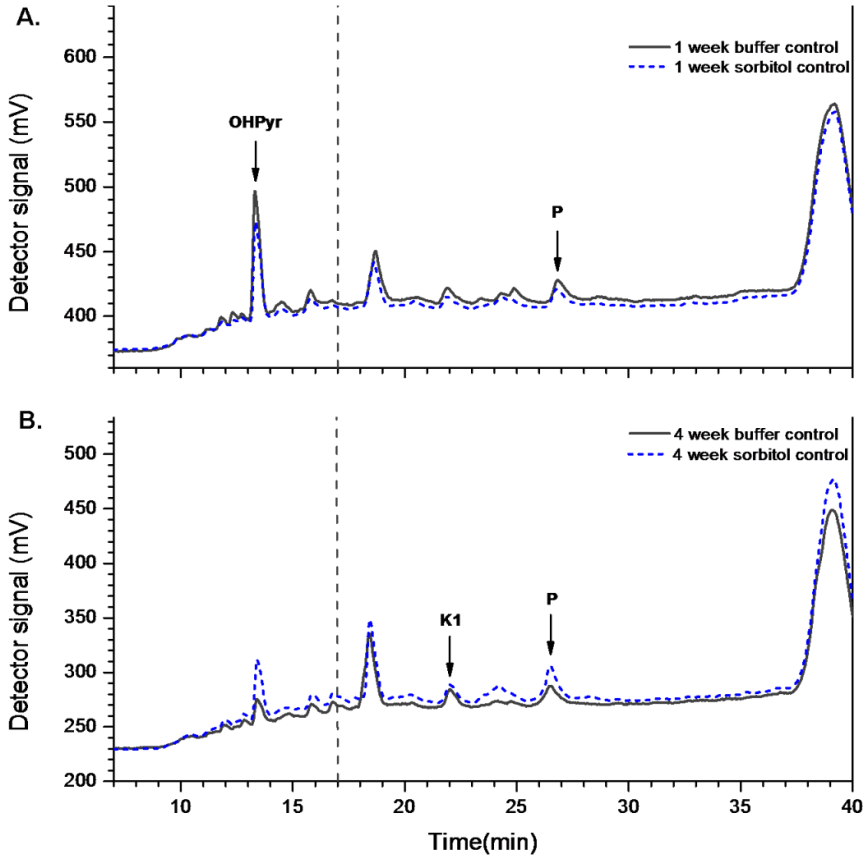


Figure 3.1 Post-hydrolysis (see section 3.2.4.1) HPLC chromatograms of individual **A.** 1 week buffer and sorbitol control and **B.** 4 week buffer and sorbitol control MTT samples. The detector signal is from a fluorescence detector (see section 3.2.4.3): from sample injection (0 minutes) until 17 minutes, excitation λ = 295 nm, emission λ = 405 nm; between 17 and 40 minutes, excitation λ = 335 nm, emission λ = 385 nm. The wavelength switchover point is indicated by the vertical dashed line. As the early chromatogram is featureless, only the traces beyond 7 minutes into the cycle are shown. Both chromatograms are represented on a full detector signal scale of 400 mV for ease of comparison and 25 μl of diluted hydrolysis sample prepared as described in section 3.2.4.3 was introduced onto the column. The peaks indicated by labels and arrows are those assigned by the Dionex Chromeleon software on the basis of least squares analysis of peaks against a window of established retention times of enzymic and non-enzymic cross-links and AGEs ±0.1 minute. Key:- *AGEs*: **K1** = Cmpd K component (Bailey et al., 1995), **P** = pentosidine; *Enzymic cross-link*: **OHPyr** = hydroxylysyl-pyridinoline.

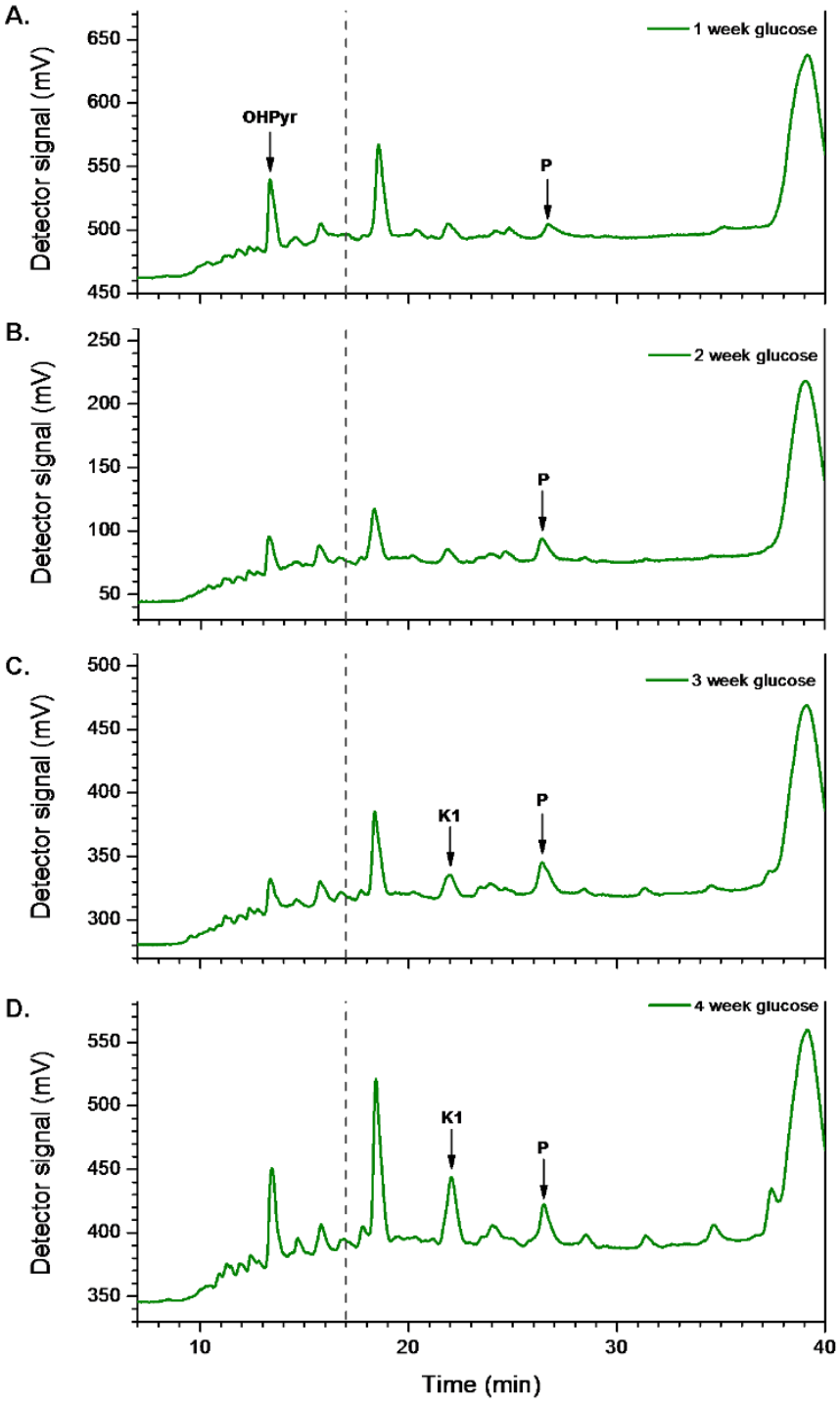


Figure 3.2 Post-hydrolysis HPLC chromatograms of individual **A.** 1 week, **B.** 2 week, **C.** 3 week and **D.** 4 week glucose-incubated MTT samples. Parameters and representations used are exactly as described in the legend for Figure **3.1**. The peaks indicated by labels and arrows are those assigned by the Dionex Chromeleon software on the basis of least squares analysis of peaks against a window of established retention times of enzymic and non-enzymic cross-links and AGEs ±0.1 minute.Key:- *AGEs*: **K1** = Cmpd K component, **P** = pentosidine; *Enzymic cross-link*: **OHPyr** = hydroxylysyl-pyridinoline.

The galactose incubations produced more noticeable alterations in the HPLC fluorescence profile of hydrolysed MTT collagen (Figure 3.3). The putative OHPyr peak was not identified by the Chromeleon software but is prominent at 13.3 minutes (average) and does not appear to have changed appreciably during the incubations, except for a reduction between 3 and 4 weeks. With regard to glycation effects, after 1 week there are already small changes compared with the controls (Figure 3.1), with three clearly defined, small novel peaks eluting after the pentosidine and a small shoulder immediately before the 18.4 minute unassigned peak (comparable with that seen for glucose after 4 weeks, seen in Figure 3.2D). After 2 weeks, these changes have become more pronounced, with an apparent small new doublet peak running at approximately 23.0 minutes and a new, unidentified fast-eluting peak evident at 11.2 minutes. Between 2 and 3 weeks some profound changes have occurred, with the unassigned peak at 18.4 minutes becoming significantly higher and the doublet, the peaks now separately resolved at 22.6 and 23.0 minutes, having also increased significantly. The fast 11.2 minute peak, the 25.0 minute and the 5 peaks between pentosidine and the broad THF gradient-associated peak at ~39 minutes have increased in height.

The chromatogram after the 4 week MTT incubation in galactose appears similar to the 3 week trace except for a sharp decrease in the unassigned peak at 18.4 minutes, despite its previous large increase between 2 and 3 weeks, where it now forms a doublet with its former shoulder.

It is noteworthy that unlike with glucose, the K1 AGE did not increase and overtake pentosidine in terms of peak height. The specific pentosidine levels for MTT collagen incubated in galactose after each time period are shown in Table 3.1 (Page 55). It can be seen from this that the change in pentosidine measured for this monosaccharide after 1 and 2 weeks showed the only significant difference between two levels measured for adjacent time periods, with the overall change between 1 and 4 weeks also being significant. For every time period, the galactose specific pentosidine level was significantly greater than that for glucose (one-tailed paired sample Student's t-test: 1 and 2 weeks, $p < 0.05$; 3 and 4 weeks, $p \leq 0.005$). It would therefore be reasonable to presume that the pentosidine increase over 1 week as compared with a notional "0 week incubation" or untreated sample would be significant.

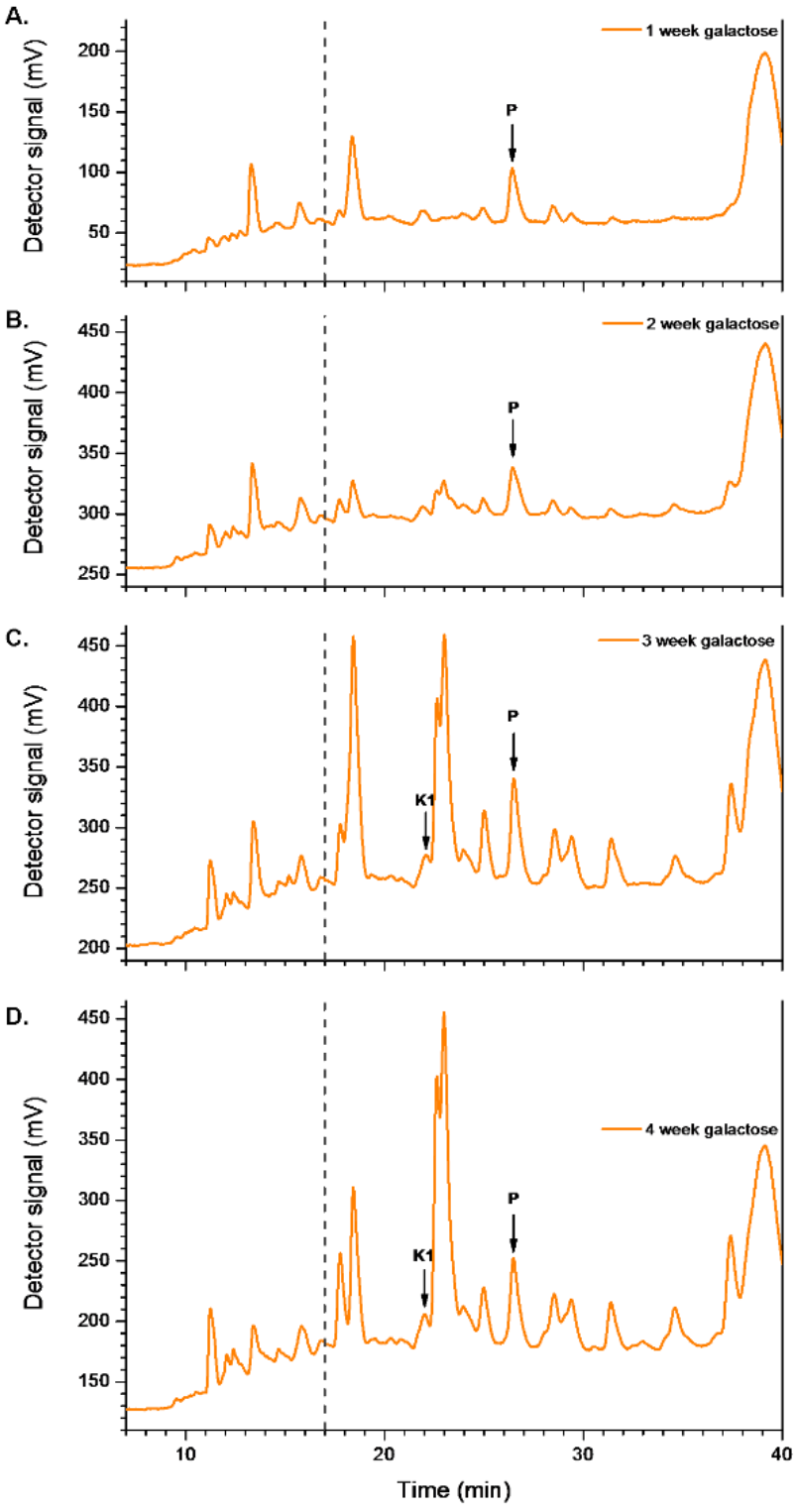


Figure 3.3 Post-hydrolysis HPLC chromatograms of individual **A.** 1 week, **B.** 2 week, **C.** 3 week and **D.** 4 week galactose-incubated MTT samples. Parameters and representations used are exactly as described in the legend for Figure **3.1**. The peaks indicated by labels and arrows are those assigned by the Dionex Chromeleon software on the basis of least squares analysis of peaks against a window of established retention times of enzymic and non-enzymic cross-links and AGEs ±0.1 minute. Key:- *AGEs*: **K1** = Cmpd K component, **P** = pentosidine.

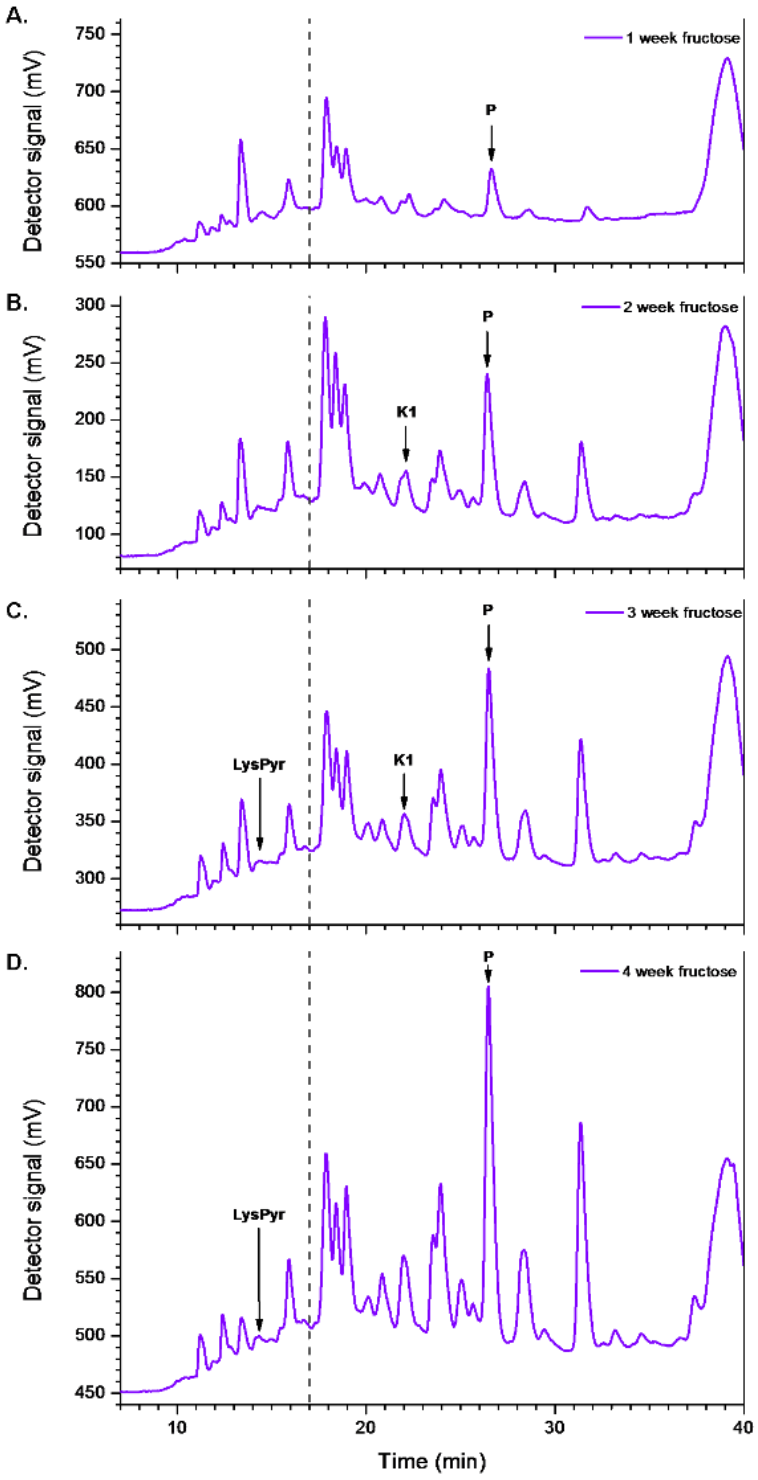


Figure 3.4 Post-hydrolysis HPLC chromatograms of individual **A.** 1 week, **B.** 2 week, **C.** 3 week and **D.** 4 week fructose-incubated MTT samples. Parameters and representations used are exactly as described in the legend for Figure **3.1**. The peaks indicated by labels and arrows are those assigned by the Dionex Chromeleon software on the basis of least squares analysis of peaks against a window of established retention times of enzymic and non-enzymic cross-links and AGEs ±0.1 minute. Key:- *AGEs*: **K1** = Cmpd K component, **P** = pentosidine; *Enzymic cross-link*: **LysPyr** = lysyl-pyridinoline.

The changes seen over all of the MTT incubation time periods in the case of fructose are even more pronounced (Figure 3.4). After 1 week, a triplet of peaks has appeared at 17.9, 18.4 and 18.9 minutes respectively. The middle one corresponds to the unassigned peak noted in the controls, glucose and galactose incubations with the 17.9 minute peak being close to the 17.8 minute shoulder seen with glucose (Figure 3.2C and D) and galactose (Figure 3.3). This remains prominent throughout the incubation time periods used, with the third peak increasing as a proportion of the other two after 3 and 4 weeks. In addition, a doublet of peaks at 23.5 and 23.9 minutes becomes obvious after 2 weeks and increases in height after the 3 week and especially the 4 week incubations. Two small peaks eluting after pentosidine, at approximately 28.4 and 31.4 minutes respectively, steadily increase in height – the later one more so – over the incubation periods studied, with other smaller peaks appearing in this region of the chromatogram especially after the 3 week incubation. A couple of smaller peaks elute before the (unassigned) putative, prominent OHPyr peak. The first of these elutes at approximately 11.2 minutes and the second is at 12.4 minutes. Both are unidentified, becoming more visible over the fructose incubations (Figure 3.3). Unlike with the galactose, where the first of these peaks increased more rapidly in height than the 12.4 minute peak, the latter kept pace with the earlier peak in the case of fructose, exceeding it after 4 weeks of incubation. The progressive increase in the pentosidine peak height over time is striking, with a definite but more modest increase in K1 also seen.

The increase in specific measured pentosidine level occurring over the incubation time periods for fructose, are shown in Table 3.1. Each increase in time of incubation showed a progressive and statistically significant increase in measured pentosidine. It is clear that the level of pentosidine after 2, 3 and 4 weeks of MTT collagen incubation in the fructose solution is markedly higher than for glucose and galactose. For all time periods, the fructose produced higher specific pentosidine levels than glucose (one-tailed paired Student's t-test: $p < 0.01$ for 1, 2 and 3 weeks; $p < 0.001$ for 4 weeks) and it is therefore a reasonable assumption that after 1 week's incubation in fructose, the pentosidine is significantly higher than the baseline "0 week" level. After the 1 week incubations, there was no significant difference between the fructose and galactose pentosidine level. However, after 2, 3 and 4 weeks, the fructose-derived pentosidine was significantly higher than that measured after the corresponding galactose incubations (one-tailed paired Student's t-test: $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively).

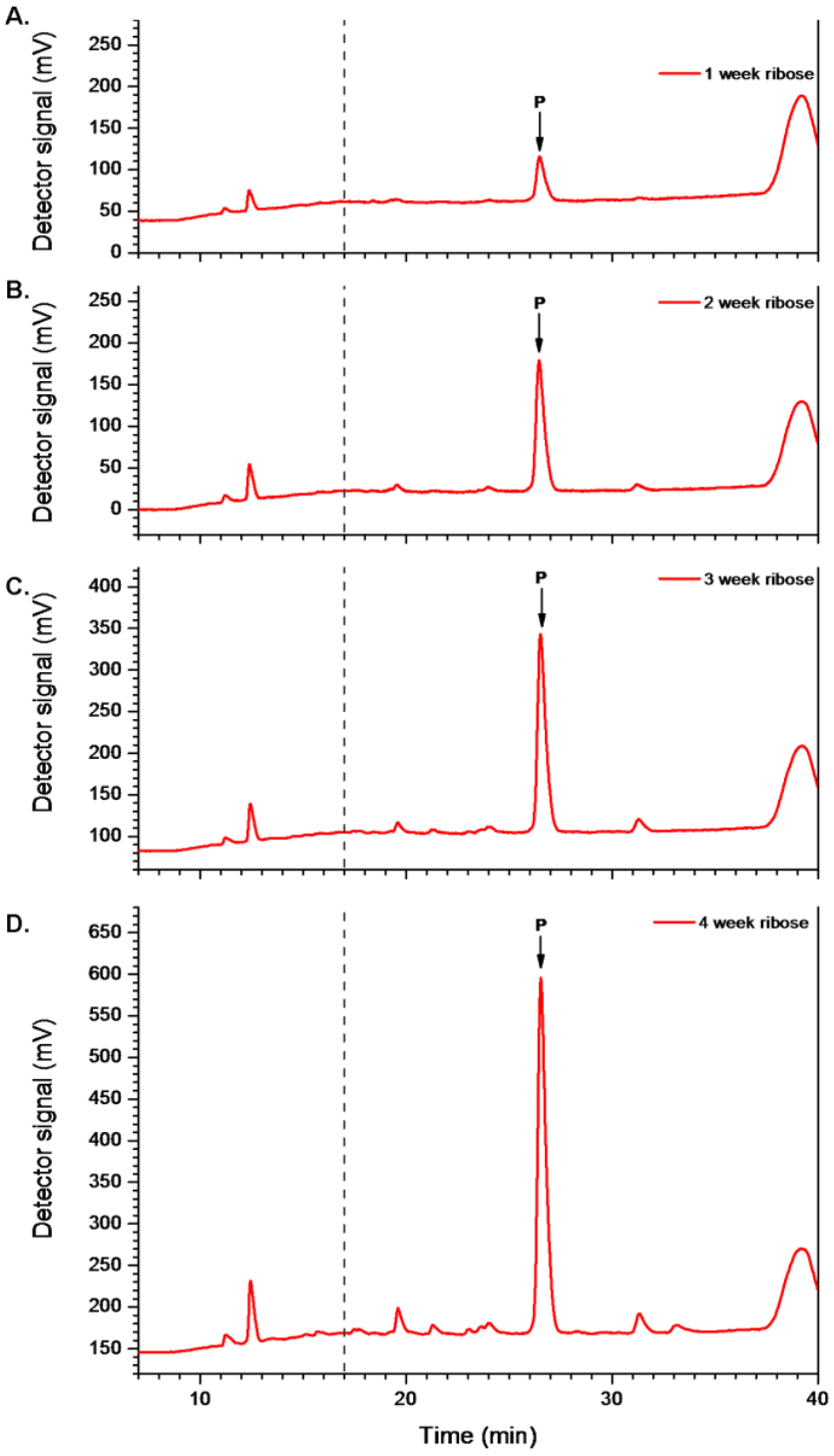


Figure 3.5 Post-hydrolysis HPLC chromatograms of individual **A.** 1 week, **B.** 2 week, **C.** 3 week and **D.** 4 week ribose-incubated MTT samples. Parameters and representations used are as described in the legend for Figure **3.1** except that here, 5 μl of the diluted hydrolysed samples were injected due to the height of the pentosidine peak and the full detector signal scale used was 600 mV. The peaks indicated by labels and arrows are those assigned by the Dionex Chromeleon software on the basis of least squares analysis of peaks against a window of established retention times of enzymic and non-enzymic cross-links and AGEs ±0.1 minute. Key:- *AGE*: **P** = pentosidine.

In the case of ribose (Figure 3.5), the HPLC chromatograms for each successive time period show a dominance of pentosidine formation. The only other peak to be comparable on all chromatograms is the unidentified 12.4 minute peak previously seen with fructose. Both of these AGE peaks show progressive increases over the successive individual time periods, though this is significantly more pronounced for the pentosidine cross-link. Other smaller peaks become more evident such that after 4 weeks, six of these were resolved at 11.2, 19.6, 21.2, 24.0, 31.4, 33.1 minutes. However, none of these retention times corresponded with any known species in the standards so could not be identified.

	Pentosidine content (pmol/nmol collagen)			
	After 1 week incubation	After 2 week incubation	After 3 week incubation	After 4 week incubation
Incubation group	Mean ± 1sd	Mean ± 1sd	Mean ± 1sd	Mean ± 1sd
Buffer control	**0.657** ± 0.037			**1.072** ± 0.155
Sorbitol control	**0.521** ± 0.150			**0.965** ± 0.153
Glucose	**0.523** ± 0.209	**0.762** ± 0.029	* **1.136** ± 0.231	† **1.191** ± 0.274
Galactose	**1.536** ± 0.696	* **2.282** ± 0.776	**2.379** ± 0.302	† **2.703** ± 0.329
Fructose	**1.259** ± 0.215	* **4.00** ± 1.155	* **6.99** ± 1.592	‡ * **8.89** ± 0.804
Ribose	**153.7** ± 51.1	** **442.8** ± 96.4	* **746.2** ± 155.8	‡ **815.2** ± 194.4

Table 3.1 Pentosidine levels for each incubation group expressed specifically with respect to collagen content. For all the glucose, fructose, ribose and galactose incubations, means and standard deviations (sds) are based on four individual samples. The buffer and sorbitol control figures are from duplicate samples. For each of the sugar treaments, the pentosidine results were compared with the previous time period using a one-tailed paired Student's t-test. * indicates $p < 0.05$; ** indicates $p < 0.01$; † and ‡ denote significant week 4 differences compared with week 1, indicating p values of <0.05 and <0.01, respectively.

The specific pentosidine levels in the ribose-treated MTT samples for all four time periods are shown in Table 3.1. In all cases, they are two orders of magnitude higher than for any of the other three glycating sugars. Unsurprisingly, for all time periods, comparison of the ribose-treated MTT collagen pentosidine content with the tissue treated with the other three glycating sugars gives a statistically significant difference in all cases (one-tailed paired Student's t-test: $p \leq 0.005$). Changes in measured specific pentosidine levels increased significantly between 1 and 2 weeks, and 2 and 3 weeks. Between 3 and 4 weeks, the change was not statistically significant despite an increase in the reported mean. It can be seen that the difference between the means of

the 3 and 4 week incubations is 69.0 pmol/nmol collagen, which is less than either of the relevant standard deviation (SD) values documented. In light of this, comparison of this with the changes between 1 and 2 weeks, and 2 and 3 weeks (289.1 and 303.4 pmol/nmol collagen, respectively) suggests that the rate of pentosidine formation is diminishing and possibly close to reaching a plateau.

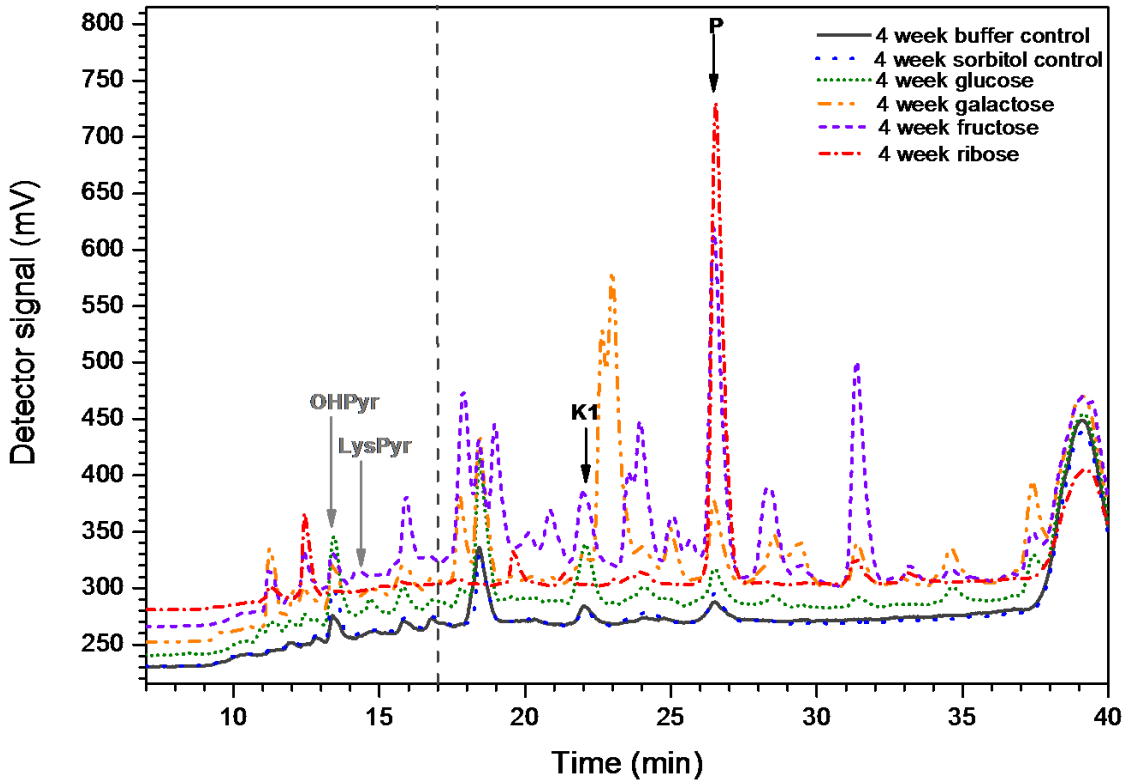


Figure 3.6 Four week HPLC chromatograms from Figure **3.1**, Figure **3.2**, Figure **3.3**, Figure **3.4** and Figure **3.5** for the control groups and the four glycating sugar incubation groups. For ease of comparison, all chromatograms are represented on the same detector signal (mV) scale. As a fifth of the volume (5 μl) of diluted hydrolysate was used for the ribose as compared with all the other groups (25 μl), the comparative ribose peaks should be envisaged as being five times higher than seen here. The baselines at the beginning (7 minutes) of each trace shown are staggered for clarity, particularly so that the smaller peaks can be seen more easily. AGEs are labelled in black, enzymic cross-links in gray.

In Figure 3.6 are presented the HPLC chromatograms for all treatment groups, including controls, after 4 weeks incubation. Comparison of the glucose and control traces confirms the presence of small novel AGE peaks after glucose treatment. All of these novel glucose peaks appear to be matched by peaks in the 4 week galactose MTT treatment group, though in the galactose chromatogram they are comparable to or of greater peak height than in the glucose trace. This is with the exception of the K1 AGE peak which, contrary to the trend for pentosidine, seems to show higher peaks for

the glucose treatment than galactose. For all four individual mice, the K1 peak was taller after the 4 week MTT incubation in glucose as compared with the 4 week galactose treatment.

The 4 week galactose treatment has predominantly produced a set of peaks with elution times the same as peaks seen after 4 weeks of treatment with glucose and/or fructose, although the relative peak heights differ (Figure 3.6). The exception is the doublet at 22.6 and 23.0 minutes, which despite not being identified here, is unique to the galactose treatment and the most significant detector response measured following treatment with this sugar.

The 4 week fructose-treated MTT has produced an HPLC trace that is more complex still, than the chromatogram from the galactose treatment. With the exception of the significant doublet on the galactose trace, all of the peaks seen for glucose and galactose show the same retention times as peaks seen after the fructose treatment. However, some of these peaks are relatively-bigger than their glucose and galactose counterparts, most notably those at 17.9, 23.9 and 31.4 minutes plus the pentosidine peak. Unique peaks for fructose are seen at 19.0, 20.8 and 23.5 minutes, the latter forming a shoulder on the 23.9 minute peak. Of the identified AGE peaks, the pentosidine has shown the biggest increase (see Table 3.1). However, the K1 peak increase appears broadly in line with that seen with glucose.

The 4 week ribose treatment chromatogram is dominated by pentosidine, with the unidentified 12.4 minute peak being small in comparison. Even so, as the peak heights on the ribose HPLC chromatogram trace shown in Figure 3.6 are effectively a fifth their actual heights as compared with the other traces, the production of these two AGEs with ribose treatment is enormous by comparison. The pentosidine results in Table 3.1 show the magnitude of this. However ribose treatment of collagen appears to produce few novel peaks detectable using this method. The only one of these which has no equivalent in any of the other treatment groups has an elution time of 19.6 minutes.

3.4 Discussion and conclusions

The data presented in this chapter have shown interesting comparisons between products of glycation in MTT collagen for the monosaccharides studied. However, they are not an entire representation of the products of glycation formed during these incubations. It is known that some AGEs are labile to the acid hydrolysis procedure employed (Eble et al., 1983; Iijima et al., 2000; Biemel et al., 2002), including the most common ones occurring *in vivo* such as glucosepane, GODIC, MODIC and DOGDIC (Biemel et al., 2002; Sell et al., 2005). Notwithstanding that, although it has been stated previously that most [then known] AGEs exhibit UV activity (Monnier, 1990), the UV fluorescence detection system is by no means conducive to direct detection of all AGEs either, including glucosepane (Sell et al., 2005) and CML (Ahmed et al., 1986; Monnier et al., 1999). However, fluorescence at various emission/excitation wavelengths in the UV and short-visible region have been shown to give comparable estimates of levels of glycation in different treatment groups (Berzhitskaya et al., 2002) and therefore selection of one set excitation/emission wavelength combination or programme seems reasonable, for the evaluation of changes in *relative* amounts of glycation products for all groups considered. Different peaks could not be assumed to be directly comparable, however.

The programme used is described by Avery et al. (2009) and is a combination of an established automated in-house method in Nick Avery's laboratory in Bristol, designed to combine detection and assay of enzymic pyridinium cross-links (Fujimoto and Moriguchi, 1978; Eyre et al., 1987; Hanson and Eyre, 1996) and pentosidine (Sell and Monnier, 1989; Monnier et al., 1999) using the same HPLC cycle. The data presented from these analyses show clearly that although there are similarities with respect to the products of glycation, the HPLC "fingerprints" of the MTT hydrolysates following 4 weeks of each different sugar treatment shown in Figure 3.6 are sufficiently unique to identify each treatment. Disregarding peak heights, all of the fluorescent products of glycation detected for glucose appear to share elution times with products from both of the other hexose sugars (galactose and fructose). There is nonetheless a clear variability between the relative peak heights of these products between the three six-carbon sugars, both with respect to the identified and unidentified peaks, and these contribute to the uniqueness of each profile.

The identity of the labelled AGE peaks is assumed from the retention times determined from analysis of the standards as no secondary detection and analysis such as mass spectroscopy was used to confirm the specific structures of the component(s) of the eluting peaks. Pentosidine is clearly a predominating product detected here, known to be formed from hexoses though to a considerably lesser extent than produced from glycation by pentoses such as ribose (Sell and Monnier, 1989; Grandhee and Monnier, 1991) and has been identified with sufficient confidence to enable it to be measured (see Table 3.1). K1 is one component of the Cmpd K complex (Bailey et al., 1995). A standard containing the other components - K2, K3 and K4 - was run but no MTT treatment peaks for these were identified in any of the treatment groups. The pyridinium enzymic cross-links (OHPyr, LysPyr) were also assigned on some of the chromatograms. With regard to the changes noted for the putative OHPyr peak, especially given this cross-link's established stability (Robins, Shimokomaki and Bailey, 1973), no mechanism for decreases in its level can be put forward except for the possibility of a mis-assignment of the peak. Even so, the reason for fluctuations in this peak's height seen with successive time periods of glucose incubation (Figure 3.2), and reductions in its height between 3 and 4 weeks of both galactose and fructose MTT treatments (see Figure 3.3 and Figure 3.4) are not known. It could be a transient fluorescent glycation product or an artefact of the experimental procedure. If not related to inter-individual variation, the reduction in this peak seen between 1 and 4 weeks in both controls would suggest the latter.

Considering the comparison between glucose and galactose in light of the "identified" peaks, the higher specific pentosidine level for galactose (Table 3.1, Figure 3.6) would be expected in light of the expected rates of reaction for the two sugars (Bunn and Higgins, 1981; Syrový, 1994; Ledesma-Osuna, Ramos-Clamont and Vazquez-Moreno, 2008). However interestingly, the K1 peak height in the case of glucose clearly showed a greater increase with respect to the control than galactose, and this was the case for all four individual samples after 4 weeks. Therefore a specific discrepancy in AGE formation between glucose and its proposed faster-acting epimer, galactose, has been identified. Inspection of Figure 3.6 shows unique peaks detected for galactose as compared with glucose and the other sugars, most notably the doublet at 22.6 and 23.0 minutes. It was considered that, of the four chromatograms in Figure 3.3, the galactose-MTT treatment chromatogram that bore the closest resemblance to the 4 week glucose trace (Figure 3.2D) was the 2 week incubation (Figure 3.3B), and an overlaid comparison of these is shown in Figure 3.7. The two chromatograms appear similar. Of the differences, the 2 week galactose treatment peak at 18.4 minutes can

be mitigated by consideration of the corresponding 1 week galactose peak (see Figure 3.3A) where this peak is taller and so more in line with the corresponding 4 week glucose peak. It should be borne in mind again that exact matching of the glucose and galactose traces in terms of stage of glycation, is difficult and was not achieved here. The other main differences, with reference to the 2 week galactose treatment trace as compared with that for 4 week glucose incubation, are the significantly lower K1 peak and the nascent doublet at 22.6/23.0 minutes.

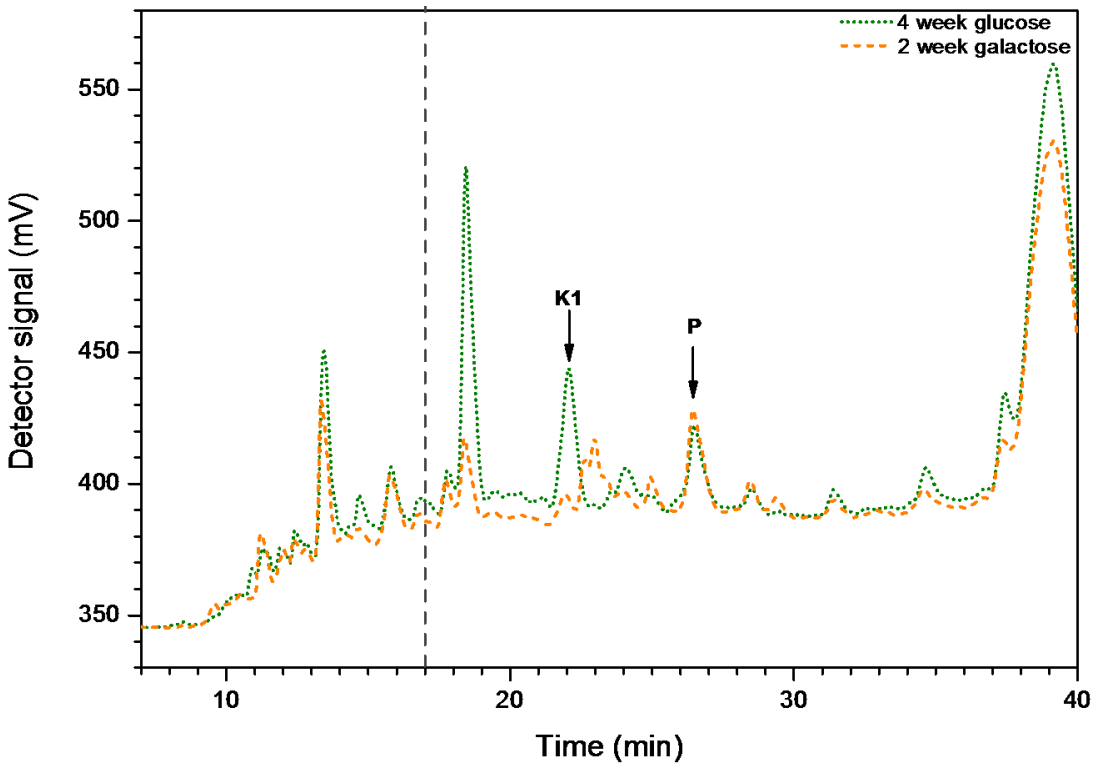


Figure 3.7 Comparison of 4 week glucose and 2 week galactose MTT treatment HPLC chromatograms. Both traces are shown on the same scale, between 7 and 40 minute elution times with 250 mV fluorescence detector signal range. The analysis was performed as described in section 3.2.4.3. The detected AGEs are labelled, based on retention times determined from standards.

Fructose treatment of MTT appears to show profoundly greater differences in the chromatogram from glucose than galactose (Figure 3.6). Although a hexose, it is a ketose rather than an aldose and it can be presumed that the shift of the reactive carbonyl group from carbon-1 to carbon-2 gives rise to a greater number of different products of glycation (Jakas et al., 2008). Clearly, it is predisposed to producing a greater amount of pentosidine under the reaction conditions chosen for this study (Table 3.1). It has previously been found that fructose glycation does result in significantly greater amounts of UV fluorophore production than glucose (Suarez et al.,

1989). This is reflected in the chromatograms shown here (Figure 3.6) and is in agreement with the differing predicted glycation reactivity of the two sugars (Bunn and Higgins, 1981). The present data do not agree with another previous study (Grandhee and Monnier, 1991), which found that glucose and fructose glycation treatments yield equivalent amounts of pentosidine. However Grandhee and Monnier (1991) did use a higher temperature (65^{o}C) and a phosphate buffer without the presence of a chelating agent to remove metal ions. It is known that phosphate can catalyse glycation (Watkins et al., 1987; Gil et al., 2005; Nemet et al., 2011) as compared with MOPS, which might itself have an inhibitory effect (Watkins et al., 1987; Nemet et al., 2011). It is also known that transition metal ions have an effect on pentosidine yield, which is dependent on the type of sugar. Removal of metal ions has been found to decrease the yield of pentosidine in the case of glucose and have the opposite effect with ribose glycation (Culbertson et al., 2003) although the experiments suggesting this were performed using a phosphate buffer system.

The ribose-treated MTT chromatograms are dominated by pentosidine (Figure 3.5 and Figure 3.6). The markedly greater amounts of this cross-link with ribose as compared with the other sugars (Table 3.1) are not unexpected (Sell and Monnier, 1989). However, it is interesting that with the exception of pentosidine and the unidentified AGE eluting at 12.4 minutes, although there are a few smaller unidentified AGE peaks the ribose incubation chromatogram is relatively featureless as compared with its galactose- and fructose-treated MTT equivalents. Despite the expectation that it is the fastest acting glycator of the four sugars considered here (Bunn and Higgins, 1981), protein-attached ribose may be more constrained in its reactions because it is smaller than its hexose counterparts or it may simply form pentosidine – a stable cross-link (Sell and Monnier, 1989) - so rapidly that other AGE-forming mechanisms have less chance to compete. As mentioned above (section 3.3.1) study of the pentosidine levels for ribose MTT treatment from each time period (Table 3.1) shows progressive significant increases in the first week of incubation and then between weeks 1 and 2, and weeks 2 and 3. The lack of a significant increase in the 4th week would seem to indicate a plateau is being reached. However the levels reported here still only constitute a maximum of <1 pentosidine cross-link per collagen molecule (<0.8% of lysine residues in MTT tissue). Slatter et al. (2008) previously published data from work on RTT that would tend to broadly support this level and suggested that during glucose treatment, pentosidine formation is lower in tendon collagen than collagen from skin by a factor of about 2.5 times, as expressed as a proportion of available lysine residues. In bovine serum albumin under the same conditions, pentosidine formation was

reported to be over 100 times higher than in tendon (Slatter et al., 2008). Therefore the plateau of pentosidine formation is unlikely to be purely a chemical diminution of potential lysine attack sites. As suggested by Slatter et al. (2008), globular proteins are not likely to be a good transferable model for the study of glycation that is structurally applicable to fibrous proteins such as collagen. The constraints of the order upon sugar molecule access (Slatter et al., 2008), positions of "glycatable" sites and the spatial feasibility of AGE and cross-link formation, as well as the changes in the ordered structure of and environment within the collagen fibril brought about by the glycation process are all likely to be important factors and have been considered previously (Tanaka et al., 1988a; Reiser et al., 1992b).

The Cmpd K AGEs were first identified in RTT collagen (Bailey et al., 1995) and have so far only been identified in RTT collagen or tropocollagen (free collagen molecules). They are therefore likely to have been formed with the involvement of hydroxylysine residues (Slatter et al., 2008) although these AGEs have yet to be characterised (Private communication from Nick Avery, 2011). Here, on the basis of its elution time, K1 appears to have been identified as a result of glucose and fructose treatment of MTT, with possible smaller amounts produced during galactose treatment (Figure 3.6). However, despite Bailey et al.'s (1995) detection of the K2 component of Cmpd K alongside K1 with glucose treatment, and K3 + K4 with ribose treatment of RTT, these three additional AGEs were not detected here. It may be that they are specific to RTT or that the non-phosphate buffer conditions employed here are not conducive to K2, K3 and K4 being formed. However, the evidence here suggests that K1 is found in MTT.

In summary, it has been shown here that all of the four sugar treatments of MTT studied over the time periods chosen showed progressive formation of novel glycation-related peaks. All of the sugars individually produced glycation-related profiles, which were specific to them. Galactose and fructose especially each produced significant novel peaks, which remain unidentified. Study of the chromatograms (Figure 3.2, Figure 3.3, Figure 3.4, Figure 3.5 and Figure 3.6) was consistent with the expected rank order of expected glycation activity, i.e. ribose > fructose > galactose > glucose (Bunn and Higgins, 1981). The measured pentosidine levels (Table 3.1) also confirmed this rank order, though was disproportionate in this regard due to the differing propensities of the respective sugar adducts to form the AGE (Sell and Monnier, 1989; Grandhee and Monnier, 1991). Allowing for its faster glycating activity, galactose did show a profile of products of glycation that resembled that of glucose. However, there

were three peaks from galactose-treated MTT which were absent after glucose treatment.

From the work in this chapter it can be seen that further studies aiming to identify some of the novel fluorescent peaks seen in this work, would be useful. The methodologies utilised could include HPLC in tandem with mass- and/or NMR spectrometry previously used to characterise products of glycation in both collagen-containing (Mikulíková et al., 2007; Slatter et al., 2008) and non-collagen-containing protein glycation models (Kawasaki et al., 1998; Biemel et al., 2002; Nemet et al., 2011). It is anticipated that such research would further insights into the differences between the chemical mechanisms involved in AGE formation following the initial reaction between collagen and the different sugars, as well as the nature of their biological effects.

From an experimental perspective, a characterisation of the discrepant peaks in the fluorescent AGE profiles resulting from glucose and galactose treatments over time periods matched for equivalence, as in Figure 3.7, might not only shed light on how minor differences in monosaccharide structure can affect progression of AGE formation, but also help confirm the viability of galactose as a faster-acting substitute for glucose for use in *in vitro* glycation studies. It is hypothesised that the nascent doublet present in the 2 week galactose trace at 22.6/23.0 minutes (Figure 3.7) is at least partially composed of an unreported member of the Cmpd K group of AGEs. Confirmation of this would require work to elucidate the structure of the components of these galactose peaks and, for the first time, K1 itself.

Although the predominant focus of this thesis is the study of structural effects of glycation on MTT type I collagen, it is considered important here to represent biochemical evidence showing that glycation has occurred during the *in vitro* monosaccharide treatments, the relative extents of glycation and the fact that the different monosaccharides are reacting and producing AGEs in mechanistically varied ways. Conversely the changes detected in the buffer and sorbitol MTT incubations were seen to be small and comparable to each other, and less than those seen for all four monosaccharides. Therefore sorbitol does not show detectable glycating activity and both of these treatments would appear to have been demonstrated to be biochemically viable controls for structural studies of MTT collagen glycation.

It is asserted that the data and findings reported within this chapter constitute an important tier of information, which provides valuable contexts in helping to interpret, and draw conclusions from, the glycation-related structural work presented in subsequent chapters.

Chapter 4: Positive staining analysis of fibrils from mouse tail tendon following different glycation treatments, using transmission electron microscopy

4.1 Introduction

The use of TEM in the study of type I collagen fibrils has commonly involved the use of positive and negative staining. Both of these were described in Chapter 1, section 1.1.4.1 and have been reviewed elsewhere in some detail by Chapman *et al.* (1990). Positive staining tends to be carried out at acidic pH and the resulting banding pattern is a reflection of the charge distribution axially along the fibril. Commonly, tungstate ions or PTA and UA are used. Anionic tungstate or PTA tends to bind to positively charged amino acid residues (Tzaphlidou et al., 1982b), especially lysine and arginine whereas UA binds to both positively-charged amino acids and negatively-charged side-chains (Tzaphlidou et al., 1982a) such as acidic glutamate and aspartate residues. The summation of charged regions along individual collagen molecules within their ordered fibrillar arrangements to give a regular repeating positive staining banding pattern along fibrils was demonstrated by Hodge and Schmitt (1960) and was instrumental in Hodge and Petruska's (1963) pivotal deduction of their model for the arrangement of collagen molecules within fibrils. The positive staining patterns have been found to be remarkably conserved for type I collagen between species (Hodge and Schmitt, 1960; Meek and Holmes, 1983; Hadley et al., 1998; Silver et al., 2001) and appear almost identical in type II collagen (Ronziere et al., 1998; Ortolani, Giordano and Marchini, 2000). This consistency means it is potentially a dependable methodology in collagen structural research. By simulation of the collagen fibril based upon the known α-chain amino acid sequences and using the correct "quarter stagger" arrangement of collagen molecules based on the known *D*-period (or progressive molecular stagger) length of 234 amino acids, it is possible to produce a histogram of predicted charge distribution. Appropriate smoothing of this to allow for resolution has been shown to produce traces that correlate well with experimental TEM/positive staining band intensity data (Meek et al., 1979; Tzaphlidou and Hardcastle, 1984).

The charged amino acids that positively stain with tungstate/PTA and UA tend to collocate along the fibrils in flexible regions with the more rigid regions in between

being more proline- and hydroxyproline-rich (Silver et al., 2001). Therefore although the relative intensities of the bands are different, a pattern of up to 12 positive staining bands has been identified as being visible by TEM after use of either stain and these are shown in Figure 4.1.

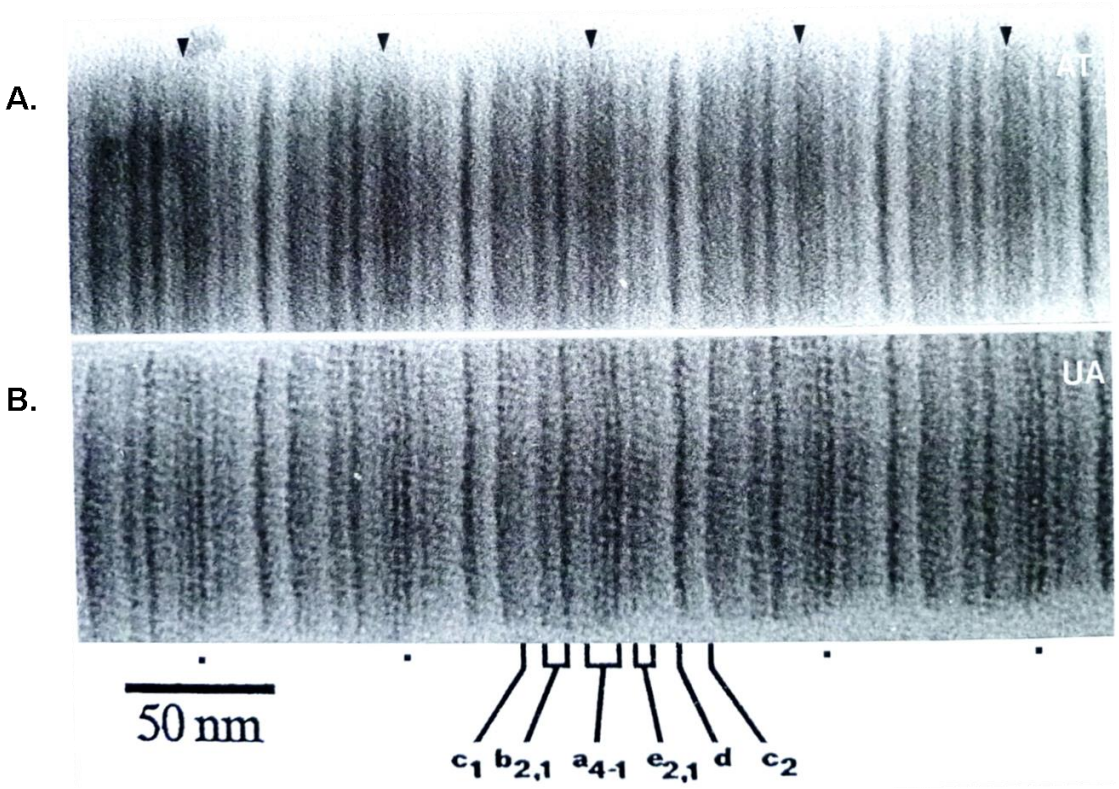


Figure 4.1 Comparison of TEM positive staining patterns along type I collagen fibrils following treatment with **A.** ammonium tungstate (AT) (1%, pH 3.2) and **B.** UA (10 μM, pH 4.4). The positions of the 11 visible positively stained bands – *a1, a2, a3, a4, b1, b2, c1, c2, d, e1, e2* – are indicated along one *D*-period beneath the micrographs. Sometimes an additional, faint *c3* band is seen between the *c2* and *d* bands. Whilst all the bands are in the same positions on both micrographs, there is some variability in their relative intensities. The arrowheads (top) are 1 *D*-period apart and indicate the *a3* bands, which are more intense for the tungstate-stained fibril as are the *b2* bands. The dots (bottom) indicate the *a1* bands, which appear more relatively-intense for the UA staining. Reprinted from Chapman, Tzaphlidou et al. (1990), © 1990, used with permission from Elsevier.

Of particular note in this work is the reaction of sugar molecules with free amine groups as the initial step in the glycation reaction (Bunn and Higgins, 1981; Monnier, 1990) and the subsequent progression of the glycation reaction and formation of AGE cross-links involving lysine and arginine residues (Iijima et al., 2000; Brady and Robins, 2001; Biemel et al., 2002; Sell et al., 2005) and so removal of positive charges along the fibrils. This would specifically affect essential functioning of the collagen-containing

tissue, e.g. interaction with other biomolecules such as PGs (Scott, 1996), loss of capacity to store energy (Silver et al., 2001). However it could also serve to provide a useful means of studying the progression of glycation.

It stands to reason that a loss of positive charge along the fibril would cause a reduction in the positive staining band intensity. Using TEM to produce electron micrographs, band intensity analysis along the fibrils would give axial plots that could yield valuable information about specific locations along the collagen *D*-period where glycation occurs. This approach was taken by Hadley, Meek and Malik (1998) in studying the effects of *in vitro* glycation by fructose along fibrils from human scleral collagen. Using PTA staining, they demonstrated for the first time that specific changes in the positive charge distribution along the fibril after *in vitro* glycation of collagen was possible. Reductions in PTA uptake were seen to have occurred predominantly in the *c1* band after 11 days' incubation, with smaller changes noted in the *b2*, *d* and *e1* band. Further, by comparison of their experimental data with a smoothed histogram of positively charged amino acids along the *D*-period in a manner demonstrated previously (Meek et al., 1979; Tzaphlidou and Hardcastle, 1984), they were able to consider the likely specific amino acid loci, which were glycated during their experiments. However, since the report of this group's work (Hadley et al., 1998; Hadley et al., 2001), this methodology does not appear to have been used to study glycation further.

Through the work presented in this chapter, it is hoped to advance use of this technique to study the changes in positive charge distribution that occur after treatment of MTT collagen *in vitro* with different sugars over different time periods. The sugars used will be glucose, fructose, galactose and ribose over 1, 2, 3 and 4 weeks as in Chapter 3 (see also Chapter 1, section 1.4.2). It is hypothesised that the preferential sites and patterns of glycation along the collagen *D*-period are different for the various sugars and notwithstanding the range of rates of reaction, that these will be identifiable by the changes in the positive charge distribution along the surfaces of the fibrils post-incubation. Therefore it is anticipated that identification of similarities in the preferred sites of glycation for the different sugars might help elucidate factors determining these sites and the functional effects of the glycation process.

4.2 Materials and methods

4.2.1 Materials

For general laboratory chemicals and consumables used throughout the experiments in this entire work, refer to Chapter 2, section 2.1. Glass microscope slides with frosted ends (M353/DF/1.2) and G300 mesh copper TEM grids (GG007/C), 3.05 mm diameter, were obtained from TAAB Laboratories Equipment Ltd (Aldermaston, Bershire, UK). PTA was obtained from Agar Scientific (Stansted, Essex, UK). The polyetherimide (PEI) (0.28%) solution in ethylene dichloride had been previously prepared for laboratory use within the School of Optometry and Vision Sciences. 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) was obtained from Sigma-Aldrich. Bovine testis hyaluronidase (H-3506) also originated from Sigma-Aldrich and was kindly donated by Dr R D Young.

In addition to that described in Chapter 2, section 2.1, computer software used in image data analysis in this section was ImageJ v. 1.46i (Rasband, 2011)

4.2.2 Coating TEM grids

Within a fume cupboard, a dedicated Coplin staining jar was filled to three-quarters full with 0.28% PEI solution in ethylene dichloride. A glass microscope slide with frosted end was quickly dipped into the liquid so that the entire clear portion was immersed and held there for ten seconds. It was then quickly removed and any excess solution allowed to drip back into the jar. The jar lid was replaced and the slide held upright inside the fume cupboard, until it could be seen to have dried completely. The dipping-drying procedure was repeated a second time on the same slide in exactly the same way.

Using a razor blade the resultant film coating on the slide was scored on all four sides at >1mm distance from the edge of the film/microscope slide such that an entire, approximate rectangle had been described. A large glass beaker was filled with dH_2O. The coated slide was held over the water with the narrow, unfrosted edge at the bottom and parallel to the surface, inclined at an angle of approximately 40^o from vertical such that the scored film-face was uppermost. The bottom edge of the slide was touched to the water surface and then dipped just below it until the leading edge of film could be seen floating on the water surface. Keeping the angle constant, the slide was steadily

immersed further into the water until the entire score-defined portion of the film had floated onto the water surface and detached itself from the slide. The slide was then removed from the liquid, being careful not to disrupt the floating film layer.

Using a pair of fine-tipped tweezers, a number of G300 mesh copper TEM grids were laid gently, shiny side down, onto the floating PEI film, such that they were at least 2mm away from the film edge and 1 mm apart. In addition, a 1 cm grid-free gap was left at one of the narrow ends of the film.

A second microscope slide, coated in the same way as the first but not scored, was then held vertically over TEM grid-free end of the floating film, such that the narrow edge was over the centre of this empty region, parallel to the water surface and the narrow end of the film. The slide was touched to the film/water surface, and then steadily and slowly pushed all the way into the water, such that the film was pushed into close contact with the second slide. When the film was withdrawn, the film was seen to have adhered to the coated slide, with placed TEM grids sandwiched between film and coated slide. The slide was left leaning within a dry place such as inside a drawer, for at least 24 hours, allowing it to dry.

After this time, employing care to avoid disrupting the PEI film lying over the copper mesh, the coated grids were cut free from the surrounding film by use of a razor using a binocular microscope for guidance. Using a pair of fine-tipped tweezers, they were then transferred in turn into a grid storage case until use.

4.2.3 Preparation of MTT

The obtaining of mouse tails and MTT dissection process were described in Chapter 2, section 2.3. All mouse tails used in the method optimisation experiments were of indeterminate age and were already in storage in Dr Justyn Regini's laboratory within the School of Optometry and Vision Sciences at -20^{o}C for an unknown period of time. The tissue used in the glycation experiments in this chapter was all from recently killed 3 month old female mice, their tails having been stored in the freezer in the interim as previously described. Tendon tissue was dissected into 0.9% NaCl solution containing 1% v/v PIC. Following complete tendon extraction from each tail, all tendon tissue was transferred into small plastic bags which were sealed, labelled and stored at -20^{o}C.

4.2.4 MTT glycation incubations

The general conditions used for the glycation incubations have been described in Chapter 2, section 2.4. For each, using fine scissors to cut fragments from the stored tendon, approximately 1 mg of MTT was weighed and transferred into 10 ml glass test tube. A volume of 2 ml of the appropriate glycation or control solution was added to each tube plus 20 μl of PIC solution. All tubes were vortex-mixed for a few seconds and checked to make sure the fragments of collagen remained fully immersed in the incubation solution at the bottom of the tube. Each was sealed with a piece of Parafilm over the top and then placed in individual holes in an electric heating block set at 37°C. Each tube was checked every one or two days and gently mixed to maintain solution homogeneity and restore any condensation to the incubation mixture.

The MTT incubations were carried out individually over 1, 2, 3 and 4 weeks in MOPS-saline buffer and the same containing 0.2M sorbitol (controls), plus buffer containing 0.2 mol/l of each of the sugars fructose, galactose, glucose and ribose.

At the end of the incubation periods the glycation or control solution was removed from each tube using a Pastette and discarded, using care to ensure the MTT fragments remained *in situ*. To the tube, was added a further 2ml of 0.9% saline solution. This was thoroughly vortex-mixed for 10 seconds, the collagen allowed to settle to the bottom and the saline decanted off to waste with a fresh Pastette, again leaving the collagen in the tube. The collagen was washed in this way, twice more in fresh saline. Another 2ml of the saline were then added to the tube and vortex-mixed again.

4.2.5 Transfer of MTT collagen fibrils onto coated TEM grids

The post-incubation MTT fragments in 2 ml of 0.9% saline were vortex-mixed and immediately poured into a circular Whatman grade 541 filter paper of 11 cm diameter that had been corrugated radially and was sitting in and supported by a small plastic funnel. The tube was checked to make sure that all of the MTT pieces had been dispensed. Once the supernatant had fully percolated through the filter paper, the tissue fragments were carefully picked off the filter paper with a pair of fine-tipped tweezers. With a pair of fine scissors, these were chopped up into pieces that were as small as practicably achievable, which were deposited into a small mortar. To the mortar, 1 ml of either dH_2O or 0.9% saline (see section 4.3.1.1) was added

proportionally per mg of MTT as originally weighed out and the tissue was continuously ground up with a pestle for a period of 5 minutes.

Immediately following the grinding stage, the ground MTT-containing suspension was taken up into a Pastette, avoiding picking up any aggregates of material that may have formed. A drop of this was deposited onto a PEI-coated (see section 4.2.2) G300 copper mesh TEM grid, which had been placed coated side-uppermost onto a plastic Petri dish. Care was taken to ensure the grid remained in position during drop dispensing such that it ended up fully immersed in the suspension drop. The Petri dish-grid-drop arrangement was covered by an inverted beaker to protect it from contamination with airborne particles and slow evaporation, and left for 1 hour (time periods of 30 and 90 minutes were initially trialled for this – see section 4.3.1.1). After this time, the beaker was removed and the drop removed by soaking it into a small piece of tissue paper, taking care that this would not come into contact with the grid. The grid was then transferred carefully using fine tipped tweezers, onto a labelled piece of filter paper contained within a small plastic Petri dish. The lid was put in place and the dish placed in a designated drawer and left overnight to dry fully.

In addition, as a “0 week” untreated baseline comparator sample, approximately 1 mg of stored tendon from a subject mouse was weighed out and chopped up into the mortar with fine scissors. This was ground up and deposited on a coated TEM grid as described above, in the same way as for the incubated MTT samples.

Remaining suspension was decanted into 1.5 ml Eppendorf tubes, which was sealed, placed in a small plastic bag and stored at -20^{o}C.

4.2.6 Positive staining of the TEM sample grids

The following procedure was optimised using staining with 1% and 2% PTA over different time periods between 2 and 30 minutes. The protocol described here was the one deemed best after visual comparison of the resulting electron micrographs, and therefore used.

In terms of setup, staining was carried out very much in line with that employed by other workers using TEM within the School of Optometry and Vision Sciences. A 1% solution of PTA in water, pH 2.2 (see Hulmes et al. (1981)) was used. To remove any insoluble debris, the PTA solution used was placed into an Eppendorf tube and spun in a microcentrifuge at 3000 rpm for 3 minutes and the supernatant filtered using a microfilter attached to the end of a plastic syringe. Via this syringe-microfilter

arrangement, a natural drop of PTA solution was placed onto a sheet of Parafilm. Using a pair of fine-tipped tweezers, the sample-containing coated grid was placed onto the drop such that it floated with the coated/sample side in contact with the stain solution. This arrangement was covered with an inverted plastic tray for protection and left for 10 minutes. After this time, using fine-tipped tweezers, the grid was transferred onto a similar drop of dH_2O - itself filtered using a microfilter - to float for 2 minutes. After this time, the grid was transferred again in the same way, to two successive drops of dH_2O, each for 2 minutes. Finally, held in fine-tipped tweezers, the grid was rinsed under a flow of several drops of dH_2O from a plastic syringe-microfilter arrangement and then placed onto a labelled filter paper in a Petri dish. The lid was placed on the dish and it was placed in a drawer and left to dry fully overnight.

4.2.7 TEM study of collagen fibrils

Stained sample grids were examined using the JEOL 1010 transmission electron microscope. The acceleration voltage was set at 80 kV for all these experiments. Grids were systematically scanned for collagen fibrils at a low magnification of 1000 – 5000x. Located fibrils were inspected at higher magnifications and ones with a clear banding pattern and minimal background contamination were selected, viewed at an appropriate magnification and photographed with the Gatan ORIUS SC1000 CCD camera attached to the microscope. All images were saved using the .dm3 format.

For electron micrographs used to produce the glycation and control data presented in this chapter, a magnification of 30,000x was used. Having initially used a live Fourier transform function available in the Gatan software to aid in focussing of images, each fibril was photographed using a through focus series such that selection of the clearest image could be carried out later for data analysis purposes. For each treatment group, it was aimed to capture micrographs of at least six fibrils.

4.2.8 Use of hyaluronidase to remove background grid contamination

A rationale was established to attempt removal of non-collagenous background contamination seen on the grids using bovine testis hyaluronidase enzyme's previously demonstrated cross-reactivity (Madinaveitia and Stacey, 1944; Lyon and Phelps, 1981) to break down and remove the dermatan sulphate and chondroitin sulphate GAG

chains attached to the PGs present in the MTT tissue (see Chapter 1, Section 1.1.5.2) and this was trialled. This treatment was carried out directly on a sample-containing TEM grid prior to staining. An aqueous drop containing 1 mg/ml of hyaluronidase in buffer containing 5 mM HEPES and 0.9% NaCl adjusted to pH 7.4 was placed on a sheet of Parafilm. Using fine-tipped tweezers, a coated TEM grid with collagen from the 3 week buffer control group deposited upon it was placed so that it floated upon the drop with the sample side in contact with the liquid. This was covered and left at room temperature for 60 minutes. After this time, the grid was washed using dH_2O and put aside to dry, all in a manner identical to that described for grid staining in section 4.2.6. A second control grid was prepared from the same sample in exactly the same way, except the buffer upon which it was incubated contained no enzyme.

Both grids were stained as described (section 4.2.6), inspected by TEM and the results compared.

4.2.9 Band intensity extraction and standardisation

For data analysis in this study, data was extracted from micrographs in .dm3 format to produce band intensity plots for all viable *D*-periods along each selected fibril. This was accomplished using ImageJ (Rasband, 2011). Fibrils were selected for use on the basis of positive staining banding clarity and least background contamination. For each fibril selected for use, the clearest image was selected from the differently-focussed images captured.

Each image was opened using ImageJ, rotated such that the axis of the subject fibril was horizontal and oriented left to right, with the positively-stained bands in the order *e2-e1-d-c2-c1-b2-b1-a4-a3-a2-a1*. The image was inverted to negative form, such that the dark, positively stained bands appeared bright. Therefore upon band intensity extraction, the more darkly stained bands would produce higher peaks on a band intensity plot. To extract band intensity data for each *D*-period, a rectangle was defined over it. The selected origin (leftmost extremity) of this area was placed over the *e1* band and the terminus (rightmost extremity) was positioned between the *d* and *c2* bands of the *D*-period immediately to the right of the subject one. Therefore a leeway-length of >1 *D*-period was defined per extracted dataset. The dataset for the selected area was then extracted using the “Plot Profile” function. The numerical data associated with the plot were copied into an Excel spreadsheet. In this way, starting at the left side of the fibril on the micrograph and moving the rectangle stepwise-right

along the *D*-periods, the band intensity data were extracted for the entire useable length of the fibril.

All datasets were trimmed to a span of exactly one *D*-period with precisely the same start and end positions such that they would all be directly comparable and interpolated so that all plots would consist of the same number of points. For this, a dataset for a single *D*-period with clearly-defined bands, which well-represented the previously-established positively-stained pattern expected (see Chapman et al. (1990)), were selected as an alignment standard. This dataset was cut to define a plot for exactly one *D*-period. Using the OriginPro graphing software, all other individual band intensity plots were first baseline-levelled by subtraction of an appropriately-sloped straight line to compensate for any background brightness gradient present on the image. The subject plot was then moved over the standard with adjustment of its length along the *x*-axis (representing *D*-period axial distance) if necessary, until as close a visual alignment as possible was achieved. The points on this plot corresponding to the ends of the standard were located and defined as its new *D*-period termini – discarding all numerical data outside these – and interpolated to 468 data points. The adjusted subject *D*-period data were then plotted against the standard trace, the latter dataset also having been interpolated to 468 points. Using a least-squares analysis, a best-fit straight line was drawn. Using the slope and intercept of this, the values of the subject dataset were scaled linearly against the standard such that slope and intercept of the correlation best-fit line became 1 and 0, respectively.

For all treatment groups in this study (section 4.2.4), in order to reduce noise from background contamination it was deemed appropriate to use 20 individual *D*-period band intensity traces consisting of five datasets per fibril from the four best fibrils analysed. Where (as in most cases) more than five *D*-periods had been considered to exhibit sufficiently high banding quality without background contaminant interference, all aligned, interpolated and scaled datasets taken from a given fibril were cross-correlated with all others. On the basis of Pearson r correlation analysis, the five *D*-periods whose data best correlated with one another were selected and used from that fibril. These five datasets were then combined in an OriginPro spreadsheet, with those from three other fibrils and a set of mean values with SDs were calculated for all 20 individual band intensity traces.

4.2.10 Calculation of difference traces for MTT sugar treatment groups

All individual and mean traces produced in section 4.2.9 were in alignment so when plotted on the same axes, all points of every plot would represent an equivalent position. Band intensity difference data were calculated from mean and SD data. First, mean band intensity data for a given sugar incubation MTT collagen group were cross-correlated with the mean data for the buffer control collagen incubated over the same time period. Using the gradient and intercept from a least-squares linear fit analysis, the sugar treatment group data were scaled against the buffer control such that slope became 1 and the intercept became 0. The scaled mean sugar treatment relative band intensity data for each *x*-axis point was then subtracted directly from that obtained for the control. The SD data for the sugar incubation were also multiplied by the above gradient to correct the statistical indicator of spread for the factor change employed in scaling the mean data. The SDs for the buffer control and sugar treatment group for each data point were converted to variance (i.e. squared), added and the square root of the sum taken as one SD for each point along the difference plot for that sugar.

4.2.11 Simulated positive staining band intensity profile for one *D*-period

The rationale and means of simulating a mouse type I collagen fibril *D*-period was described in Chapter 2, section 2.5. Represented on an Excel spreadsheet embedded in the OriginPro graphing software, the simulated fibril created contained N- and C-telopeptides in the hairpin-folded conformation described (Chapter 2, section 2.5) and was used here to chart the distribution of positively charged amino acid (lysine and arginine) residues. This was done by scoring every amino acid position along each collagen molecule of the fibril as 1 for every “K” or “R” and 0 for all others and then summating the scores across the simulated fibril. This summation could be plotted as a histogram, which would repeat every 234 amino acid positions in line with this being the *D*-periodicity. Using a Savitsky-Golay algorithm, this was smoothed into a curve to achieve comparability with the untreated MTT positive staining mean trace. The curve was moved step-wise over the one *D*-period experimental trace to achieve best alignment correlation. Having overlaid a single *D*-period of the smoothed, simulated banding pattern over the experimental data plot, different smoothing window sizes

were tried in the Savitsky-Golay smoothing function to achieve closest resemblance between the two.

4.3 Results

4.3.1 Optimisation of method

4.3.1.1 Dispersion of collagen and deposition onto TEM grids

For the purposes of producing fibrils that were discernable by TEM viewing, various means of breaking the tendon tissue up were initially trialled (section 4.2.5). Three of these succeeded in producing dark aggregated clumps of matter with no fibrils evident, they being: use of a manual Potter-Elvehjem homogeniser; freezing in liquid nitrogen followed by treatment with a dismembranator; and sonication. Only the mortar and pestle method described in section 4.2.5 was successful in producing fibrillar material visible on the grid post-deposition, hence the adoption of this mechanical dispersion method.

Dispersion of the collagen into dH_2O and 0.9% saline were tried along with varied grid deposition times of 30, 60 and 90 minutes. With regard to the optimal time period for allowing the fibrillar material to deposit on the coated grids, 60 minutes gave a satisfactory population density as viewed under the TEM. With the 30 minute time, it was found to be difficult to find fibrils and after 90 minutes, the grids appeared crowded. Therefore a 60 minute deposition time was selected. With respect to mechanical dispersion medium, Figure 4.2 shows a comparison of positively stained MTT fibrillar appearance under the TEM following dispersal into dH_2O and 0.9% saline. This demonstrated the superior clarity of banding and less contaminating material adhering to the fibrils after use of the saline and so this medium was chosen for use in the remainder of these experiments.

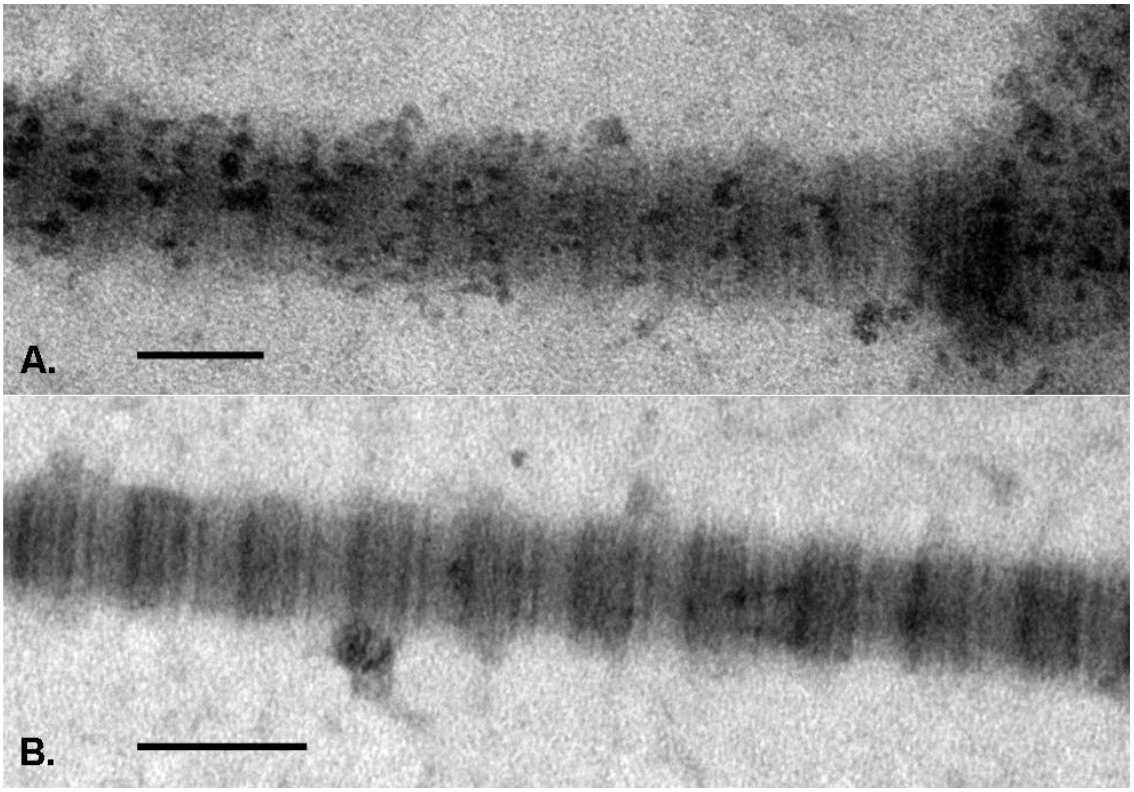


Figure 4.2 Two MTT collagen fibrils as seen under the TEM after mechanical dispersion in **A.** dH_2O and **B.** 0.9% saline. It can be seen that there is more attendant non-collagenous material associated with the fibril on the grid after dispersal in water than in the saline. The repeating positive staining banding pattern is therefore more clearly visible with saline use. Both samples were stained with 2% PTA, pH 1.9 (unbuffered) for 30 minutes. Micrographs taken at a magnification of 20,000x. The scale bar at the bottom of each micrograph represents 100 nm.

4.3.1.2 Staining

The possibility of improving the staining procedure was investigated, as stated in section 4.2.6. Concentrations of 1% and 2% PTA in dH_2O (unbuffered), pHs 2.2 and 1.9 respectively. Grids of MTT collagen, dispersed and deposited as optimised in section 4.3.1.1, were stained with both concentrations of PTA, for 2, 5, 10 and 20 minutes and after washing and drying, these were inspected under the TEM. Grids stained with even 2% PTA for 2 minutes showed faint banding and over 5 minutes, the results were variable. Staining over 10 minutes with both concentrations gave good clear banding patterns and it was decided to use 1% PTA solution to stain grids over 10 minutes (see Figure 4.3).

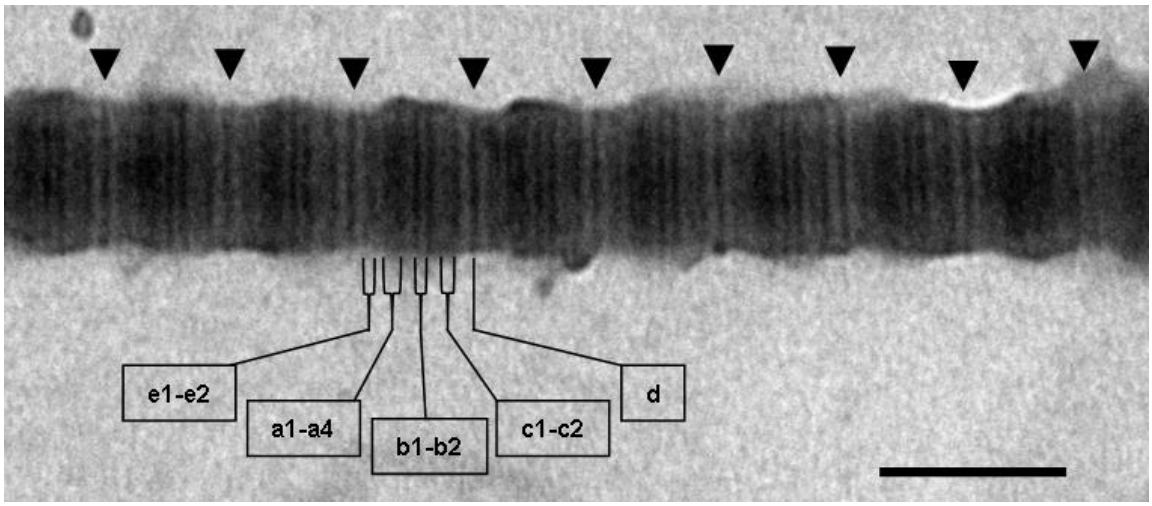


Figure 4.3 Result of staining MTT collagen on a grid, with 1% PTA solution (pH 2.2) over a 10 minute incubation time as seen under the TEM. The positive staining bands are clearly visible and as expected (see band labelling below fibril). Arrowheads above fibril point at successive *d* bands and indicate *D*-periodic spacings. Taken at a magnification of 30,000x. The scale bar, bottom right, represents 100 nm.

4.3.1.3 Hyaluronidase treatment

Treatment with of TEM grids containing MTT collagen with hyaluronidase prior to staining, as described in section 4.2.8 did not appear to reduce background contamination as compared with the control. Indeed some of the enzyme treatment grids appeared to show increased background contamination. It was speculated that this could be due to the enzyme binding to GAG chain components of PGs, which were poor substrates or even inhibitors of the enzyme (Lyon and Phelps, 1981). The additional background protein component might then take up the PTA stain. Hyaluronidase treatment was therefore not utilised any further in these experiments.

4.3.2 Band analysis of fibrils from untreated MTT

The band intensity analysis of the untreated MTT collagen fibrils selected is summarised in Figure 4.4 (see section 4.2.9, also). The resultant mean trace was overlaid and aligned to best correlation with the smoothed histogram trace (section 4.2.11), which predicted the distribution of positive charge along the *D*-period based on the simulated fibril produced as described in Chapter 2, section 2.5. Figure 4.5 shows the comparison between the experimental mean band intensity trace and that predicted from the histogram after smoothing. The Pearson r value for the correlation between

the Savitsky-Golay-smoothed histogram data shown and the experimental trace was 0.808.

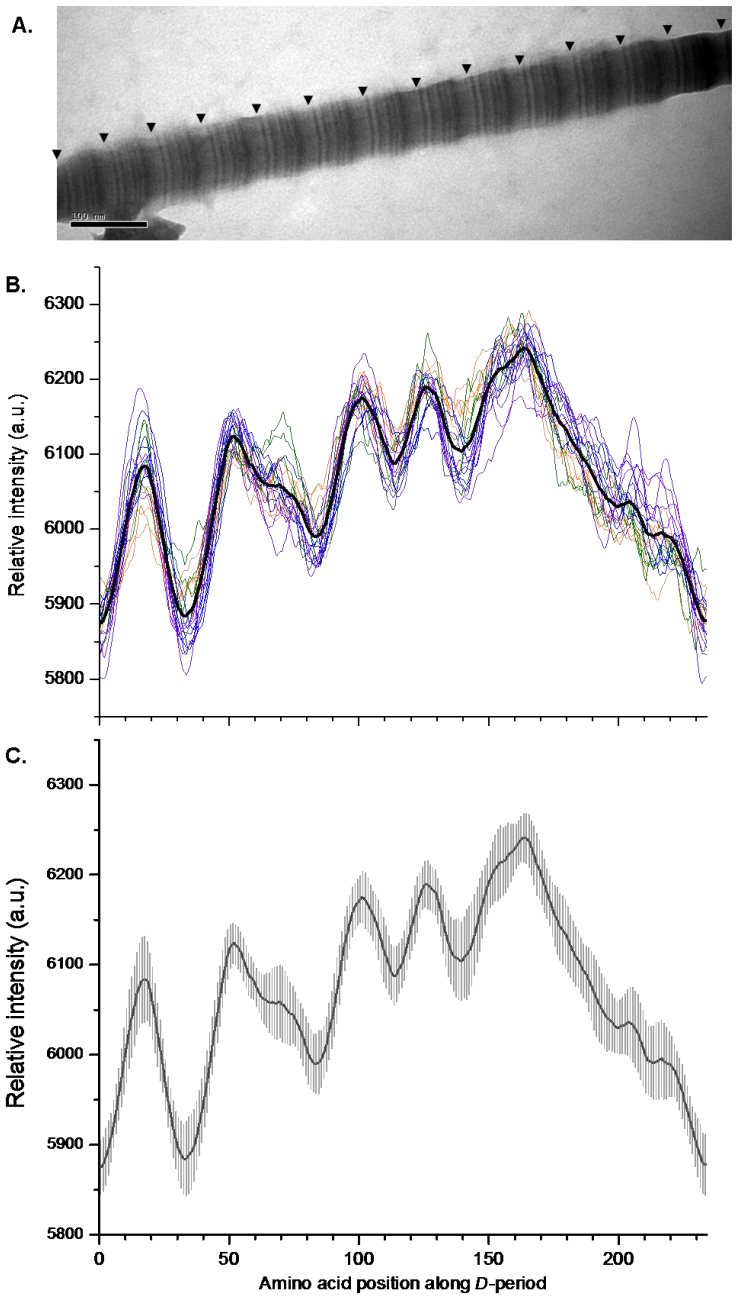


Figure 4.4 **A.** Electron micrograph of a PTA positively-stained untreated MTT collagen fibril, taken at 30,000x magnification, with the scale bar at bottom left showing 100 nm; **B.** individual positive band intensity traces (thin, coloured lines) are shown, n = 20, representing the five best cross-correlating *D*-period datasets from four fibrils (see section 4.2.9), each colour representing the datasets from a given fibril (blue representing traces from the fibril shown in **A**) and with the thick black line representing the mean data trace; **C.** mean trace from **B** with the closely spaced vertical lines along its length representing ±1 SD. All data traces represent the band intensity distributions between the points equivalent to those denoted by the black arrowheads (added) along the fibril shown in **A**.

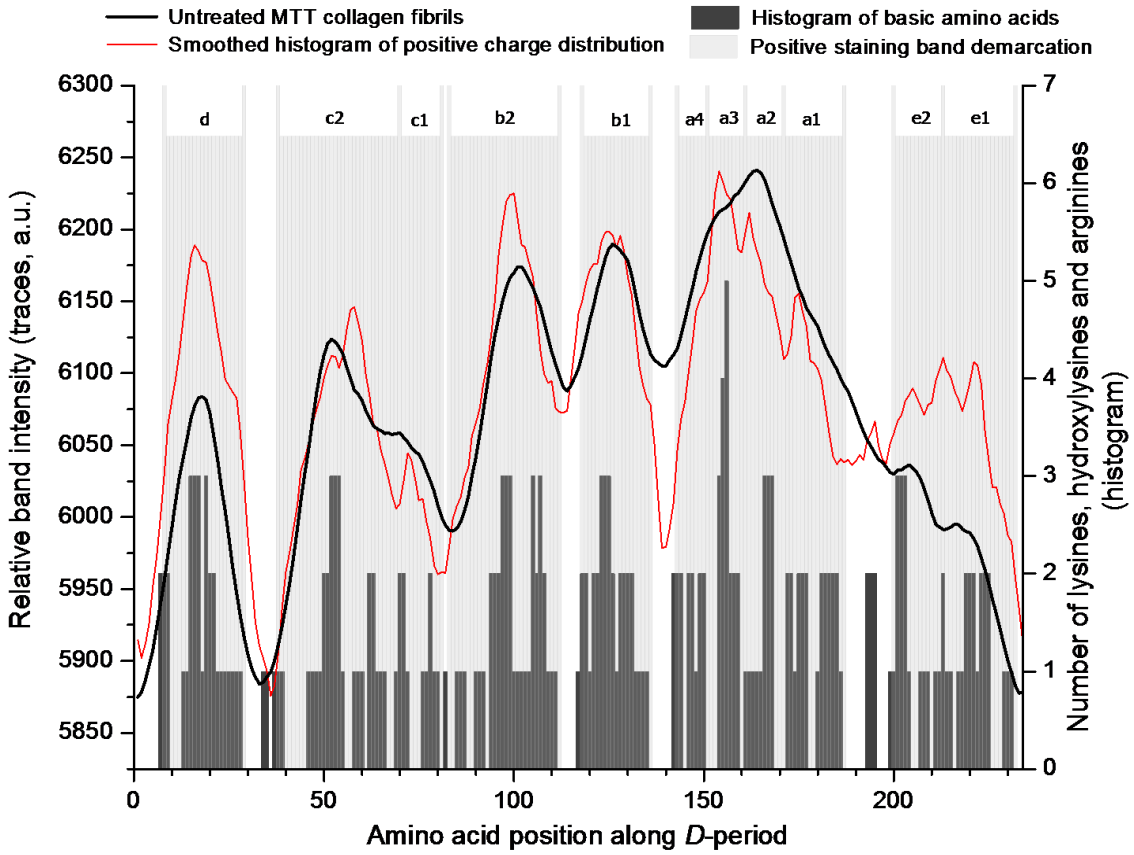


Figure 4.5 Mean band intensity trace (n = 20) for untreated MTT collagen fibrils superimposed onto a histogram of lysine and arginine (positively charged amino acids) distribution and a trace representing the smoothed histogram. The experimental and predicted data are optimally aligned to best correlation. The smoothed histogram trace was linearly-scaled against the experimental data enabling them to be representatively plotted against the same relative band intensity ordinate. Smoothing was carried out using a Savitsky-Golay function, employing a 2nd order polynomial with a window size of 34 points along the abscissa. The positive staining band demarcations as best fit these plots are shown in gray and labelled (top). Direction of fibril: N→C is left to right.

4.3.3 MTT glycation incubations

Even before data extraction and analysis of the micrographs, it was clear from visual inspection that the relative band intensities of certain glycation treatments had changed markedly. Examples of these are shown in Figure 4.6, in comparison with their respective buffer controls. The 1 week galactose treatment was the earliest obvious change detected in this way, which was unexpected given the fact that both fructose and ribose were both expected to glycate at a higher rate than this sugar (Hayward and Angyal, 1977; Bunn and Higgins, 1981). The 1 week galactose result was checked, first by re-depositing frozen excess suspension of collagen from that treatment group

onto a fresh TEM grid, staining and viewing and subsequently by repeating the entire incubation. Inspection of the grids from both repeat experiments confirmed the magnitude of this observed change.

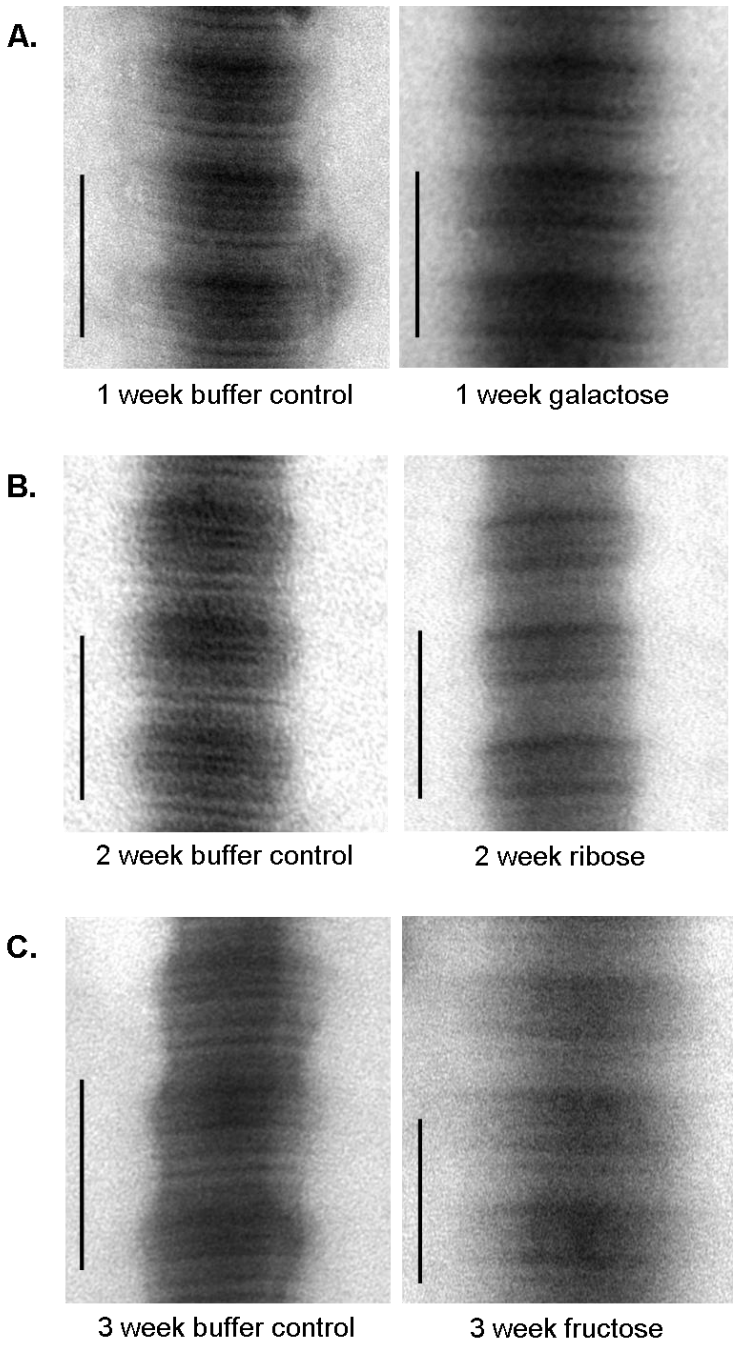


Figure 4.6 MTT collagen fibrils from three glycation incubations (on the right) compared with the buffer controls incubated for the same time period (left). For all three glycation treatments, it was evident from the micrographs that the banding pattern had changed even prior to band intensity analysis. These comparisons are for **A.** galactose incubation over 1 week, **B.** ribose incubation over 2 weeks and **C.** fructose incubation over 3 weeks. Three entire *D*-periods are shown for each sample. All micrographs were taken at a magnification of 30,000x. The vertical bars at the lower left of each image represent 100 nm.

In the case of the 3 week fructose (Figure 4.6C), it can be seen that the edges of the fibril are less distinct than the others. This phenomenon was unique for this sugar after this time period and it was noted that most of the fibrils had become more diffuse, as shown in Figure 4.7. Interestingly, this apparent feature of 3 week fructose treatment was not seen after 1, 2 or – surprisingly - 4 weeks of fructose treatment or after any time period of incubation with any of the other sugars.

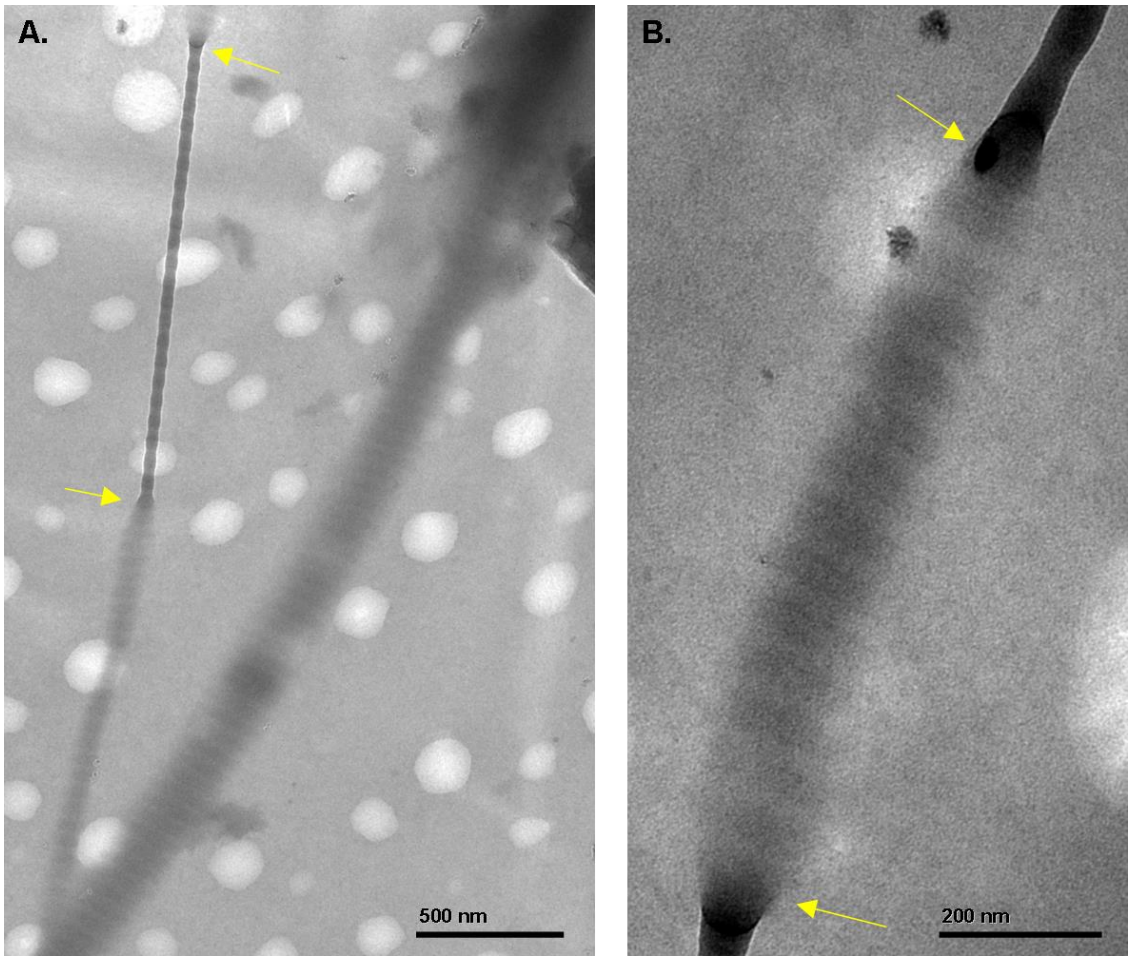


Figure 4.7 Micrographs of MTT collagen fibrils incubated for 3 weeks in fructose. **A.** was taken at 5,000x magnification and **B.** taken at 20,000x. Scale bars are at the bottom-right of both images. The yellow arrows indicate transition points between lengths of fibril with normally-packed morphology and a more diffuse and expanded appearance.

Also worthy of note is the observation that after 2, 3 and 4 weeks incubation in the ribose solution, fibrils were seen to be generally straighter and have less other, non-collagenous material associated with them on the TEM grid. It was noted that these samples were notably less resistant to mechanical dispersion in the mortar with the pestle than the other groups. The 3 and 4 week ribose incubations also showed

development of visible brunescence (Brennan, 1989a; Laroque et al., 2008). This was faint after 3 weeks, and more obviously after 4 weeks.

The extracted mean traces for all treatment groups after 1, 2, 3 and 4 weeks incubation are shown in Figure 4.8. Each plot represents the mean ±1 SD of 20 *D*-period band intensity traces from a single grid (see also, section 4.2.9), therefore from an individual mouse. MTT tissue from two mice was used in these experiments. Samples from one mouse were used for all the 1 and 4 week incubations, and also for preparation of the untreated "0 week" grid (section 4.2.5). The MTT used for all of the 2 and 3 week incubations were from a second individual.

In the first instance, it appears from Figure 4.8 **B-E** that over the four weeks of incubation, the relative band intensities in the buffer control traces did progressively change. From this, the buffer solution does not appear to be inert and may have had an effect on the positive charge distribution itself. It was considered important therefore, that analysis of each sugar's positive staining intensity trace for each time period was performed with due regard to the same incubation time for the buffer control.

The glucose band intensity traces following the four time period incubations are shown in Figure 4.8, **F-I**. Comparing these with the buffer control traces, they appear to be similar with no specific, identifiable differences. This is reflective of the relatively slow rate of glycation anticipated for this sugar (Hayward and Angyal, 1977; Bunn and Higgins, 1981).

For galactose (Figure 4.8, **J-M**), the most striking difference as compared with the buffer control is seen after 1 week of incubation (see also Figure 4.6A). There has been a considerable loss of intensity in the *b1* and *b2* bands as well as a reduction in the *d* band. The *e1* and *e2* bands also have almost completely disappeared. The 2 week galactose treatment trace is difficult to differentiate from the buffer control, showing an apparent recovery of the reduced or lost bands from the 1 week incubation trace. The 3 and 4 week incubation traces (Figure 4.8 **L** and **M**) show subtler differences mainly on the basis of differences in relative peak heights but with all bands essentially present, though after 3 weeks, there does appear to have been a comparative loss of *a1* and the *e1* and *e2* bands.

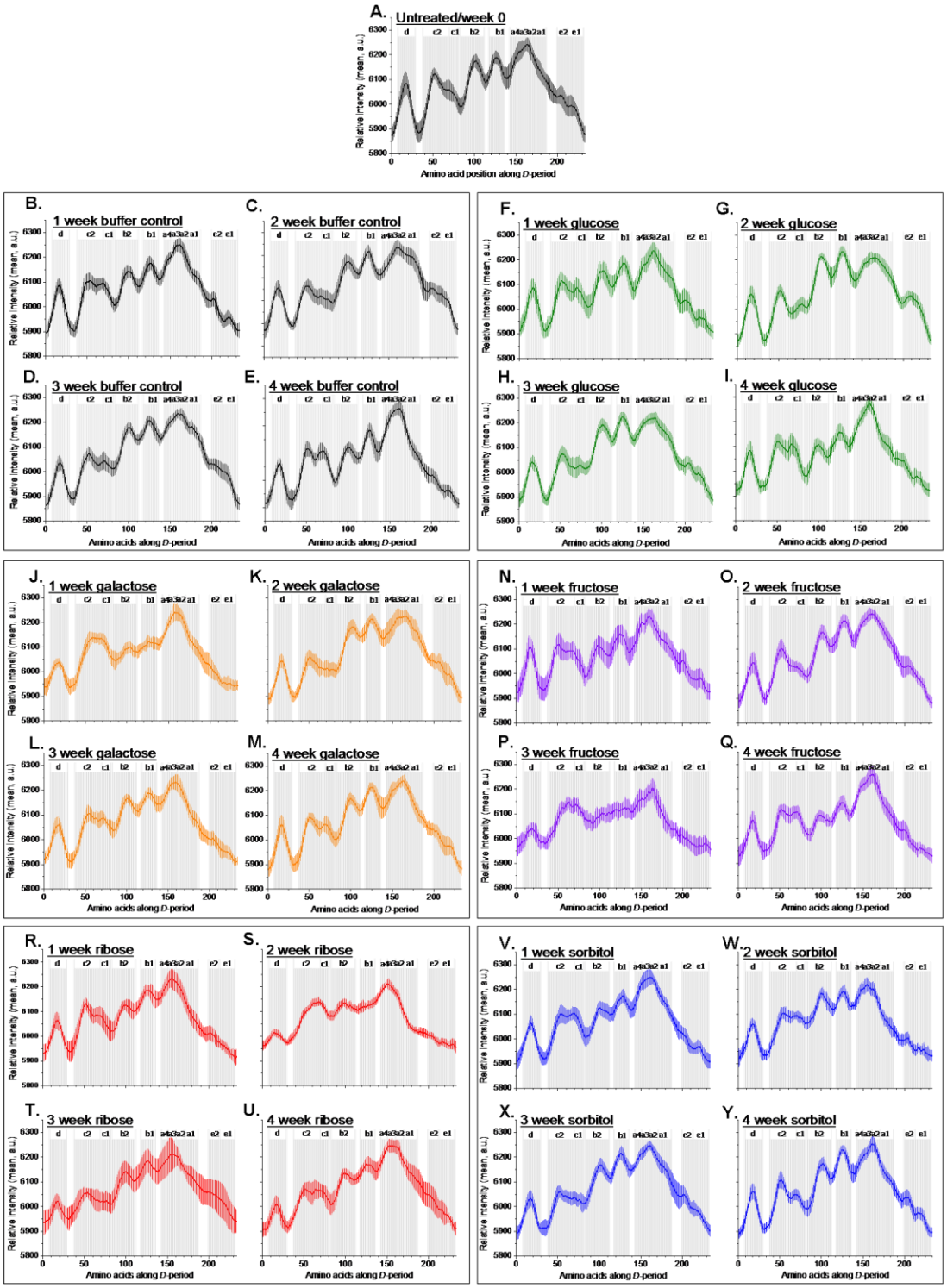


Figure 4.8 Mean TEM positive staining band intensity traces for one collagen *D*-period for all groups studied: **A.** untreated (0 week incubation) collagen, and then for 1, 2, 3, and 4 weeks incubation respectively, **B**, **C**, **D**, **E** in glycation buffer solution, **F**, **G**, **H**, **I** in 0.2M glucose (green traces), **J**, **K**, **L**, **M** in in 0.2M galactose (orange traces), **N**, **O**, **P**, **Q** in 0.2M fructose (purple traces), **R**, **S**, **T**, **U** in 0.2M ribose (red traces) and **V**, **W**, **X**, **Y** in 0.2M sorbitol (blue traces). All traces are averages of 20 individual *D*-period datasets from the same grid/individual mouse. The coloured region on either side of each line represents ±1 SD. Band positions are shown as vertical gray strips and are the same as in Figure **4.5**. Orientation of bands: N→C is left to right.

In the case of fructose (Figure 4.8, **N-Q**), the trace shapes appear to be almost identical to the corresponding buffer control traces with the exception of the 3 week trace (Figure 4.8, **P** versus **D**, see also Figure 4.6C). This fructose treatment group appears to have lost definable *b1*, *b2*, *e1* and *e2* bands, with a comparative reduction in *a1* band intensity. It also shows a general loss of peak definition. This is interesting when considering the more diffuse fibril morphology also seen for the 3 week fructose-treated fibrils (Figure 4.7).

With regard to ribose (Figure 4.8, **R-U**), the 1 week incubation appears almost identical to the buffer control for that time period. However at 2 weeks ribose incubation (Figure 4.8, **S**), there has been a significant change analogous to that seen for the 3 week fructose treatment (see also, Figure 4.6B). After the 3 and 4 week ribose incubations (Figure 4.8, **T** and **U**) respectively, there appears to be a progressive relative restoration of the peaks lost after 2 weeks but in both cases, there is still an apparent diminution of general peak definition as compared to their buffer control counterparts (Figure 4.8, **D** and **E**).

The sorbitol incubations (Figure 4.8, **V-Y**) were included as a non-glycating osmotic control, as used previously by Tanaka et al. (1988a). Although unlike galactose, fructose and ribose, none of the sorbitol incubations produced a striking loss of peaks, there were subtler changes in relative peak heights as compared with the buffer control, which were evident for all the time periods except the 1 week incubation (Figure 4.8, **V**). For example, after the 2 week sorbitol incubation (Figure 4.8, **W**) there was a relative loss of *a1* and the *e1-e2* bands.

For the more definitive location and analysis of band intensity changes, difference traces were calculated as described in section 4.2.10. All difference traces showing significant band intensity changes are shown in Figure 4.9 to Figure 4.13 for glucose, galactose, fructose, ribose and sorbitol, in order.

For glucose (Figure 4.9), there is just one region of location-specific positive charge removal of borderline-significance after 2 weeks of incubation. This corresponds with amino acid positions 159 and 160 within the *a3* band, in the gap region immediately adjacent to the boundary between overlap and gap regions (i.e. close to the C-telopeptides). No significant reductions in relative positive staining band intensity were seen with glucose, following the 1, 3 or 4 week incubations.

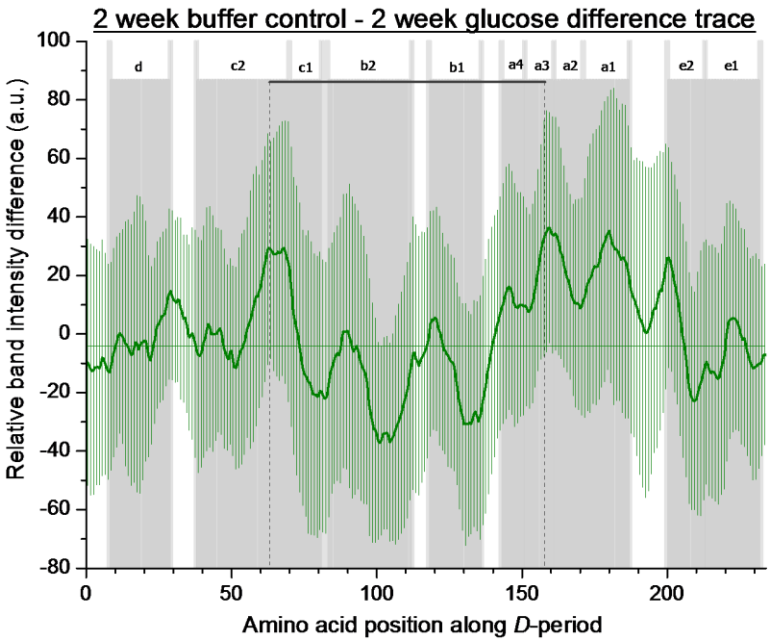


Figure 4.9 Mean band intensity difference trace for the 2 week glucose incubation over one *D*-period of MTT collagen fibrils. The trace represents a subtraction of the mean 2 week glucose band intensity trace from its 2 week buffer control equivalent after linear scaling of the former against the latter. The highest peaks therefore would represent the largest reductions in band intensity as a result of the glycation treatment over the time period in question. The vertical green lines above and below the mean represent ±1 SD calculated as previously described (see section 4.2.10). The horizontal coloured line is placed at a value along the ordinate representing the lowest value of "mean +1 SD" along the plot, which is found at amino acid position 101 along the *x*-axis in this case. Any clear area between the mean trace minus 1 SD and this horizontal line is considered a significant change. The positive staining band locations are also identified. The vertical dashed lines indicate the locations of the gap-overlap boundaries, the solid horizontal black line at the top of the band indicator-shading showing the extent of the overlap region.

In the case of galactose (Figure 4.10), significant site-specific changes were seen after 1, 3 and 4 weeks, each time period showing a different pattern of loss of positive staining. During the 1 week incubation, the most significant effects were in the *b1-b2* and *d* bands, with a more modest loss of *a4* (see also Figure 4.6A). After 3 weeks, the most significant loss of positive stain was from the *a1* band with a continuing region of loss into the *e2-e1* bands and a smaller magnitude of loss from the *b1* band. Following the 4 week galactose incubation, intensity reductions were seen in the *a2-a3-a4* and *c1-c2* regions with a more modest loss of *e1*. No significant band-specific losses were seen with this sugar after the 2 week incubation.

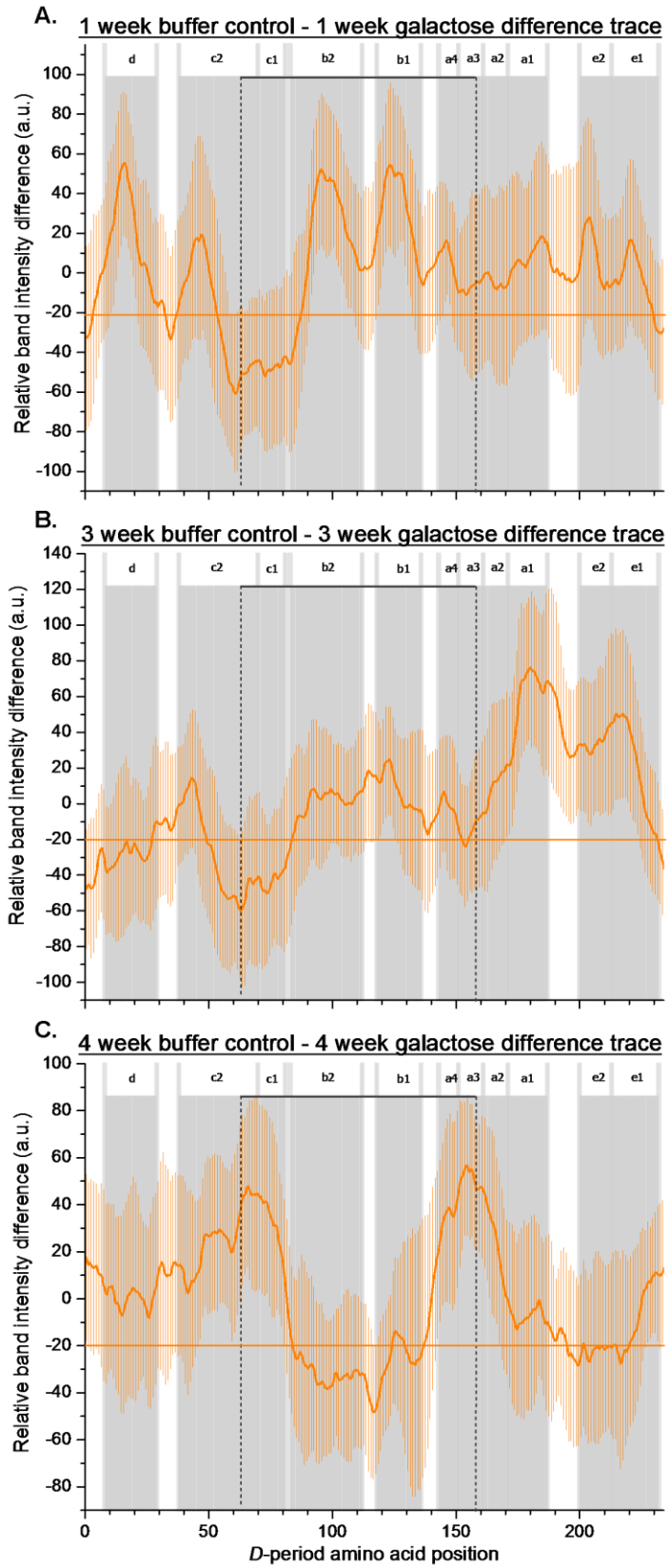


Figure 4.10 Mean TEM positive staining band intensity difference traces showing significant detected changes after galactose treatment of MTT collagen fibrils, i.e. over the **A.** 1 week, **B.** 3 week and **C.** 4 week incubations. All lines and notations are as in Figure **4.9**.

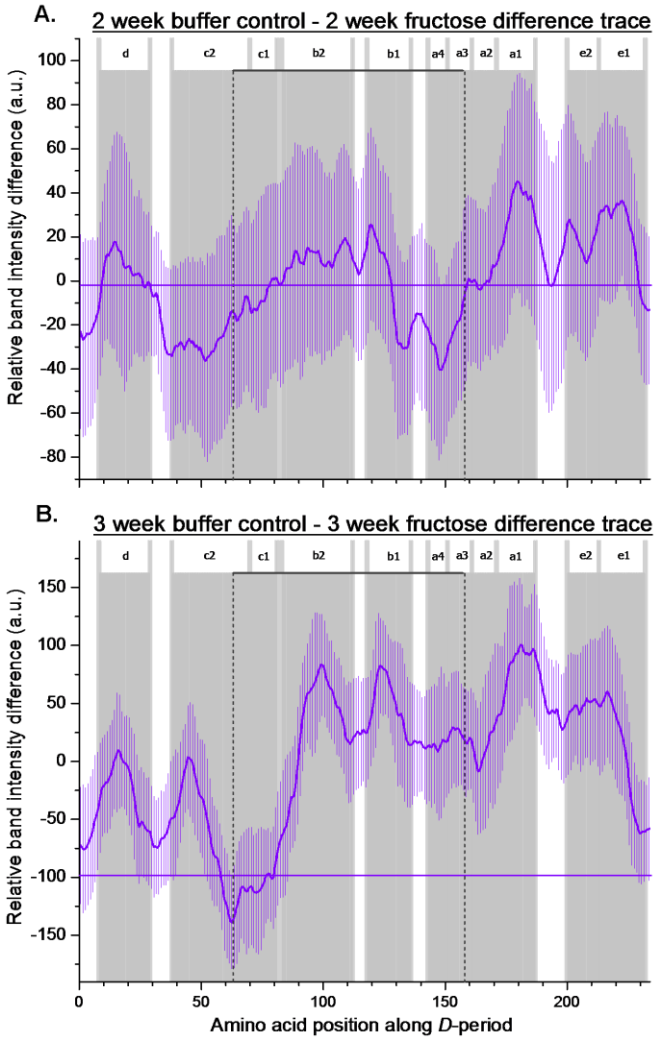


Figure 4.11 Mean TEM positive staining band intensity difference traces showing significant detected changes after fructose treatment of MTT collagen fibrils, i.e. over the **A.** 2 week and **B.** 3 week incubations. All lines and notations are as in Figure **4.9**.

With fructose (Figure 4.11), after the 2 week incubation there was a small loss of intensity from the *e1* band of borderline significance. After 3 weeks, there were widespread reductions in relative band intensity across the *a1→a4*, *b1-b2* and *e1-e2* bands, with more modest reductions in the *c2* and *d* bands (see also Figure 4.6C and Figure 4.7). Indeed here the only visible positive staining band unaffected was *c1*. Neither the 1 week nor 4 week fructose incubations produced any discernable changes.

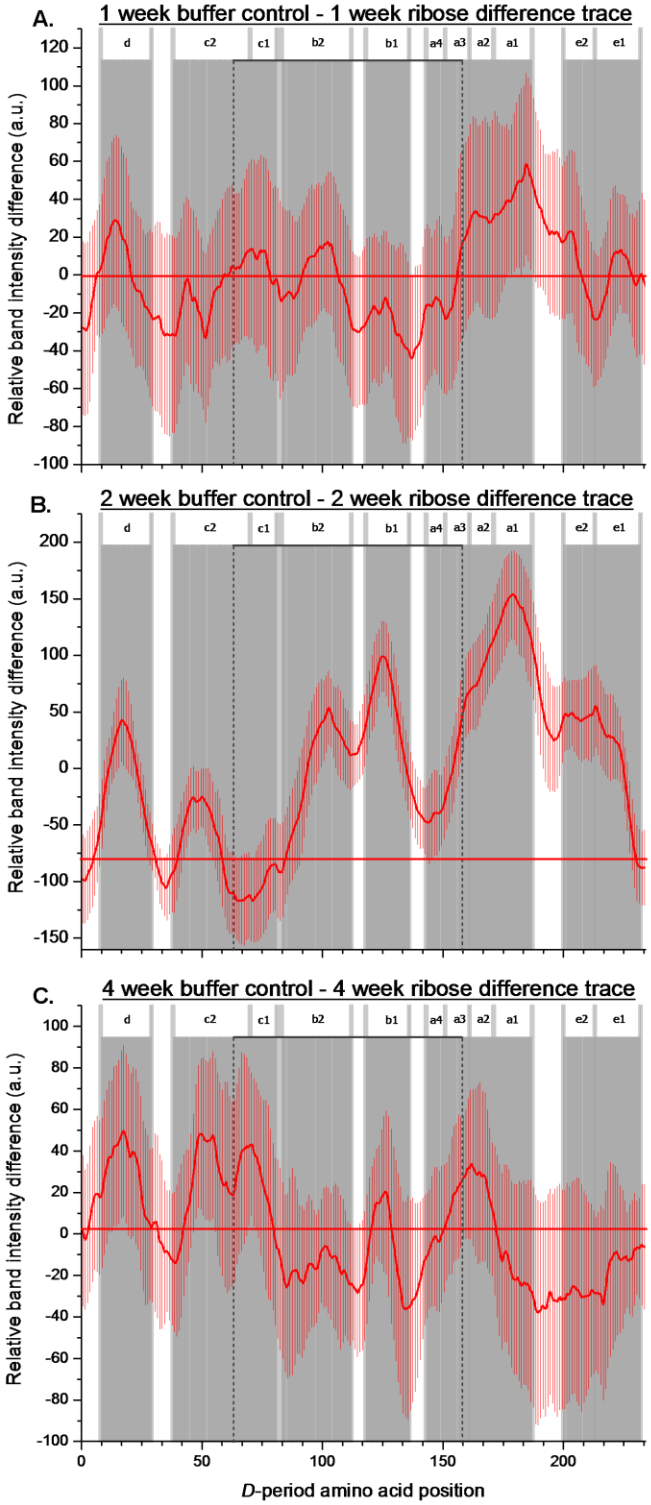


Figure 4.12 Mean TEM positive staining band intensity difference traces showing significant detected changes after ribose treatment of MTT collagen fibrils, they being over the **A.** 1 week, **B.** 2 week and **C.** 4 week incubations. All lines and notations are as in Figure **4.9**.

For ribose (Figure 4.12), after the 1 week incubation a small effect in the *a1* band was measured with no other significant changes apparent. After 2 weeks changes resembling those of fructose after 3 weeks (Figure 4.11B) are seen, the only band remaining unaffected being *c1* (see also Figure 4.6B). After 4 weeks, small changes are evident in the *c1*, *c2* and *d* bands. In the case of the 3 week ribose incubation, no significant location-specific reductions in band staining intensity were measured.

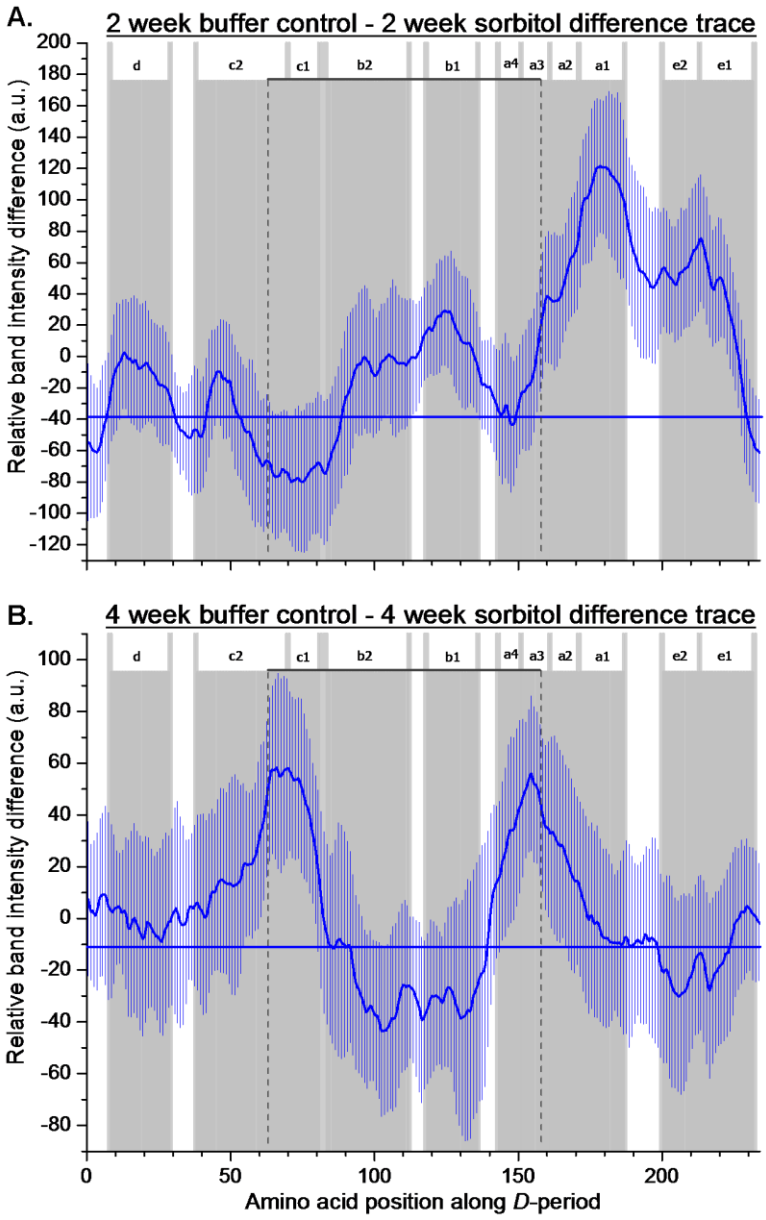


Figure 4.13 Mean TEM positive staining band intensity difference traces showing significant detected changes after sorbitol treatment of MTT collagen fibrils, they being over the **A.** 2 week and **B.** 4 week incubation times. All lines and notations are as in Figure **4.9**.

Although initially included as a non-glycating sugar-alcohol control, significant location-specific band intensity changes along the *D*-period were also measured after 2 and 4 week incubations with sorbitol (Figure 4.13). The changes after 2 weeks resulted in a difference trace similar in shape to the 3 week fructose (Figure 4.11B) and 2 week ribose (Figure 4.12B) traces, with large changes encompassing the *a3*→*a1* region and continuing through to the *e2-e1* bands and a more modest change in the *b1* band. There was also a minute change in the *b2* band, with another small borderline change in the *d* band. The 4 week sorbitol incubation produced a difference trace uniquely similar to the 4 week galactose plot (Figure 4.10C), with the measured changes of note being in the *a1*→*a3* and *c1-c2* bands.

Pearson r values for correlations

	2 week glucose	1 week galactose	3 week galactose	4 week galactose	2 week fructose	3 week fructose	1 week ribose	2 week ribose	4 week ribose	2 week sorbitol	4 week sorbitol
2 week glucose	x	0.232	0.152	0.495	0.095	0.063	0.480	0.148	0.219	0.265	0.555
1 week galactose	0.232	x	0.459	0.559	0.381	**0.781**	0.134	0.655	0.152	0.498	0.643
3 week galactose	0.152	0.459	x	0.542	0.658	**0.836**	0.387	**0.783**	0.657	**0.887**	0.518
4 week galactose	0.495	0.559	0.542	x	0.583	0.595	<0.050	0.446	0.672	0.427	**0.946**
2 week fructose	0.095	0.381	0.658	0.583	x	0.562	0.555	0.649	0.385	0.692	0.486
3 week fructose	0.063	**0.781**	**0.836**	0.595	0.562	x	0.290	**0.879**	0.519	**0.825**	0.622
1 week ribose	0.480	0.134	0.387	<0.050	0.555	0.290	x	0.516	0.055	0.541	0.126
2 week ribose	0.148	0.655	**0.783**	0.446	0.649	**0.879**	0.516	x	0.266	**0.933**	0.463
4 week ribose	0.219	0.152	0.657	0.672	0.385	0.519	0.055	0.266	x	0.404	0.619
2 week sorbitol	0.265	0.498	**0.887**	0.427	0.692	**0.825**	0.541	**0.933**	0.404	x	0.421
4 week sorbitol	0.555	0.643	0.518	**0.946**	0.486	0.622	0.126	0.463	0.619	0.421	x

Table 4.1 Cross-correlation matrix showing the Pearson r values for the correlations between all the glycation incubations difference traces that showed significant changes in the positive staining band intensities along the MTT *D*-period (see Figure **4.9** to Figure **4.13**). All Pearson r values of >0.70 are in bold red.

Table 4.1 shows the cross correlations between all the difference traces deemed to show significant changes. An r-value of 0.70 was arbitrarily chosen as a cut-off point for significance of these comparisons and so eight correlations are highlighted as being the strongest. Broadly, these confirm much of what is evident visually inasmuch as there is a four-way relationship of similarity between the traces' shapes between the 2 week ribose (Figure 4.12B), 3 week fructose (Figure 4.11B), 3 week galactose (Figure 4.10B) and also 2 week sorbitol (Figure 4.13A) incubations. Broadly, these reflect a pattern in which the favoured band intensity changes appear to be concentrated in a continuous line from the *b2* band, through the *b1* band and across the four *a* and two *e*

bands. These are more emphatic and significant in the cases of the 2 week ribose and 3 week fructose, whereas the region of change is discontinuous in the cases of the 3 week galactose and 2 week sorbitol incubations. Nonetheless, the mean trend line holds. All four of these traces also show that the *c1* band and the half of the *c2* band closest to the *c1* band to be the least favoured part of the *D*-period for positive charge removal over these time periods in these respective incubations. The major differences between these difference traces relate mainly to the portion of the *c2* band distal to the *c1* band, plus the *d* band. These regions show significant changes after the 2 week ribose and 3 week fructose incubations. These changes are smaller and of borderline significance after the 2 week sorbitol incubtation and not significant after the 3 week galactose incubations. However, for the galactose experimental data it should be noted that the discrepancies between the 3 week galactose incubation and particularly the 2 week ribose and 3 week fructose data are largely offset by consideration of the 1 week galactose difference data (Figure 4.10A). Addition of the 1 and 3 week galactose difference traces results in a dataset, which shows a better correlation with the 2 week ribose data (Pearson r value = 0.847) and especially with the 3 week fructose data (Pearson r value = 0.948). Therefore site specific band intensity changes that occurred over 2 weeks with ribose and 3 weeks with fructose appear to have taken place in two separate stages in the case of galactose, i.e. over the first week and then, subsequently, in week three. This accounts for the good correlation (Pearson r value = 0.781) between the 1 week galactose and 3 week fructose difference traces, as well as a reasonable correlation between the 1 week galactose difference data and the 2 week ribose data (Pearson r value = 0.655).

The strongest correlation represented in Table 4.1 is between the 4 week galactose (Figure 4.10C) and 4 week sorbitol (Figure 4.13B) difference traces, with a Pearson r value of 0.946. These represent an altogether different trace shape, unseen for any of the other sugar incubations and time periods. These show a loss of positive staining over the *c1* band, a portion of the *c2* band proximal to the *c1* region and the *a4*→*a2* bands. In the case of the 4 week galactose, there is also a small peak in the *e1* band region.

All the significant change means are plotted in Figure 4.14, aligned and on the same axes for comparison. Inspection of this indicates a general rank order of magnitude of region- or site-specific changes of ribose > fructose > sorbitol > galactose > glucose. The one detected significant (borderline) change for a glucose incubation, i.e. after 2 weeks, is barely visible in comparison to the other sugars, being present as a green

speck sitting just above the *x*-axis over the amino acid position 160 tick. It is also clear from this representation of these data that almost the entirety of the ribose and fructose site-specific changes are detectable after just one time period of incubation, i.e. 2 and 3 weeks respectively and so presumably take place predominantly over these time periods. The sorbitol similarly appears to effect most of its region-specific changes over the 2 week incubation period, though an apparently-greater proportion of changes occur with this substance after 4 weeks, mainly in different parts of the positive staining pattern, as compared with the aforementioned ribose and fructose incubations. Conversely, the changes detected for the galactose incubations are relatively uniform in magnitude, with a different pattern of site-specific changes detected after the 1, 3 and 4 week incubation period and collectively, resulting in changes over at least a portion of every observed positive staining band.

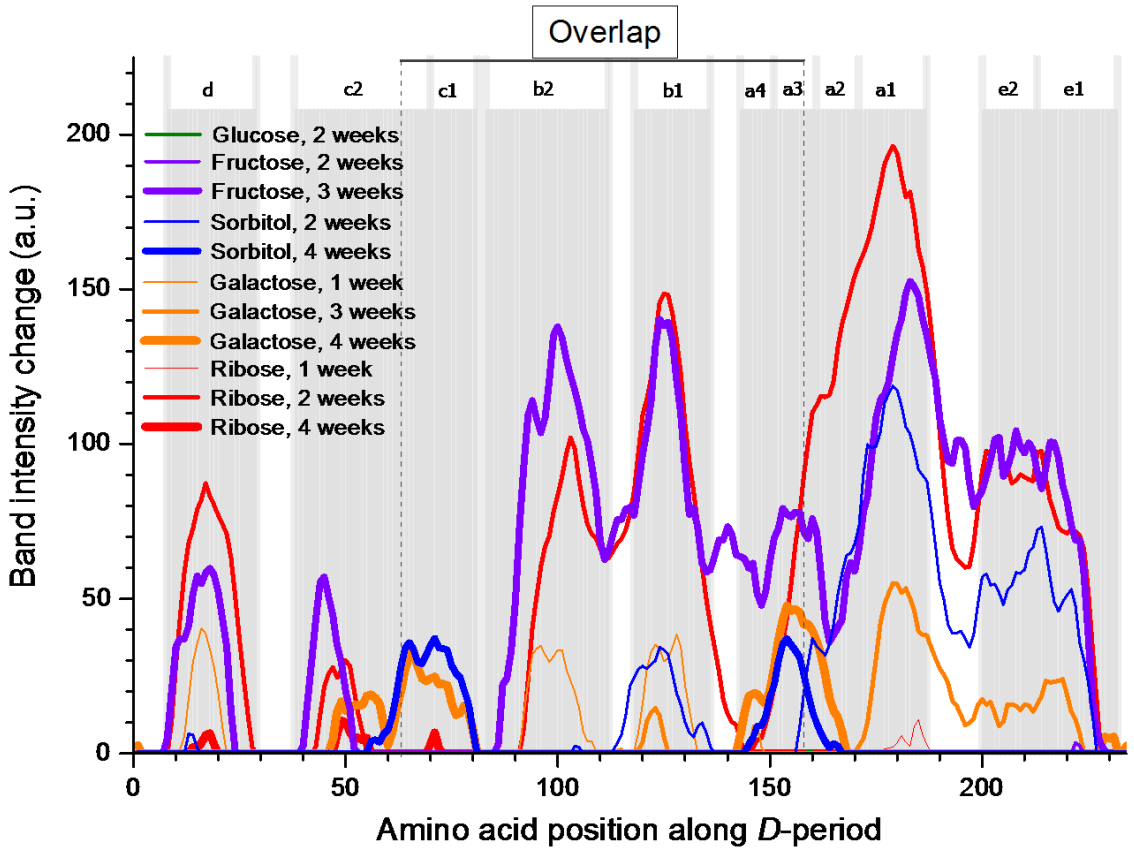


Figure 4.14 All significant changes in positive staining band intensity measured for the glycation treatment groups, including the sorbitol "control" incubations, calculated from the scaled means as described in section 4.2.10. The plots are colour-coded for each sugar, with increasing thickness of progressive traces representing longer incubation times – see key. Treatment groups not included showed no significant changes. All traces are represented on the same ordinate (a.u.) scale. For orientation along the *D*-period, the locations of the positive staining bands and overlap region are shown.

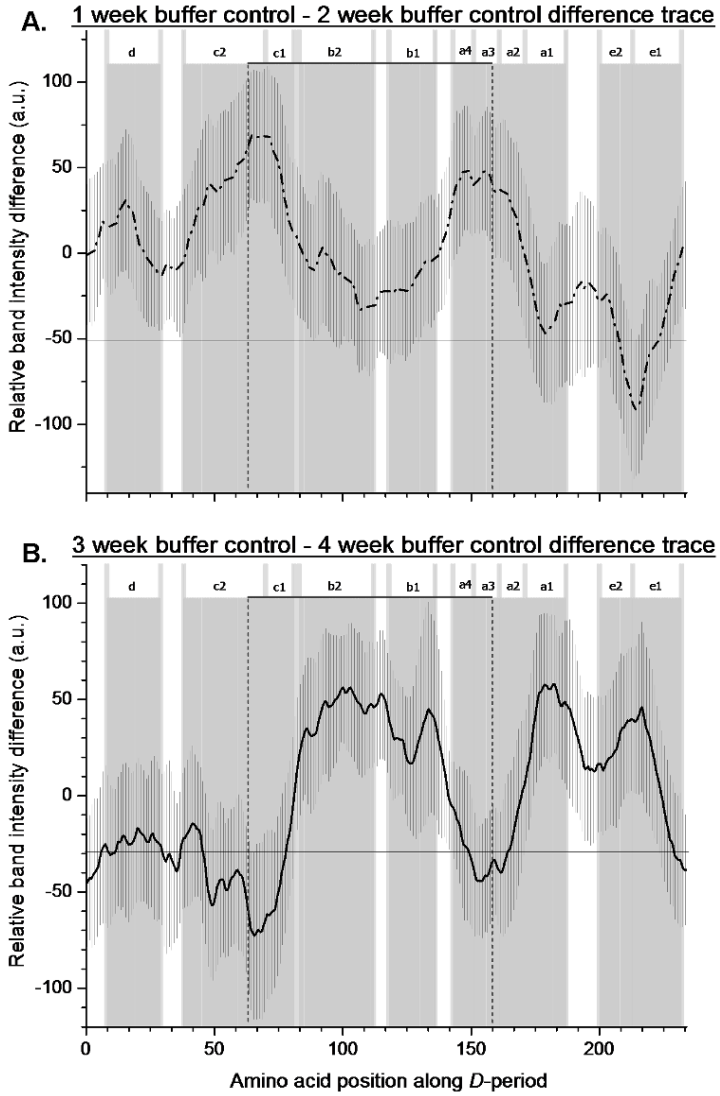


Figure 4.15 Difference traces showing the changes in the positive staining banding pattern of a single *D*-period of an MTT collagen fibril *between* **A.** 1 and 2 weeks and **B.** 3 and 4 weeks of incubation in the glycation buffer. These were calculated by subtracting the mean data for the 2 week buffer control incubation from the 1 week, and the 4 week buffer control incubation from the 3 week, respectively. The vertical lines appended along the mean trace represent ±1 SD. Bands and the overlap region are indicated as in Figure **4.9** to Figure **4.13**. For all datasets used in calculating these traces, n=20.

Changes in the buffer control MTT *D*-period positive staining banding pattern have already been described here and are evident in Figure 4.8A-E. For completeness, consideration of the changes occurring with this control treatment over successive time periods is presented in Figure 4.15. Here, it is important to note that unlike for the glycation difference data presented in Figure 4.9 to Figure 4.13 that all represent

changes over the entire periods, to which they pertain – and corrected for the changes in the buffer control, by virtue of their means of calculation (see section 4.2.10) – these buffer control difference data represent changes from one time period to the next. Significant differences, determined in the same way as described in the caption for Figure 4.9, were only seen to exist in the buffer control MTT traces representing the changes between the 1 week and 2 week, and 3 week and 4 week band intensity traces respectively. The traces representing the differences between 0 weeks (untreated MTT) and 1 week, and 2 week and 3 week showed no significant changes and so are not presented.

4.4 Discussion and conclusions

The experiments and data presented here are methodologically similar to earlier TEM positive staining work published by Hadley et al. (1998; 2001), wherein they showed successfully that positive staining of collagen fibrils on the basis of charge distribution, previously demonstrated and described by other workers (Hodge and Schmitt, 1960; Meek et al., 1979; Tzaphlidou et al., 1982a; Tzaphlidou et al., 1982b; Tzaphlidou and Hardcastle, 1984) could be used to study removal of positive charge during glycation.

As well as seeking to expand the scope of this study beyond that of the previous experiments (Hadley et al., 1998) by use of other sugars over a greater number of time periods to enable variations of the site-specificity of the different sugars and the changes of this over time to be ascertained, other differences in approach were employed here. Most obviously, the tissue source of the fibrils was tendon from mouse tail rather than sclera from human eyes. In addition, as described in Chapter 2 (section 2.2), careful consideration was given to the composition of the buffer in which the glycation incubations would take place, as described previously (Rodda, 2000). Unlike in Hadley et al's (1998) work, it was decided that phosphate buffer would not used because of the possibility of catalysis of glycation at certain sites (Watkins et al., 1987; Gil et al., 2005). MOPS was selected as a reasonable choice instead. It was also considered important here to include buffer control MTT incubations, so that any changes in the positive charge distribution that might result from interaction with the buffer components could be taken into account in assessing glycation differences. The sugars were used at a concentration of 0.2M in line with other collagen glycation work previously published (Tanaka et al., 1988a; Bai et al., 1992; Meli et al., 2003) (see also section 2.4). Therefore an additional osmotic control with a sugar-like substance

present was also included, to redress any effects of colligative property differences between buffer alone and the buffer with an additional 0.2M solute, minus glycating ability. Sorbitol – a sugar alcohol, so with no reactive carbonyl group - was chosen, as it had been used in previous glycation work and not produced any changes detectable by X-ray diffraction (Tanaka et al., 1988a).

4.4.1 Buffer effects

In view of the results presented in section 4.3.3, certain considerations are important before the effects of the sugars themselves can be evaluated. Firstly, with regard to the buffer control MTT incubations, it is clear from the data presented in Figure 4.15A and B that between 1 and 2 weeks, and 3 and 4 weeks, there are separately two distinct sets of changes along the fibril that appear to show preferential modification of the charge distribution over certain regions, by the buffer. Comparing these two traces, it is so obvious that either one can be inverted and fitted with reasonable closeness over the other, that it would seem superfluous to present it here. Predictably, there is a good negative correlation between them with a Pearson's r value of -0.783. Therefore modifications initially detected show preference for a certain pattern of sites – and actually show a striking similarity in shape to the 4 week galactose (Figure 4.10C) and 4 week sorbitol (Figure 4.13B) difference traces (Pearson's r values: 0.778 and 0.790, respectively) – between 1 and 2 weeks, with a complementary pattern of change between 3 and 4 weeks. Simplistically, it looks as though positive charge is being removed at preferred sites between 1 and 2 weeks, and then at the alternative sites over the 3 to 4 week time period. In considering this, the first factor that must be considered is that tissue from one individual mouse was used to generate all the data for the untreated MTT, 1 week and 4 week MTT incubations and, due to the limited amount of tendon available from a single mouse tail, tissue from a second mouse was used for the 2 and 3 week incubations. Therefore, although the tissue came from two 3 month old female CD-1 mice that were killed at the same time and the tissue stored and treated in the same way prior to the experiments, the possibility that the fibrils showed an inter-individual variation which was detected by these experiments cannot be discounted. However, study of the mean positive staining plots for the untreated ("0" week) and 4 week buffer control sample (Figure 4.8, traces A and E), both obtained from MTT from the same individual mouse, shows a clear buffer control-related change over the 4 weeks. The calculated difference trace between these groups showed a significant reduction in positive charge in the *b2* band after 4 week buffer incubation as

compared with the untreated MTT. Therefore it appears that even if there is variation between the MTT fibrils from the two mice, the evidence presented here suggests that there are still buffer-mediated effects. The two major questions here are **1.** by what mechanism is the buffer bringing about these changes and **2.** what effect is this mechanism having on the glycation reactions themselves?

Regarding question **1**, all the buffer components have to be considered. The NaCl component is present naturally and is used in this study at a physiological concentration. Ions are known to interact with tendon collagen (Cameron et al., 2012) and it has been shown that the varying concentrations of NaCl can affect collagen fibrillar order and therefore function (transparency) in cornea (Kostyuk et al., 2002; Regini et al., 2004). However here, NaCl is not in itself appreciably altered in concentration as compared with the the *in vivo* environment from which the tendon was excised. Therefore although they are likely to be involved in any electrostatic mechanism, it is suggested that Na^+ and Cl^- ions are not in themselves, causative agents in the buffer control changes detected.

The buffering agent, MOPS, is a zwitterionic Good's buffer (Good et al., 1966). It is generally regarded as suitable for a wide range of biochemical applications including cell culture applications, and used as such (www.sigmaaldrich.com, 2013a). MOPS does not appear to be particularly susceptible to causing various interferences suspected and investigated in biochemical research (Yu et al., 1997; Zhao and Chasteen, 2006) unlike other members of this class of buffers. The only mechanism by which it can be foreseen that MOPS might bring about the observed changes is electrostatic. At the pH used (pH 7.4), the sulphonate group would be expected to be ionised, with a proportion of the buffering tertiary amine groups of the molecule protonated. It is proposed here therefore, that the sulphonate portion of the molecule becomes electrostatically attracted and attached to protonated lysyl amine groups, thereby shielding them from staining by PTA. This might call into question the adequacy of the washing step, although three washes in 0.9% saline solution plus 1 hour in the same solution during grid deposition might generally be regarded as sufficient.

Other components of the glycation buffer are present in smaller quantities and possible mechanisms for their bringing about the buffer control effects are less obvious. DTPA is known as a cross-linking agent, but only in its anhydride form which was not employed here, as well as being an efficient chelating agent (Buffa et al., 2011), which was its purpose of inclusion in the buffer mixture (See Chapter 2, section 2.2).

Similarly, sodium azide is commonly used in comparatively small quantities as a preservative to prevent bacterial growth in research (Watkins et al., 1987; Reiser et al., 1992b; Reddy, 2004) and generally in laboratory reagents. The PIC was specifically designed for use with mammalian cell and tissue extract and is used as specified by the manufacturer (www.sigmaaldrich.com, 2013b), in small quantities.

Regarding the effect of the buffer on glycation reactions (question **2**, above) if any, this is difficult to ascertain from these data. Assuming the MOPS molecule itself is the causative factor for the observed changes, as is considered most likely, much would depend on the affinity of binding between this and the positively charged amino acids along the fibril plus the effectiveness of any resultant shielding caused, in preventing glycation.

4.4.2 Sorbitol effects

Sorbitol was included as a second control, as previously by Tanaka et al. (1988a) and therefore expected to be a non-glycating sugar analogue. In the event, the changes measured were significant and appeared to correlate well with significant changes detected with other sugars (Table 4.1, Figure 4.14). The possible reasons for the changes seen after sorbitol incubation are: **1.** the sorbitol is being oxidised into a glycating sugar during the glycation incubation; **2.** the sorbitol is binding to the collagen molecules and remaining there during the washing step, thereby selectively reducing the effectiveness of the positive PTA staining procedure and so the resultant banding pattern; or **3.** the sorbitol is having an osmotic effect, thereby affecting the positive staining pattern along the fibril.

Considering the first possibility, sorbitol is widely used in the food industry and known to be chemically stable. The presence of sorbitol in tissues is linked to the so-called “polyol pathway”, in which the sugar alcohol is a metabolic intermediary between glucose and fructose. It has been shown in rat lens that sorbitol is converted to fructose by this pathway, and the amount present is correlated to both fructose levels and the increased degree of fructose-mediated glycation (“fructation”) of the lens crystallin proteins that occurs in diabetes (Kawasaki et al., 1998). However, a hypothesis that an enzymic *in situ* conversion of sorbitol to fructose is occurring here would have to be discounted due to the fact that the distribution of the enzymes involved in this pathway is restricted to certain tissues, which do not include tendon (Gabbay, 1975). An alternative hypothesis, in which chemically-stable sorbitol is spontaneously oxidised in

the reaction solution does not seem likely, as a mechanism or cause for this cannot be envisaged. Any metal ions that might potentiate such a reaction would have been removed by DTPA chelation. Even if by some means the sorbitol were converted to another substance capable of glycating the collagen, it would be unlikely to be a rapid process. If it *were* enzymic and being converted to fructose, it should be noted that the changes seen with sorbitol are more rapidly evident, i.e. after just 2 weeks (Figure 4.14) than with fructose, which only produced location-specific changes of anything like a comparable magnitude after 3 weeks. More to the point, purely within the bounds of this entire study, chemical conversion of sorbitol to a more reactive sugar would also be inconsistent with the AGE results presented in Chapter 3, section 3.3.1 (see Figure 3.1), where little evidence of formation of fluorescent products of glycation even over the full 4 week incubation period were seen for the sorbitol "control". Actual glycation by either sorbitol or some derivative is therefore discounted.

Possibility **2** would involve the attachment of sorbitol to the collagen fibril preferentially at certain sites, for it to result in the specific changes detected. All the sugars used in this study are "polyolic" and therefore likely to share a preference for some binding sites along the fibril, principally through hydrogen bonding. It has been reported that sorbitol binds especially well to collagen with serine residues in the *X* position of the Gly-*X*-*Y* collagen triplet (Usha et al., 2006). It is in any case, reasonable to presume that both this and the glycating sugars used would generally bind better, via hydrogen bonding, to more hydrophilic regions along the fibril. A graphical comparison of the 2 week sorbitol difference trace along one MTT *D*-period and smoothed traces representing the distribution of *X*-serines and hydrophilicity are presented in Figure 4.16. The similarities are visually evident. The correlation between the 2 week sorbitol difference trace and the smoothed *X*-serine and hydrophilicity traces, give Pearson's r values of 0.522 and 0.657, respectively. This suggests that general hydrophilicity is a more important factor than *X*-serine distribution in the changed positive staining pattern. Although Usha et al. (2006) reported optimal thermodynamics and predicted a conformational change of sorbitol from its normal linear to a folded form on binding to collagen, it is unclear whether this would confer any degree of resistance to the washing step. Following incubation of bovine flexor tendon in unbuffered 0.2M glucose solution followed by a drying step, Cameron et al. (2012) suggested that they had found evidence that subsequent washing of the tissue using deionised water left residual glucose in the tendon that could not be removed by the washing. This may give an insight into what is occurring here with sorbitol.

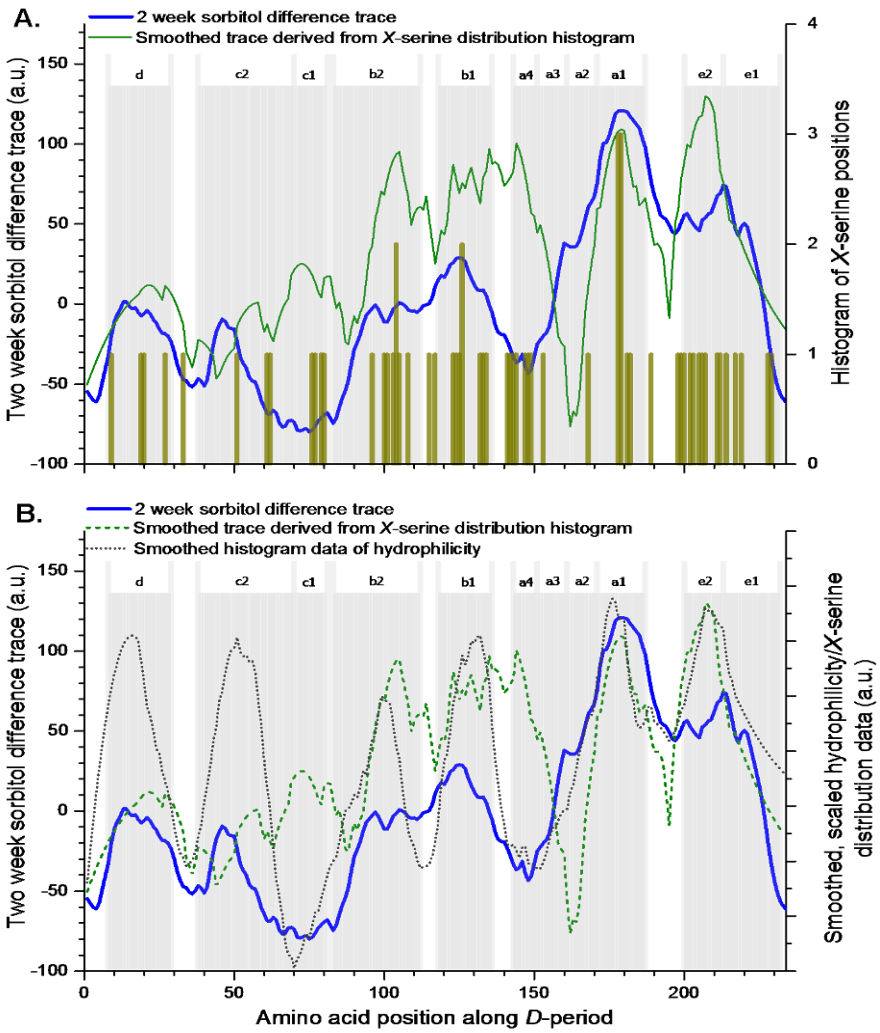


Figure 4.16 **A.** Comparison between 2 week sorbitol difference trace along one MTT collagen fibril *D*-period (as from Figure **4.13**A) and a histogram of *X*-serine distribution plus its smoothed equivalent; and **B.** 2 week sorbitol difference trace plus smoothed X-serine distribution trace from (**A**), with smoothed trace of hydophilicity along the *D*-period. The latter was calculated by producing a histogram of hydrophobic amino acid distribution (taken to be valine, leucine, isoleucine, phenylalanine, tyrosine, methionine (Nozaki and Tanford, 1971) and proline), subtracting it from its highest value (i.e. 7) for each point thereby linearly inverting it, and then smoothing it to give the trace (black dotted line) presented. All amino acid distributions were based on the simulated fibril (Chapter 2) and smoothing was performed in the same way as for the histogram of basic amino acids presented in Figure **4.5**, i.e. using a Savitsky-Golay smoothing algorithm, 2nd order, with a 34 amino acid window size. Positive staining bands are overlaid (gray areas) and labelled.

The **3**rd possibility, i.e. the sorbitol is having an osmotic effect, might be supported by previous work (Kuznetsova, Chi and Leikin, 1998), which found that sorbitol tended to be excluded from collagen fibrils, thereby producing an osmotic differential between the inside and outside. However the means by which this would affect the relative intensity of the staining pattern in a way that is independent of the gap/overlap regions is

unclear. Further, Kuznetsova et al. (1998) reported that glucose and fructose were not excluded and that the difference was likely to be as a result of the relative conformational flexibility of sorbitol, which is a linear (chemically 100% acyclic) molecule. On the other hand, glucose - and presumably the other sugars – are in equilibrium with ≥99.3% in the cyclic form (Hayward and Angyal, 1977), therefore of a more homogeneous form in solution. Interactions between these and the surface of the collagen fibril/molecules are considerably less likely therefore, to favour certain conformations of the sugars and bring about differentials that will affect entry into the fibril as rapidly. Nonetheless, with regard to this point, it can be seen that the good correlations between the 2 week sorbitol, 2 week ribose, 3 week fructose and 3 week galactose difference traces would seem to indicate that the cause of the site specific sorbitol positive staining intensity changes after 2 weeks is not primarily osmotic. Therefore the changes in relative band intensity seen for sorbitol are likely to be mediated by the polyol binding followed by resistance to being removed from the collagen during the post-incubation washing stage. The PTA binding to these regions is presumably then directly blocked by the bound residual sorbitol.

4.4.3 Glycation and binding by the glycating sugars

With respect to the glycating sugars, due attention should first be paid to glucose. Ironically, as its low rate of participation in glycation reactions is reflected in these results (see Figure 4.9 and Figure 4.14), it showed its viability as a negative control under the conditions chosen for these experiments, as was originally anticipated for sorbitol. The complete lack of site specific changes of any convincing significance along the *D*-period for all four time periods of incubation with glucose offered reassurance that the difference traces generated for the other sugars were not showing artefactual changes that were bound to occur merely as a result of the presence of any sugar or sugar-like substance in the solution. Along with the relative lack of AGEs detected in Chapter 3, section 3.3.1, it demonstrated why glucose is often either not used in glycation research *in vitro* (Bai et al., 1992; Hadley et al., 1998; Reddy et al., 2002; Pageon and Asselineau, 2005) or used in concert with other sugars (Tanaka et al., 1988a; Fujimori, 1989) to ascertain the nature of changes, despite its physiological relevance.

Ribose and fructose both appeared to exert almost all of their detectable location-specific band intensity effects over 2 and 3 weeks, respectively. In both cases, these changes were obvious from visual examination of the electron micrographs (Figure

4.6). With respect to fructose, corresponding changes in the fibrillar morphology were also noted after 3 weeks of incubation, inasmuch as most of the fibrils appeared expanded and more diffuse (section 4.3.3, Figure 4.7). It is unclear whether this effect is a genuine consequence of the fructose treatment or whether it occurred during the glycation incubation or afterwards during mechanical dispersion of the collagen prior to grid deposition. It is interesting to note that the effect was not also seen after the 4 week fructose incubation. As the positive staining pattern was preserved and useable, appearing to give a comparable degree of precision to that seen with the other treatment groups (section 4.3.3, Figure 4.8P compared with Figure 4.8A-O, Q-Y), this grid was not repeated. It cannot yet be ruled out therefore, that this 3 week fructose change is an experimental artefact. In the case of ribose, it was noted that after the 2, 3 and 4 week incubations, the fibrils became straighter and less tangled on the grid, with less adherent non-collagenous matter. However, no loss of integrity across fibrils was evident.

It is reasonable to presume that, assuming the 3 week fructose morphology change is genuine, both these observations relate to glycation effects. The ribose changes most probably relate to blocking of some of the many binding sites along the fibril by the glycation process, as considered in Chapter 1, section 1.3.3 (See Di Lullo et al. (2002), Sweeney et al. (2008)), resulting in less non-collagenous matter remaining attached to the fibrils observed. The fructose changes by comparison, if genuine, would be interesting as they do not appear to persist after 4 weeks and are unmatched by observations seen with the ribose-treated fibrils. If the changes relate to blocking of binding sites, they are presumably a different pattern of such, perhaps more related to the interactions of collagen molecules to others in order to maintain fibrillar integrity. In addition, this could be the result of a post-glycation loss of stability of the collagenous structure itself (see section 1.3.3), as suggested by the work of Reigle et al. (2008). It could be hypothesised that between 3 and 4 weeks, further AGE cross-link formation might act to join neighbouring molecules together to a point where the process reduces the fibrils' propensity to fall apart during the mechanical dispersion procedure. Nontheless, further experiments are needed to confirm the 3 week fructose observation.

4.4.3.1 Distribution of TEM positive staining changes after ribose and fructose MTT incubations

Overall, with regard to the 2 week ribose and 3 week fructose difference data, these patterns of change correlate well with one another as well as with the changes seen for the 2 week (non-glycating) sorbitol incubation (Table 4.1). Study of Figure 4.14 suggests that between the *a2* and *e1* bands, these changes closely follow the pattern demonstrated by sorbitol over 2 weeks and in view of the discussion above, are influenced by the distribution of hydrophilic residues and possibly to a lesser extent, *X*-serine residues. In other regions of the *D*-period, specifically the *b1* and *b2*, *c2* and *d* bands, the changes seen with both sugars over these time periods were proportionally higher than sorbitol, and so other factors seem to have come into play. Assuming that these are more specifically glycation-related, two factors that have previously been suggested are proximity of additional lysines (Watkins et al., 1987; Johansen et al., 2006; Kitamura et al., 2011) or acidic residues (aspartate and glutamate) (Shapiro et al., 1980; Johansen et al., 2006) proximal to preferentially glycated lysine residues. In the case of nearby lysines, there is a danger of postulating a self-fulfilling prophecy inasmuch as the greater the concentration of lysines along a primary sequence or in a region, the more likely one is to find at least one of them to be preferentially glycated. On the basis of this, a concentration effect may exist. It can also be foreseen that further removal of basic amino acids is more likely due to formation of AGEs, where there are more lysines (or arginines) present, concentrated close to an initial glycation event. Therefore demonstration of catalysis or augmentation of glycation via proximity of lysines – the very residues that are being glycated – is potentially problematic.

The proximity of potentially glycation-catalysing acidic amino acid side-chains to preferentially glycated lysine residues has also been considered previously, in both non-collagenous protein (Shapiro et al., 1980) as well as type I collagen itself (Reiser et al., 1992b). Of the four preferentially-glycated lysines that Reiser et al. (1992b) identified, all have an acidic amino acid three or less residues away in the same molecule. This does not exclude the possibility of a carboxyl group close to a lysine by lateral proximity within the *D*-period rather than being close in the primary sequence, exerting a catalytic effect, as has been shown previously in haemoglobin (Shapiro et al., 1980).

A.

	Pearson's r values							
	Predicted basic amino acids							
	Basic amino acids including histidines		Lysines and arginines		Lysines		Arginines	
	1	**2**	**1**	**2**	**1**	**2**	**1**	**2**
2 week sorbitol	0.630	0.051	**0.701**	0.123	0.554	**0.472**	0.574	-0.477
2 week ribose	0.747	0.277	**0.812**	0.332	0.643	**0.674**	0.664	-0.422
3 week fructose	0.554	0.122	**0.651**	0.193	0.458	**0.220**	0.585	-0.025

B.

	Pearson's r values					
	Predicted acidic amino acids					
	All acidic amino acids		Aspartates		Glutamates	
	1	**2**	**1**	**2**	**1**	**2**
2 week sorbitol	**0.763**	**0.656**	0.647	0.416	0.676	0.602
2 week ribose	**0.853**	**0.706**	0.652	0.443	0.792	0.655
3 week fructose	**0.713**	0.441	0.465	0.253	0.704	**0.446**

C.

	Pearson's r values			
	Hydrophilicity		X-serines	
	1	**2**	**1**	**2**
2 week sorbitol	**0.651**	**0.779**	0.570	0.342
2 week ribose	0.642	**0.634**	**0.678**	0.224
3 week fructose	0.530	0.518	**0.647**	**0.589**

Table 4.2 Tables of correlations (Pearson's r values) between the positive staining difference traces for 2 week sorbitol, 2 week ribose and 3 week fructose incubation treatments and distribution of different types of amino acid predicted from the simulated fibril *D*-period **1.** between amino acid positions 1 and 139 (bands *d*→*b1*, N to C) and **2.** between amino acid positions 140 and 234 (bands a4→e1, N to C). These regions are identified as "**1**" and "**2**" for each amino acid/type of amino acid in the tables. Table **A.** shows the correlations with the smoothed histogram trace for basic amino acids, Table **B.** shows the correlations for the smoothed histogram traces for the acidic amino acids and Table **C.** shows the correlations with hydrophilicity and *X*-serine distribution. In each table, the best correlation for each treatment group for each *D*-period region is shown in bold, with the best correlation overall for each treatment group per region being highlighted in red in the appropriate table. Histograms of predicted positions for each amino acid or amino acid type were smoothed using a 2nd order Savitsky-Golay smoothing function with a 34 amino acid window size (see section 4.3.3 and Figure **4.5**.).

In Table 4.2 are presented data correlating the MTT *D*-period positive staining difference traces from the 2 week ribose and 3 week fructose treatment groups, plus 2 week sorbitol for comparison, with smoothed histogram data showing the predicted distribution of basic, acidic and hydrophilic amino acids. An *X*-serine group was also included. Although the two regions considered – i.e. **1.** the *d* band to the *b1* band

(encompassing the N-terminus), left to right, and **2.** the *a4* to the *e1* band (encompassing the C-terminus), left to right – were somewhat arbitrarily defined, both contained portions of *D*-period, in which significant changes had been detected for the 2 week ribose and 3 week fructose glycation treatments. A greater proportion of region **2** was so affected than region **1** and so it may be anticipated that any differences in the correlations seen between the two regions might at least in part, reflect their differing overall predispositions to being glycated as opposed to just binding the monosaccharides.

Between the two regions, the patterns of correlation are different. For region **1**, the strongest correlation for all three groups was with the acidic amino acids, collectively. This being the case for the sorbitol also, might suggest that this implies an involvement of aspartate and glutamate collectively in binding of sugars to the protein in addition to any catalytic activity. Similarly, the 2 week ribose group's best correlation for region **2** is also with the acidic amino acids, with the 2 week sorbitol group's being moderately good and 3 week fructose's being considerably weaker. Sorbitol also showed good correlations with hydrophilicity, especially for region **2**. Whilst for region **1** fructose showed a reasonably good correlation with the acidic amino acids (glutamate significantly more than aspartate), it showed no such good correlations within region **2**, with its best overall correlation here – uniquely - being with *X*-serine distribution. It was also evident that in region **1**, with regard to the distribution of basic amino acids, that the localisation of positive staining changes showed its strongest correlation for all three treatment groups, with lysine and arginine collectively, and actually arginine distribution appears to show a slightly stronger relationship with the banding changes than lysine. This is different in region **2**, where the arginine distribution does not appear to be positively correlated to distribution of the changes, and where lysine shows the best relationship, even if not strong in itself.

This analysis does appear to show consistency within region **1** with regard to what sorts of amino acids influence changes in relative positive staining intensity along the *D*-period with slightly less such consistency in region **2**. This disparity does appear more evident with the 3 week fructose group though, with the 2 week sorbitol treatment being relatively closely aligned with the 2 week ribose. The overall resemblance between these three groups is contrary to what might be expected, as both ribose and fructose would be expected to be fast-glycating sugars – with ribose being expected to react approximately 16x faster than glucose, and fructose, 7.5x faster (Bunn and Higgins, 1981) – with sorbitol exerting its changes without forming covalent

attachments to amino acids or AGE formation. However again, the correlations between the difference traces of these three groups were good (section 4.3.3 and Table 4.1) and it seems that the overriding similarities are driven by preferred binding sites for the sugars and sorbitol rather than glycation reactivity. Some of the differences between these are likely to be due to glycation reactions and AGE formation, but as the kinetic differences between the ribose's and fructose's glycating activities are not precisely known and therefore precise equivalent times between them are difficult to ascertain, exact comparisons in this regard and in comparison with sorbitol would be difficult to perform.

4.4.3.2 Positive staining changes after MTT incubations in ribose and fructose compared with galactose

It was also considered in section 4.3.3 that addition of the TEM positive staining difference traces for the 1 and 3 week galactose incubations produced a trace which showed comparatively good correlations with the 2 week ribose and 3 week fructose. Of course, the correlation of this derived trace with the 2 week sorbitol was also quite strong (Pearson's r value = 0.829). Galactose would only be expected to glycate about 4.5x as fast as glucose (Bunn and Higgins, 1981) so significantly more slowly than ribose and fructose. Therefore these two separate galactose groups would seem to represent a slowed down version of any location-specific changes which would be seen with the other two sugars. It is interesting in view of this that both ribose and fructose appeared to produce only non- or borderline-specific changes in the week or weeks leading up to their 2 and 3 week difference traces, respectively, whereas galactose after 1 week begins to show changes sufficiently substantial that they were visually obvious from the original electron micrographs (see section 4.3.3 and Figure 4.6). This does not rule out glycation having taken place – both the chromatograms and pentosidine results from Chapter 3 (section 3.3, Figure 3.4, Figure 3.5 and Table 3.1) confirm that AGE formation was detectable from 1 week's incubation onwards for both ribose and fructose, and more so than for galactose – but show that in the case of ribose and fructose, any changes were more general and less site specific, therefore not detected by this method. This plus these two sugars' relatively high reactivity might suggest that they were reacting quickly at first along the *D*-period and only when a certain depletion of basic amino acids (lysines) had occurred, were binding and possibly, glycating at more specific locations. It could be hypothesised that binding of sugars along the fibril would be a determinant of site of reaction, depending on some

balance between reactivity and ability of the sugar to adopt a conformation which more predisposes it to binding at specific (ergo detectable) locations along the fibril/*D*-period.

4.4.3.3 Distribution of the galactose MTT incubation TEM positive staining changes

With regard to this, an analysis of correlations of detected site-specific changes for galactose with the distributions of various amino acids or types of amino acid is presented in Table 4.3. General interpretation of these correlations within bands has to be exercised with caution, due to their relatively small size when compared with the two regions. For example the *a4* band as defined here is only eight amino acids long and therefore particularly susceptible to influence of amino acids from neighbouring bands via the smoothing algorithm used to calculate their distribution. It is also devoid of lysine and aspartate residues, and so a negative correlation might be expected if attempting to correlate any significant difference peaks to these amino acids.

	D-period band/region	Basic amino acids including histidines	Lysines and arginines	Lysines	Arginines	All acidic amino acids	Aspartates	Glutamates	Hydrophilicity	X-serines
1 week galactose	1	0.619	0.692	0.610	0.506	0.745	0.520	0.717	0.677	0.458
	2	-0.040	0.079	-0.014	0.038	0.168	0.094	0.173	0.231	0.541
	d	0.906	0.915	0.875	0.615	0.805	0.635	0.842	0.936	0.534
	b2	X	0.931	0.708	0.865	0.908	0.816	0.898	0.789	0.554
	b1	0.883	0.884	0.855	0.696	0.882	0.893	0.857	0.526	0.029
	a4	X	X	X	0.007	X	X	-0.024	0.354	0.867
3 week galactose	1	0.225	0.322	0.110	0.416	0.471	0.349	0.443	0.231	0.474
	2	-0.207	-0.126	0.093	-0.346	0.392	0.208	0.422	0.746	0.514
	b1	0.711	0.709	0.732	0.337	0.529	0.437	0.595	-0.231	-0.541
	a1	X	-0.403	-0.675	0.812	-0.266	-0.548	0.403	-0.118	0.462
	e2	X	-0.063	-0.537	0.539	-0.096	0.009	-0.301	0.025	-0.129
	e1	0.955	0.934	0.963	0.908	0.969	0.974	-0.641	0.863	0.925
4 week galactose	1	-0.388	-0.481	-0.307	-0.462	-0.587	-0.228	-0.658	-0.273	-0.609
	2	0.632	0.561	0.259	0.510	-0.148	-0.072	-0.169	-0.791	-0.504
	c2	0.483	0.395	0.149	0.672	0.085	-0.098	0.182	-0.336	0.755
	c1	0.909	0.879	-0.292	0.966	-0.859	0.535	-0.899	-0.963	0.473
	a4	X	X	X	0.927	X	X	0.960	-0.832	-0.270
	a3	0.722	0.957	-0.411	0.845	-0.487	-0.497	-0.473	-0.546	0.655
	a2	X	0.975	-0.470	0.940	0.849	0.353	0.959	-0.969	-0.967
	e1	-0.936	-0.920	-0.942	-0.891	-0.950	-0.954	0.620	-0.842	-0.908

Table 4.3 Correlations between the difference traces for the three treatment groups of galactose, for which significant changes were measured - i.e. 1 week, 3 week and 4 week galactose treatments (see Figure **4.10**) – and predicted distributions of different amino acids/types of amino acids. The correlations are shown for each treatment group, with the same classes of amino acids that were considered in Table **4.2** within regions **1** and **2** as before (highlighted in gray in the table) and for each band for which any significant difference was identified (highlighted in orange in the table). An “X” in a position denotes where a particular amino acid is not present in the identified region or a given group does not apply as a result.

Study of this table reveals that as one moves from 1 week galactose incubation to 3 weeks and to 4 weeks (also included), increasing the treatment time period changes the pattern of correlations that exist. After 1 week's galactose incubation, although there is some variation between bands, good correlations exist between the difference traces and the distributions of the basic and acidic amino acids, with general preferences for lysine and glutamate. The relationships with hydrophilicity varied from strong to weak, depending on the band and *X*-serine dependence varied between no correlation (*b1* band) and strong (*a4* band). In the case of the 3 week galactose incubation group, there are fewer positive correlations – indeed, some apparently significant negative correlations – with a greater degree of inter-band variation in preference for changes associated with lysine or arginine, or glutamate or aspartate. Only the *e1* band showed any strong correlation with hydrophilicity and *X*-serine distribution.

Although not actually considered above, the 4 week galactose treatment group shows further change. There are no longer any notable positive correlations between difference traces within given bands and lysine distribution, with all but one band (*e1*) showing a moderate to strong positive correlation between the experimental traces and arginine. There is great variability in the correlation of changes in this group with the acidic amino acids. There are no strong positive correlations with aspartate, with the only strong positive correlations in the group being with two bands (*a4* and *a2*) for glutamate distribution. Perhaps the most significant change in this group is the complete loss of any apparent association between the difference trace for the 4 week galactose incubation and hydrophilicity. Indeed, this correlation is strongly negative for four of the bands (*c1, a4, a2, e1*), which might imply depletion of the hydrophilic regions. The other changes might be indicative of the same phenomenon, though in the case of the reduced association with lysine, the progressive removal of these residues by the glycation process are likely to be a factor.

Whether an indicator of binding location or glycation itself, it has been established that site- or region-specific changes do occur and that there are consistencies of these between different sugars. As it has been broadly ascertained that a window size of 34 amino acids with a 2nd order Savitsky-Golay function brings about a smoothing of the basic amino acid distribution histogram that most closely resembles the experimental mean native MTT collagen trace for 1 *D*-period, precise identification of the amino acids whose charge removal gives rise to changes in the positive staining difference traces following sugar treatment is difficult. Hadley et al. (1998) took the approach of

removing a number of amino acids from the region they found to be most affected, until the correlation improved between their original simulated collagen trace and their fructose-treatment experimental data. This still involves a degree of speculation and guess work. With the galactose data here, where there is a significant and well-defined difference peak which concurs with that of another sugar and the correlations in Table 4.3 point to specific co-distributed amino acids (or amino acid types) which might give clues to specific locations, narrowing down the affected loci might be possible even with the limited resolution used.

4.4.3.4 Preferred glycation sites

The difference peak in the *d* band in the 1 week galactose difference trace corresponds closely with those seen in the 2 week ribose and 3 week fructose traces, and with a small sorbitol peak of borderline significance (Figure 4.14). These three glycating sugar peaks point to an amino acid in almost exactly the same position. The highest point of the peak for the galactose treatment is position 16 along the simulated *D*-period, for fructose, position 18 and for ribose, position 17. From Table 4.3, there is a stronger association between the experimental difference trace with lysine than arginine and with glutamate than aspartate. The proposed, favoured amino acids for glycation-related positive charge removal here are lysine triplets at positions 416 and 884 in the α1 and α2 chains, with a further lone lysine at position 651 in the α2 chain. These are shown in a representation of the simulated fibril *d* band primary sequence structure (Figure 4.17).

All of these lysines are in perfect mutual alignment across the fibril in addition to being in alignment with, and potential-close proximity to, the three glutamates in position 650. Therefore they seem to be strong contenders for preferentially glycated residues. However the triplet of lysines at position 648 are each only two residues away from the three glutamates at position 650 and in reasonable alignment with the two α1-chain glutamates (and one α2-chain aspartate) at position 413, and the lone lysine along α2-chain at position 654 is proximal to two α1-chain aspartates and aligned to acidic amino acids in all the other three alignments of collagen molecule present – including an α1- and α2-chain glutamate triplet at position 419 – therefore these have to be considered possible candidates for preferential glycation by galactose over this time period.

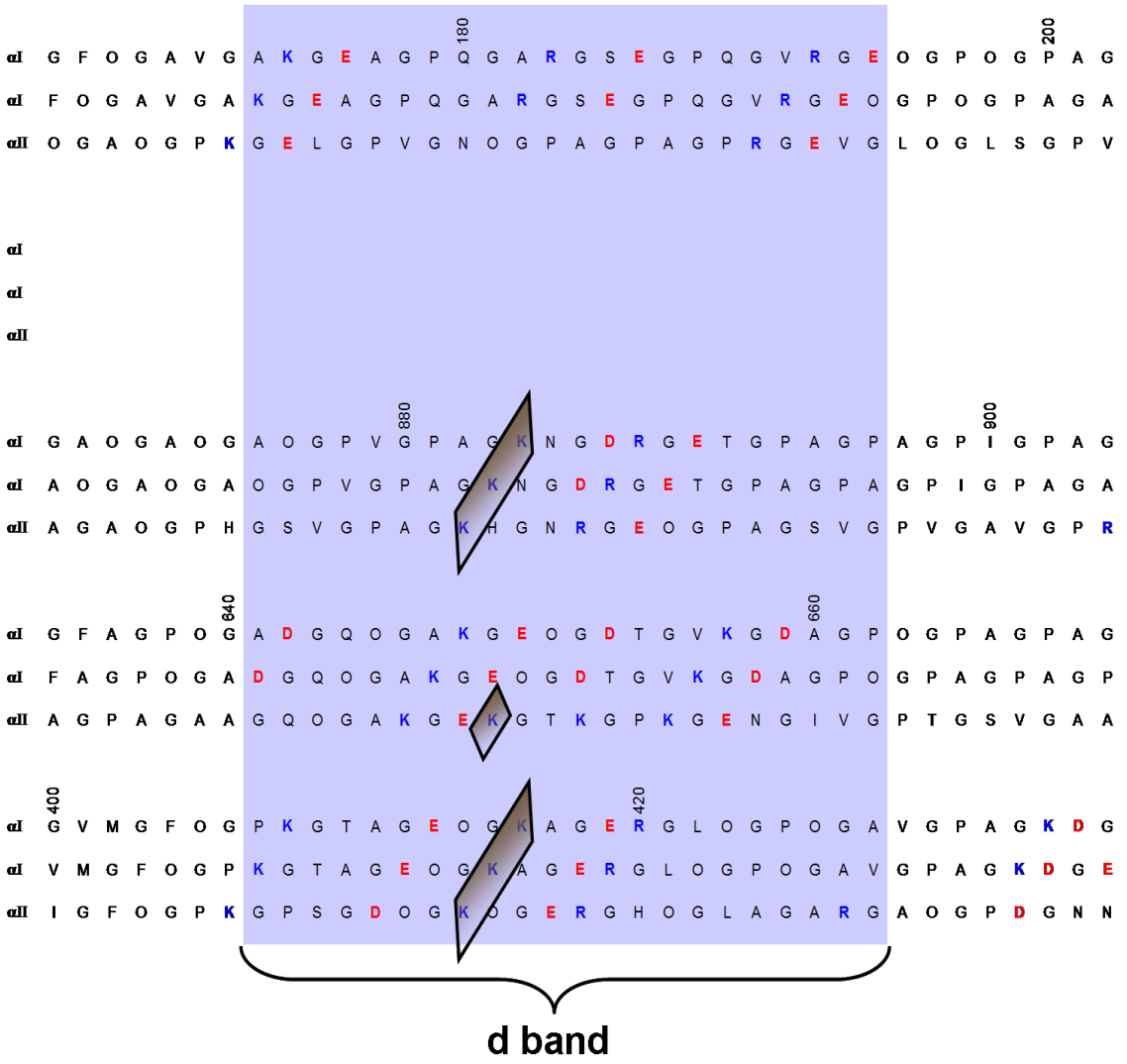


Figure 4.17 Representation of the portion of the simulated fibril containing the *d* band, showing the primary structure of each α-chain (identified on the left). All these molecule segments are triple helical in reality, represented in linear fashion for clarity and progressively aligned with an offset of 234 amino acid residues. They all extend beyond the diagram, on either side. The numbers denote triple helix numbered amino acid position (N→C) of the uppermost α1 chain, and these are equivalent along a line down-left. As the *d* band lies within the gap region, one of the molecules (the second molecule down, here) is not present. Each amino acid is represented using standard one letter notation, with "O" representing hydroxyproline. Lysine and arginine (K and R) are highlighted in blue; the acidic amino acids, aspartate and glutamate (D and E) are highlighted in red. The enclosed lysines are proposed preferentially-glycated residues after the 1 week galactose incubation.

Within the *b2* positive staining band, another significant difference peak exists post-1 week galactose incubation with only a very small 2 week sorbitol peak of borderline significance (Figure 4.14). A 2 week ribose peak and a 3 week fructose peak are present at the same location, which are proportionately larger in comparison than the galactose peak. The 1 week galactose difference peak is split at the top, one peak

being at amino acid position 96 along the simulated *D*-period with the other being at position 100. The correlation analysis in Table 4.3 suggests a strong co-distribution of these differences with basic and acidic amino acids as in the *d* band. However unlike the *d* band, there appears to be a stronger association of these changes with arginines as compared with lysines and also a greater parity of aspartates with glutamates as well as less dependence on the hydrophilicity.

In Figure 4.18 it can be seen that there is a denser packing of basic amino acid residues in the region of the *b2* band, in which glycation positive charge removal-related changes are most likely to occur. However unlike the *d* band, most of these are arginines: of six triplets of basic amino acids wholly or partially in the highlighted region, only two contain lysines: three lysines at positions 264 (α1- and α2-chains) and a lysine doublet at position 729 in the α1-chains, with an arginine at the corresponding α2-chain locus. It is generally presumed that the initial glycation reaction (i.e. attachment of sugar to amino group) occurs on lysine residues (Reiser et al., 1992b; Wess et al., 1993), with arginine as well as other lysines being involved as a result of secondary reactions with reactive intermediates to form further products and cross-links (Iijima et al., 2000; Biemel et al., 2002; Sell and Monnier, 2004). One of these is pentosidine (Sell and Monnier, 1989), which is a lysine-arginine cross-link and found to have been formed after the 1 week galactose incubation in Chapter 3, section 3.3.1 (see Table 3.1). Therefore here, lysines in the triplet at position 264 or the doublet at 729 are likely to be glycated and the secondary products might participate in further reactions and AGE/cross-link formation with nearby arginines, thereby removing further positive charge. Here, the lysines at locus 264 would be favoured, due to their precise location in the middle of the identified region of significance and its better alignment and proximity to potentially catalytic carboxylic acid groups (Shapiro et al., 1980) of aspartates and predominantly, glutamates.

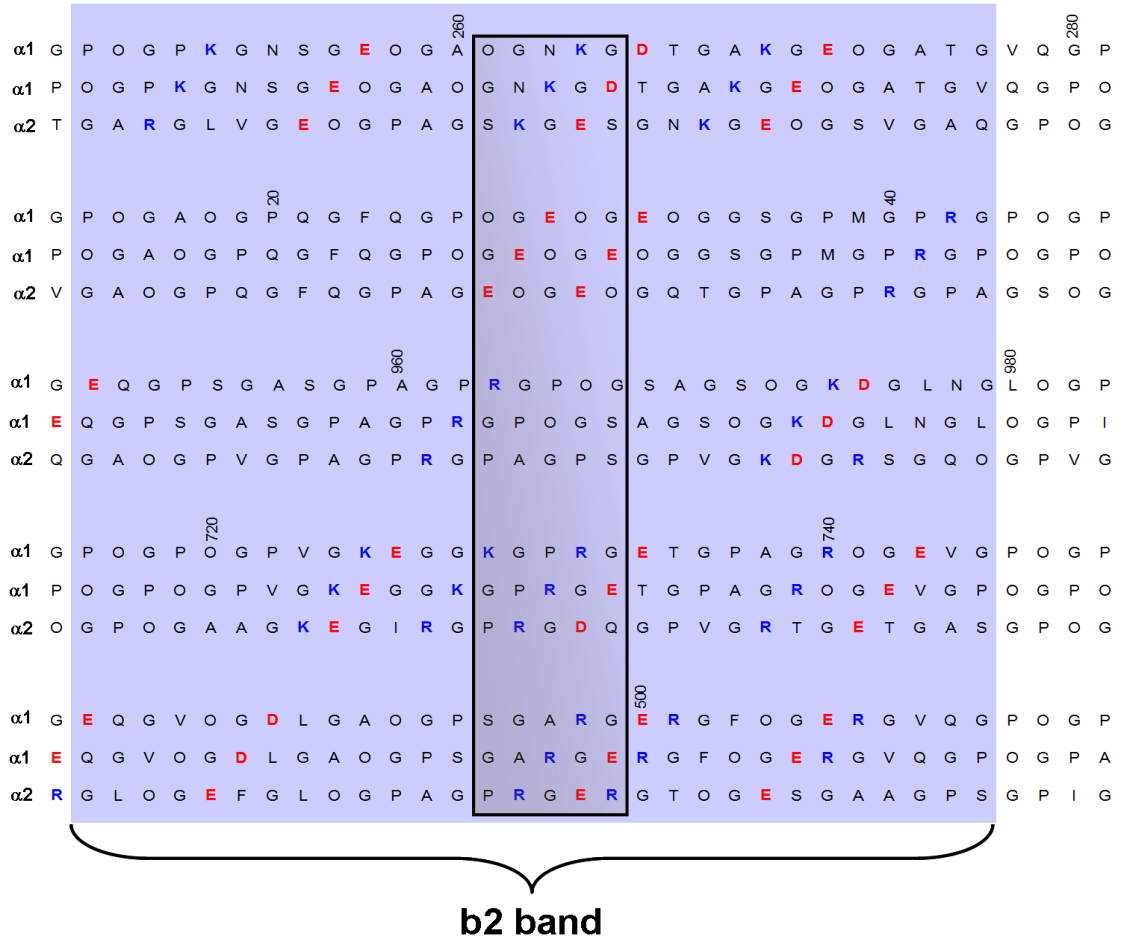


Figure 4.18 Representation of the *b2* positive staining band from the simulated fibril, in a manner similar to Figure **4.17**. This band is in the overlap region so all five molecular alignments are seen. Based on the 1 week galactose difference peak, the most likely region for glycation-related removal of positive charges is shown by the large rectangle. See text.

Attempts to locate the preferred site of glycation in the *b1* region are less easy to justify as there is a small 2 week sorbitol difference peak overlaying the 1 week galactose difference trace (Figure 4.14) suggesting that other binding mechanisms as previously discussed, may be factors in causing the changes observed here. However, specific loss of this (*b1*) as well as the *b2* and *d* bands, were all particularly evident even from the original micrographs for this treatment (section 4.3.3, Figure 4.6). This location (Figure 4.19) contains a triplet of lysines at triple helix position 290 that is within 3 amino acids along the primary sequences of its molecule's α-chains, of two consecutive triplets of glutamates. These in turn are precisely aligned with three more triplets of acidic amino acids and a further triplet of lysines at position 756. This is the highest concentration of acidic side-chain amino acids in the *D*-period. Therefore both of these lysine triplets are suggested as further sites for preferential glycation.

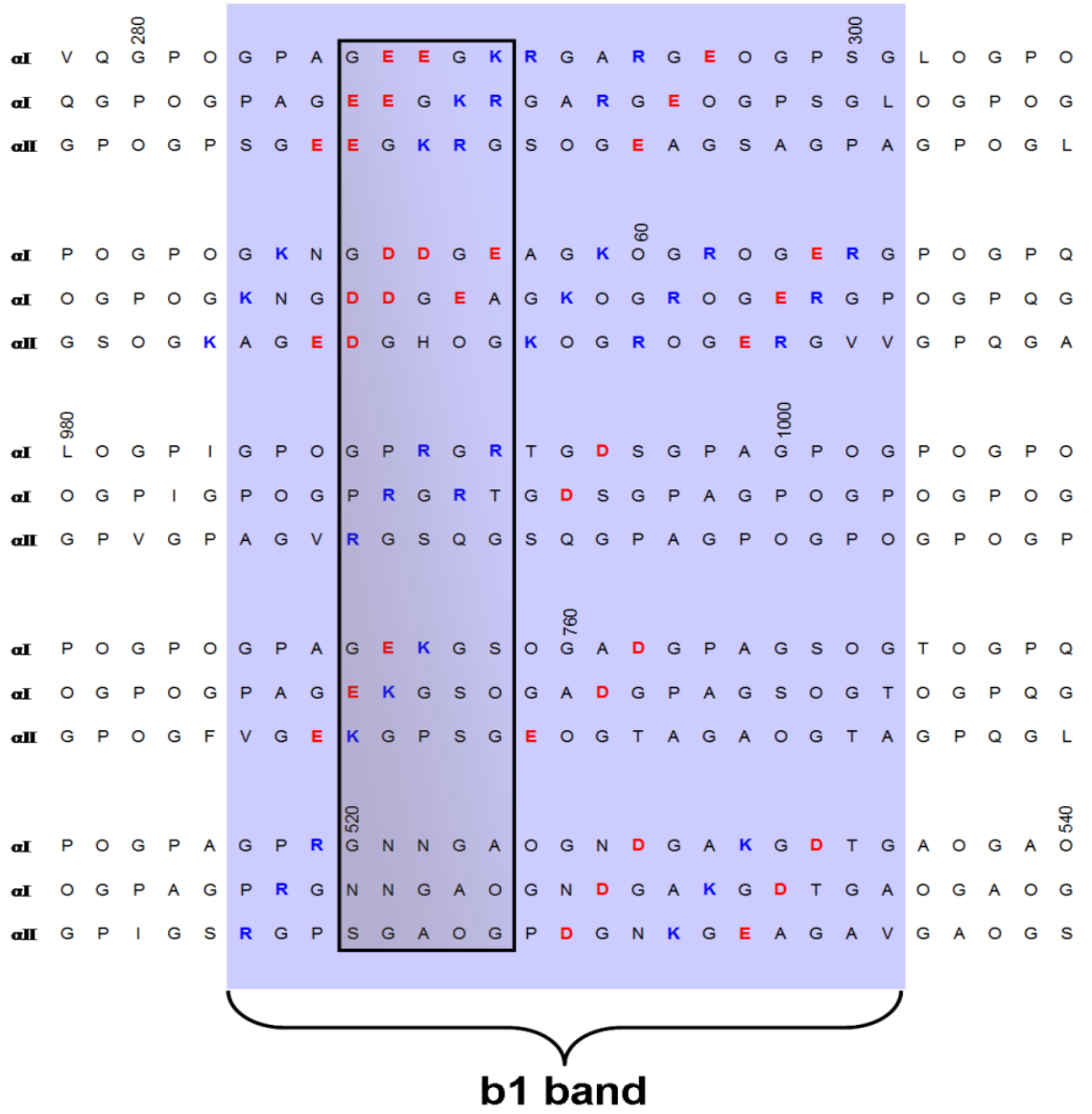


Figure 4.19 Representation of the *b1* positive staining band from the simulated fibril, in a manner similar to Figure **4.17** and Figure **4.18**. This band is in the overlap region so all five molecular alignments are seen. Based on the 1 week galactose difference peak, the most likely region for any glycation-related removal of positive charges is shown by the large rectangle. See text.

These analyses of potentially favoured glycation sites have been biased in favour of the 1 week galactose treatment due to the fact that this sugar has shown earlier detectable site-specific changes along the *D*-period than either fructose or ribose. This is fortuitous given its chemical resemblance to glucose (Syrový, 1994), which is quantitatively by far the most important and relevant sugar present in the ECM. Glucose's possible evolutionary selection as the monosaccharide of choice for use as a metabolic energy source in biological systems due to its comparatively slow rate of glycation (Bunn and Higgins, 1981) is emphasised in this study, by virtue of it having

produced no significant detectable positive staining changes over any of the glycation time periods used (see Section 4.3.3, page 84 and Figures 4.9 and 4.14). It has therefore established itself as a relatively good negative control. Noting the biochemical similarities, and bearing in mind the discrepancies, between the fluorescent AGEs formed after treatment of MTT with glucose and galactose as discussed in Chapter 3 (Section 3.4 and Figure 3.7) in addition to the two sugars' structural closeness, it is reasonable to suggest that galactose is a viable, faster-glycating alternative to glucose in *in vitro* glycation studies, resulting in more relevant glycation reactions and changes in collagen's structure than either fructose or ribose. Therefore the bias towards a preferential study of early sites of change during galactose treatment here is further justified.

Comparison of these results with other studies that have attempted to locate preferential sites and regions of glycation can be complicated somewhat by the mix of methodologies and approaches employed. Le Pape et al. (1984) conducted work both *in vivo* and *in vitro* and found some regions were preferentially glycated, including on or near with the hydroxylysine residues associated with the C-telopeptide regions but noted that results from diabetic rats *in vivo* showed a different pattern of specificity than was seen with *in vitro* glycation. Subsequent *in vivo* work by another group concurred that glycation does occur preferentially on hydroxylysine residues aligned with the N- and C-telopeptides (Wess et al., 1993). Conversely, from *in vivo* work using diabetic rats, Brennan (1989b) reported that no region along the collagen molecule was significantly more predisposed towards glucose attachment and glycation than any other, though presented evidence that the α2-chain seemed to become glycated to a greater extent than the α1-chains. Certainly, the results presented here do suggest that there are preferred sites of glycation along the *D*-period, but none of these were shown to be located close to either telopeptide in the *c2* or *a3* regions.

Other work has been successful in proposing specific preferred amino acid loci for glycation (Wess et al., 1990; Reiser et al., 1992b; Hadley et al., 1998). Reiser et al. (1992b) used *in vitro* glycation with glucose to locate four favoured sites of glycation: lysines at triple helix positions 434, 453, 479 and 924, all located roughly in the *c1-c2* region. Using neutron diffraction, Wess et al. (1990) also identified lysine 434 as a possible preferred *in vivo* glycation site in RTT with an additional one at position 855 (positions 220→222 along the simulated fibril – see Figure 4.14) in the middle of the *e1* band with Hadley et al. (1998) being in agreement re lysine 479. Most recently, Mikulíková et al. (2007) studied *in vitro* glycation of bovine collagen with ribose and

glucose using electrophoretic and HPLC separation techniques coupled to mass spectrometry, and found the preferential sites for glycation with glucose to be lysines 327 (*a2* band) and 855 with ribose showing the same preferred sites plus three more, i.e. lysines 573 (*a1* band), 684 (*c2* band) and 342 (*a1* band). Although a PBS buffer was used here (Mikulíková et al., 2007) as in the other *in vitro* studies cited (Le Pape et al., 1984; Reiser et al., 1992b; Hadley et al., 1998), only lysine 855 was in agreement with a preferred site determined by another study, i.e. Wess et al.'s (1990).

It has to be said that the data from this study do not add to the weight of evidence that any of these loci are preferred glycation sites inasmuch as none of the amino acid locations proposed here are in agreement with the above-cited research. However, Reiser et al. (1992b) studied preferentially glycated lysines on cyanogen bromide digest fragments of collagen which in total, only represented 25% of the length of the collagen molecule in what turned out to be – in this study – the most change-poor region of the collagen. Using similar methodology to that employed here, Hadley et al. (1998) found the major site affected by glycation to be the *c1* region. That region showed the weakest changes in this study, though it does contain just a single lysine triplet, at triple helical position 479, which Hadley et al. (1998) identified as being preferentially glycated. However, other aspects of their study were broadly in agreement with the results and inferences presented here, inasmuch as the *b2* and *d* bands were shown to have exhibited significant changes.

Due to the non-specific binding effects especially over the *a3-a2-a1-e2-e1* region, it has proven difficult to produce evidence for or against any previously-suggested sites in this region - e.g. lysines 327, 342, 573 and 855, mentioned above (see Wess et al. (1990), Mikulíková et al. (2007)) – from the data in this chapter, due to interferent non-glycation binding obscuring them.

Although carried out *in vitro*, Reiser et al.'s (1992b), Hadley et al.'s (1998) and Mikulíková et al.'s (2007) studies differed methodologically from this one, in using phosphate buffer, which is known to affect glycation sites and kinetics (Watkins et al., 1987; Lapolla et al., 1996), with MOPS apparently exhibiting a lesser effect (Watkins et al., 1987). Two of them also used higher concentrations of sugars than that used here, and over different time periods, i.e. ~2.8 M glucose over 24 hours in the case of Reiser et al. (1992b) and 0.5 M fructose over 5 and 11 days in the case of Hadley et al. (1998). Mikulíková et al. (2007) used 0.1M glucose and ribose over 7 days. Therefore noting the changes measured here and their variations with time, there are potential difficulties in assessing their comparability with these other studies.

The buffer used in this study was seen to produce changes that could affect the results (see section 4.3.3 and Figure 4.15). Although they used similar methodology to that employed here, it is possible that the fact that Hadley et al. (1998) did not allow for these occurring with their PBS buffer, relying on an untreated collagen sample as their control, could have distorted their results.

Wess et al.'s (1990; 1993) studies were *in vivo* and the degree to which they are comparable directly with *in vitro* experimental data is open to debate and this has been considered previously (Le Pape et al., 1984; Brennan, 1989a; Wess et al., 1993).

4.4.4 Summary

In conclusion, from the data obtained in this study it has been possible to suggest a number of possible favoured glycation sites. These are the triplet (at corresponding 2x α1- and 1x α2-chain loci) lysines at triple helix positions 416 and 884 in the *d* positive staining band, at positions 264 and possibly 729 in the *b2* band, plus potentially at positions 290 and 756 in the *b1* band. A lone preferred lysine has also been identified at α2-chain locus 651 in the triple helix, also in the *d* band. These do not agree with preferred glycation sites identified in other studies (Wess et al., 1990; Reiser et al., 1992b; Wess et al., 1993; Hadley et al., 1998; Mikulíková et al., 2007) but perusal of their primary sequence positions in the light of other potential factors such as concerted presence of other basic amino acids, acidic amino acids and generally, hydrophilicity would tend to suggest they are reasonable candidates for being favoured sites. Of course, site-specific conformational and steric considerations are unavailable and therefore could not be taken into account.

The discrepancies identified between the findings of this and other studies are likely to be as a result of differing methodological approaches in terms of whether the studies to which it is being compared were conducted *in vivo* (Wess et al., 1990; Wess et al., 1993) or *in vitro* (Reiser et al., 1992b; Hadley et al., 1998; Mikulíková et al., 2007), and in the latter case, the nature of the buffer constituents and sugars used and their concentration and time period, as considered on the previous page. Aspects of the differing analytical methodologies employed could also be a factor. Wess et al.'s (1990; 1993) neutron diffraction work was non-destructive inasmuch as deuteration-reduced intact tendon was studied. Conversely Reiser et al.'s (1992b) work studied collagen digest polypeptide fragments, using HPLC separation and amino acid compositional analysis and sequencing to locate preferred glycation sites. Similarly Mikulíková et al.

(2007) studied collagen digest polypeptide fragments by use of HPLC and capillary electrophoresis coupled to mass spectrometry to identify these sites. All of these studies had obvious limitations. Wess at al.'s (1990; 1993) work was of a comparatively low resolution, so suggestion of preferred amino acid sites of glycation was largely speculative. On the other hand, although the studies of Reiser et al. (1992b) and Mikulíková et al. (2007) produced specific preferred glycation locations, Reiser et al.'s (1992b) work looked at a limited length (about 25%) of the collagen molecule, with Mikulíková et al. (2007) having matched identified pepetide fragments with 77% of the entire primary sequence. Nonetheless the rationales for the production of the data presented in these other studies appear sound, therefore no experimental basis for the discrepancies between the preferred glycation sites suggested or determined by those and this study are obvious, except for the differences in the glycation protocols employed.

The only other comparable TEM study is that of Hadley et al. (1998). The discrepancies between the preferred glycation sites suggested in that work and those communicated here are similarly likely to have a basis in the difference between the *in vitro* glycation protocols utilised, especially their use of PBS buffer as opposed to the MOPS buffer used in these experiments. However as discussed on the previous page, Hadley et al.'s (1998) study used untreated collagen as a control as opposed to buffer-treated collagen control used here so the possibility of PBS-related changes, analogous to the MOPS buffer-induced changes detected here (see Section 4.4.1), was not discounted and these may have affected this group's results.

Other apparently non-glycation-related changes in the positive staining profile appeared to be a consequence of binding effects, due either to buffer and direct non-covalent sugar binding to the collagen fibril. It is possible that the non-glycating polyol used (sorbitol) has shed light on a mechanism by which the location of glycation specificity might be determined, inasmuch as the fibril is able to hold a sugar or sugar-like molecule in certain locations. It has yet to be elucidated how the buffer effects might affect sugar binding and glycation specificity, but it is considered that other buffer systems might show comparable effects to those seen with MOPS and therefore this should be taken into account when utilising this form of analysis.

Because of the difficulties inherent in standardising the time periods between sugars due to their differing reactivities, direct comparison of the changes and assessment of their similarities is difficult. Accepting that glucose brings about changes that are too slow for detection via this method and approach, there are nonetheless convincing

correlations between the earliest significant site-specific changes seen with the galactose, ribose and fructose (Table 4.1), but with some visually apparent differences. There is an interesting question as to why there are few or no changes of significance seen in the 1 week ribose and 1 and 2 week fructose data, whereas some of the same changes evident later with these sugars are seen with the slower-glycating galactose after only 1 week's incubation. From the data presented in Chapter 3, section 3.3, both ribose and fructose had clearly been actively glycating over the first week of the experiment. This can be explained by the early changes brought about by these two more reactive sugars being less specific than with galactose. Therefore, the hypothesis that the different sugars glycate with different specificities appears to be true when considering the temporal kinetic route to a given overall pattern of similar change. Mikulíková et al. (2007) produced evidence suggesting that glucose and ribose would ultimately react with the same sites and that the only difference in that respect is the fact that ribose is faster-acting. However, their investigation looked at lysines modified to the CML AGE, which is not a cross-link. The specificities of any cross-link AGEs formed would involve further lysines and arginines and probably be more highly dependent on the preferences of the sugars for forming different AGEs, which appears to be variable according to the data from Chapter 3, section 3.3.

The positive staining-TEM methodology used here to assess patterns of positive charge removal during glycation and AGE formation has shown itself to be a useful tool but potentially prone to interferences which make interpretation challenging. As we still have an incomplete picture on the basis of biochemical and other evidence, as to the location of the preferentially glycated sites along the collagen fibril (Reiser et al.'s (1992b) 25% complete picture is still cited as being definitive (Di Lullo et al., 2002; Sweeney et al., 2008)), this technique is potentially useful in helping to identify more sites and regions along the *D*-period, which might warrant further investigation.

Chapter 5: Construction of an axial one-dimensional electron density map for untreated mouse tail tendon collagen using X-ray diffraction and isomorphous replacement techniques

5.1 Introduction

The theory of X-ray diffraction has been described in depth in various texts (e.g. Cantor and Schimmel, (1980); Glusker, Lewis and Rossi, (1994)). The methodology hinges on the interaction of X-rays with, and therefore scattering by, centres of electron density. The diffraction pattern recorded is a representation in reciprocal space, of the ordered arrangement of regions of differing electron density in a crystalline (or near crystalline) structure. The smallest repeating unit of this is called the *unit cell* and along the collagen fibril, this represents the *D*-period.

It is well known that the arrangement of type I collagen molecules within fibrils and the crystallites (microfibrils) contained within them has a high level of order axially (Hodge and Petruska, 1963). This is evidenced by the sharp axial or meridional reflections that are collected during X-ray diffraction experiments using tendon tissue from a variety of animal species (Bear, 1944; James et al., 1991; McBride Jr et al., 1997) (see Chapter 1, section 1.4.1.2 and Figure 1.9). Conversely, the equatorial pattern shows more diffuse reflections indicative of the variability of the packing of collagen molecules across the *D*-period and especially within the gap region (Wess et al., 1998b; Orgel et al., 2006) where there is known to be greater disorder than in the more constrained overlap region.

5.1.1 X-ray data requirement to produce an electron density map

In order to successfully deduce the axial one-dimensional electron density map of MTT, the sharp meridional reflections are required. As stated in Chapter 1, section 1.4.1.2, the two types of information made available during an X-ray diffraction experiment, theoretically, are the intensities and therefore *amplitudes* of each reflection and their *phases*. In practice, amplitudes are straightforward to measure from the recorded X-ray diffraction pattern, whereas phases cannot be measured directly and so represent an inherent difficulty within X-ray crystallography.

5.1.1.1 Amplitudes

If the intensity or relative brightness of a reflection can be measured, appropriately corrected and integrated accurately, then the square root of the intensity, I_l, for the reflection of order l gives the amplitude, A_l for that order:

$$A_l = \sqrt{I_l} \qquad \text{-} \qquad \textit{Equation 1}$$

Each reflection's amplitude is a consequence of the distribution of electron density within the repeating units of the diffracting axial structure. The amplitudes alone cannot be used to calculate an electron density map but they can still yield useful information about the magnitudes and vector spacings between centres of electron density within a crystal. The greater the number of orders there are with measurable amplitudes, the greater the resolution of eventual electron density that would be achieved. With a collagen *D*-period (unit cell) of 67 nm length, 10 orders would result in an eventual electron density map of 6.70 nm resolution, 20 orders would give a resolution of 3.35 nm and 100 orders would give a resolution of 0.67 nm.

A summation of the amplitude information for all the orders can be achieved by addition of all their cosine waves. Figure 5.1 illustrates this process, known as "Fourier synthesis". Effectively, each successive harmonic corresponds to successive diffraction orders, so it can be seen that by including amplitude information from higher harmonics the resolution in the synthesis increases. As only amplitude data was used (i.e. no phase information), the result is a "Patterson plot". This can give useful information as the peaks represent vectors between all possible combinations of centres of electron density within the unit cell or *D*-period of the experimental sample. Patterson plots are always centrosymmetric and have their largest peaks at 0% and 100%, which represent every individual point of electron density's vector distance from itself. All of the Patterson plot data is contained in the 0 – 50% (or 50 – 100%) section of the plot, so often just half of the plot needs to be shown. Although not electron density maps, they can yield useful data about the number, magnitude and positions of electron-dense centres within a sample.

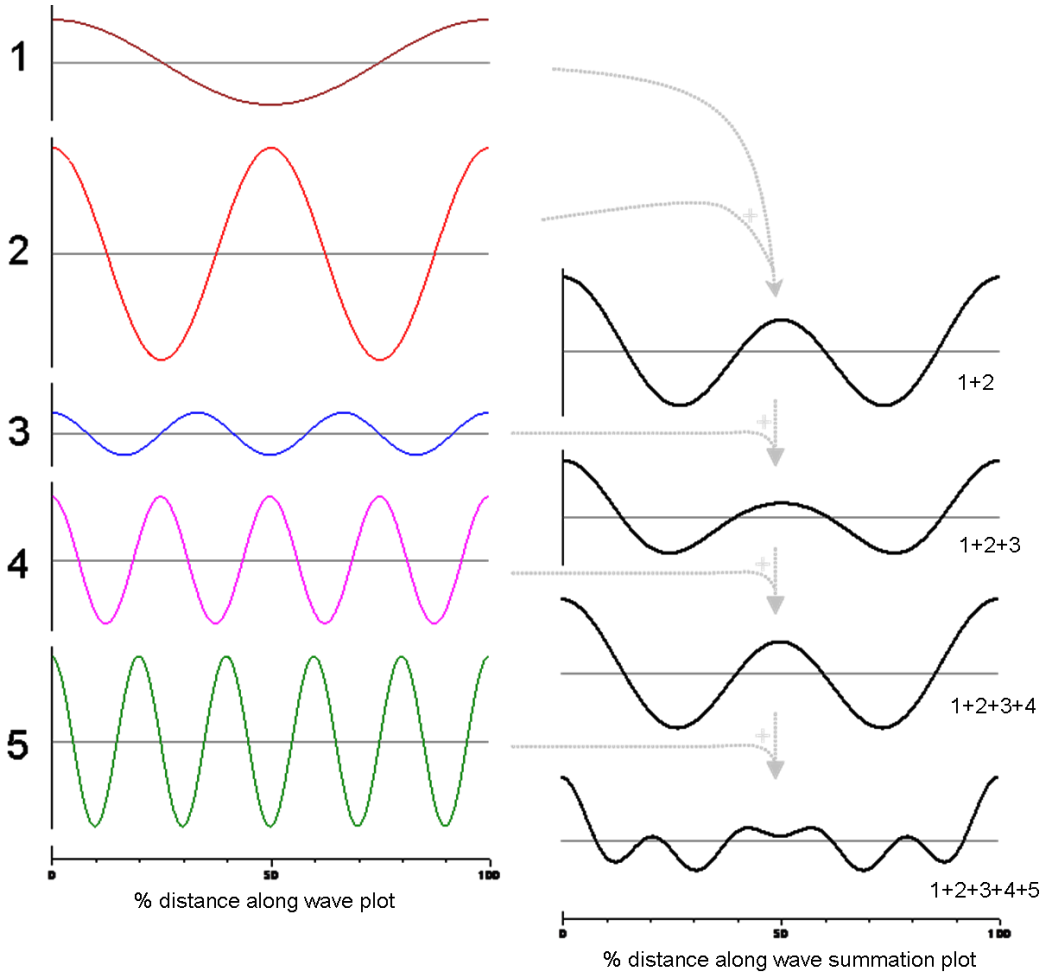


Figure 5.1 Demonstration of the progressive summation of the amplitudes of five orders of reflections without phase information. The coloured waves represent the individual orders 1→5, order 1 consisting of one complete cosine wave, order 2 consisting of two, etc. The amplitudes of these waves are represented on the same vertical scale by the peak heights and are: 2 units for order 1; 5 units for order 2; 1 unit for order three; 3 units for order 4; 4 units for order 5. The black traces on the right are progressive summations, starting at the top with order 1 + order 2, representing the effect of addition of progressive orders on the plot. These plots as applied to squared actual amplitude data are known as Patterson plots.

A one-dimensional Patterson plot can be represented by

$$P(z) = 1/\mathrm{L} \sum_{l} |F(l)|^2 \cos 2\pi(lz) \qquad \text{-} \qquad \textit{Equation 2}$$

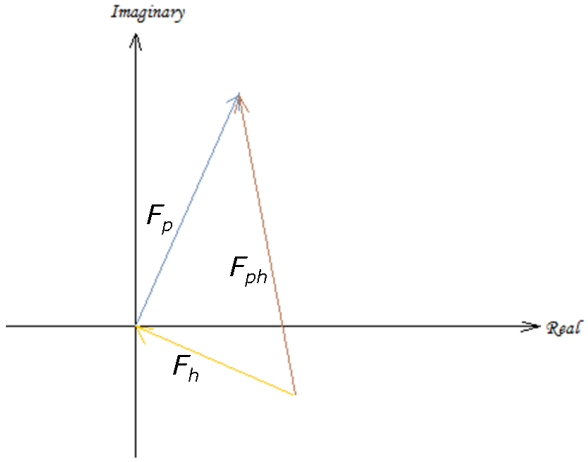
along the fibrillar axis, $|F(l)|$ being the amplitude of meridional order l, L representing unit cell (*D*-period) length with z being the fractional Cartesian coordinate along the unit cell.

5.1.1.2 Phases

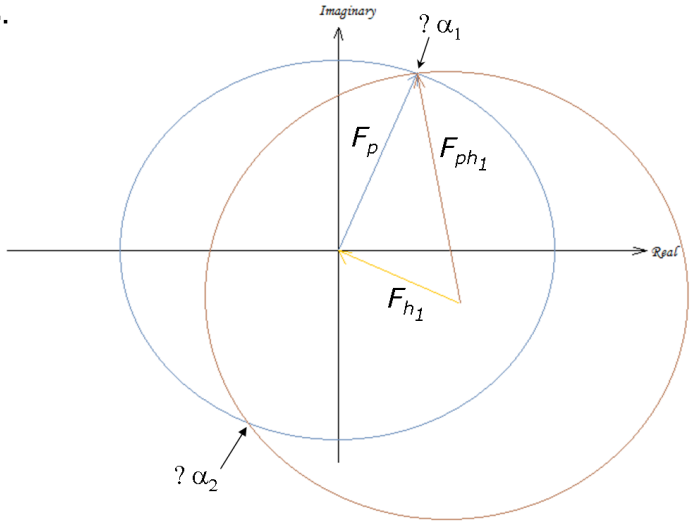
Unlike amplitudes, there is no direct way to measure the phase angle from X-ray diffraction patterns therefore the phase information is lost. This gives rise to the *phase problem.* Estimation or measurement of the phase angle for each reflection is an indirect and deductive process.

The approach taken to measure phases depends on the type and size of molecule under consideration and a number of methods are considered elsewhere (Glusker et al., 1994). In the case of type I collagen, work to determine phase angles for the reflections has been carried out using observed X-ray data alongside models of electron density distribution based on primary sequence of collagen molecules within fibrils (Hulmes et al., 1977) or by using data from banding or staining patterns recorded using electron microscopy (Chandross and Bear, 1973; Meek et al., 1981) based on assumptions regarding correlation of electron density with incorporation of stain into fibrils. Both approaches have drawbacks. Meek et al. (1981) conceded that the electron microscopy approach is limited by the resolution of the TEM, which using the negative stain uptake approach of in fibrils from corneal stroma was equivalent to 9 orders (7.4 nm). In addition unlike the hydrated tissue generally used in X-ray diffraction experiments, fibrils viewed by TEM are in the dried state and therefore likely to vary in terms of the spacings of regions along the *D*-period. Methods based entirely on models are susceptible to errors in the assumptions upon which the models are based. Hulmes et al. (1977) produced several models of RTT collagen electron density distribution. They selected the best one(s) based on calculated amplitudes and phases derived from these models, compared with experimental amplitudes with the calculated phases. However, though there were two models that showed a good fit with the experimental amplitude-based electron density data, none of the models fitted precisely.

A.

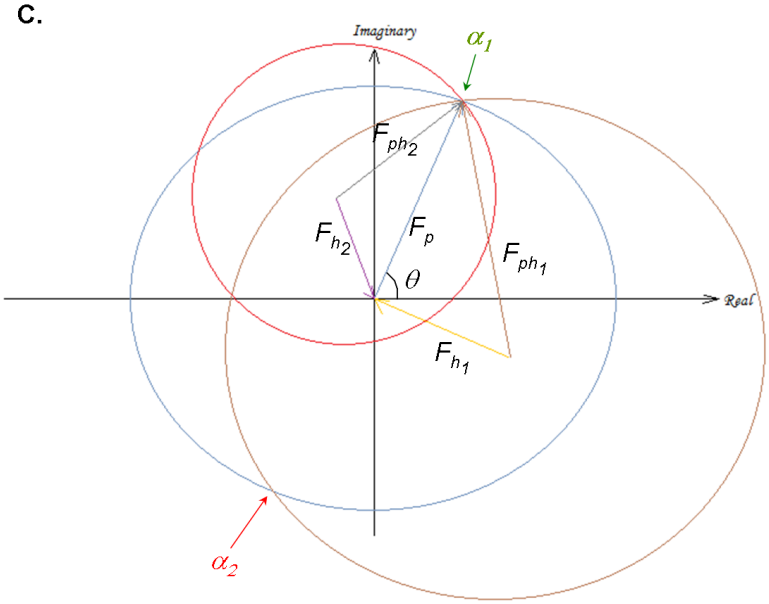


B.



Two possible phases from the data: α_1 & α_2

C.



α_1 identified as the **correct phase,** θ

Figure 5.2 (Previous page) **A.** An Argand plot of amplitudes and phases (structure factors) for a given order, for untreated collagen (F_p), a heavy atom (F_{h1}) and collagen after isomorphous replacement with heavy atom *h1* (F_{ph1}). Lengths of the structure factor vectors are proportional to their respective amplitudes, with their anticlockwise angular deviation away from an axis along or parallel to the "real space" axis (0 or 2π radians) representing each structure factor's phase angle. Ideally, this resolution of vectors results in a closed triangle. In practice, only the amplitudes for F_p and F_{ph1} are known, with both amplitude and phase being known for the heavy atom component. **B.** Harker diagram for untreated native collagen and its isomorphous derivative. The amplitude component of F_p is represented by the radius of a circle with its centre at the origin of real and imaginary space. A second circle with its radius being the amplitude of the *h1* collagen derivative, centre offset from the first circle by the *h1* structure factor vector, is drawn. The two circle intersects represent two possible phase angles, $\alpha 1$ and $\alpha 2$, one correct and the other incorrect. **C.** Use of a second collagen derivative using heavy atom *h2* produces a third circle that intersects with only one of the two circle intersects. This point around the F_p circle is identified as the correct phase.

Bradshaw et al. (1989) demonstrated the use of *isomorphous replacement* as a model-independent means of phase determination in type I collagen fibrils. It is based on the use of heavy atoms that are known to bind to proteins at specific sites. With X-ray diffraction data collected from untreated native and isomorphically-replaced native collagen-containing tissue respectively, as the binding sites and electron densities of the heavy atoms are known, it is possible to narrow down the phase probabilities for each reflection. The method of double isomorphous replacement for phase determination in proteins was first described by Harker (1956) and an illustration of the method is shown in Figure 5.2. It is based on known (measured) amplitudes for untreated and isomorphically-replaced collagen and known structure factors (amplitudes and phases) for the heavy atoms used – separately – for the isomorphous replacement experiments. This enables plotting of the data in the complex plane (real and imaginary space). Plotting of the three structure factors on an Argand plot ideally results in the production of a closed triangle (Figure 5.2A). Of course, the phases are not known for the untreated and isomorphically replaced collagen. Use of the Harker construction with just one isomorphic derivative results in two possible phases being identified at the circle intersects, one correct and the other incorrect (Figure 5.2B). Using a second heavy atom isomorphic derivative enables the correct phase to be determined (Figure 5.2C).

In order to apply this method, the non-isomorphic and isomorphic derivative data must be scaled so that their amplitudes are comparable for phase probability determination. This can be done using Patterson plots. If the binding sites of each heavy atom along the *D*-period are known, a predicted *difference Patterson plot* can be calculated. Where

appropriate scaling factors have been applied to the amplitudes of all the reflections to allow for differences in the samples, especially sample thickness, an experimental difference Patterson plot can be calculated, represented by

$$P(z)_{Diff} = P(z)_{Deriv} - P(z)_{Native} \quad \text{-} \quad \textit{Equation 3}$$

where $P(z)_{Native}$ and $P(z)_{Deriv}$ represent the untreated and isomorphically-replaced collagen Patterson functions and $P(z)_{Diff}$ is the experimental difference Patterson plot. The difference Patterson plot represents all possible vectors – originating from the origin – between heavy atom binding sites along the *D*-period. Appropriate scaling of $P(z)_{Native}$ and $P(z)_{Deriv}$ relative to one another would result in an experimental difference Patterson plot close to the predicted function. Therefore the closeness of the predicted difference versus experimental heavy atom Patterson plots can be used as a measure of the effectiveness of scaling prior to phase determinations. It is important that a heavy atom that binds to an appropriately distributed amino acid is selected. A Patterson plot for the native protein (or its derivative) alone is complex due to the thousands of electron dense centres (atoms) distributed along and within it. Therefore, the plot becomes a smooth curve rather than a series of peaks. Similarly, choice of heavy atom is important. Selection of a heavy atom-containing compound such as PTA has been of use in studies requiring the enhancement of fainter reflections in X-ray diffraction patterns (Fraser and Macrae, 1981; Wess et al., 1995). However, here this would result in a complex and unusable heavy atom (tungsten) difference Patterson plot due to the high number of lysine and arginine residues present. Heavy atoms that bind to amino acids, comparatively few of which are present per molecule or *D*-period are ideal as characteristic difference Patterson plots with defined and easily predictable peaks are achievable. In this study, gold and iodine have been selected as used previously ((Bradshaw et al., 1989; Orgel et al., 2000; Orgel, 2000). Gold binds to methionine and histidine, and iodine binds to tyrosine and histidine (Tsomides and Eisen, 1993; Glišić, Rychlewska and Djuran, 2012). Comparatively few of each of these amino acids are found per collagen *D*-period. Specifically, there are 22 methionines, 12 histidines and 13 tyrosines (UniProt, 2012).

With this procedure having been completed for both derivatives, phase determination can be carried out using the Harker construction method.

5.1.2 Production of an electron density map for MTT collagen

With the structure factors (amplitudes and phases) determined for each reflection, it is possible to produce a one-dimensional electron density map using a Fourier synthesis as described by

$$\rho(z) = 1/\mathrm{L} \sum_{l} |F(l)| \cos 2\pi(lz - \alpha_l) \qquad - \qquad \textit{Equation 4}$$

where $\rho(z)$ is the electron density along the *D*-period axis and α_l is the relative phase angle for each meridional Bragg reflection. Note that the fundamental difference between this and the Patterson function (Equation 2) is the inclusion of phase information, which locates each diffracted wave into the correct position relative to the others (Figure 5.3).

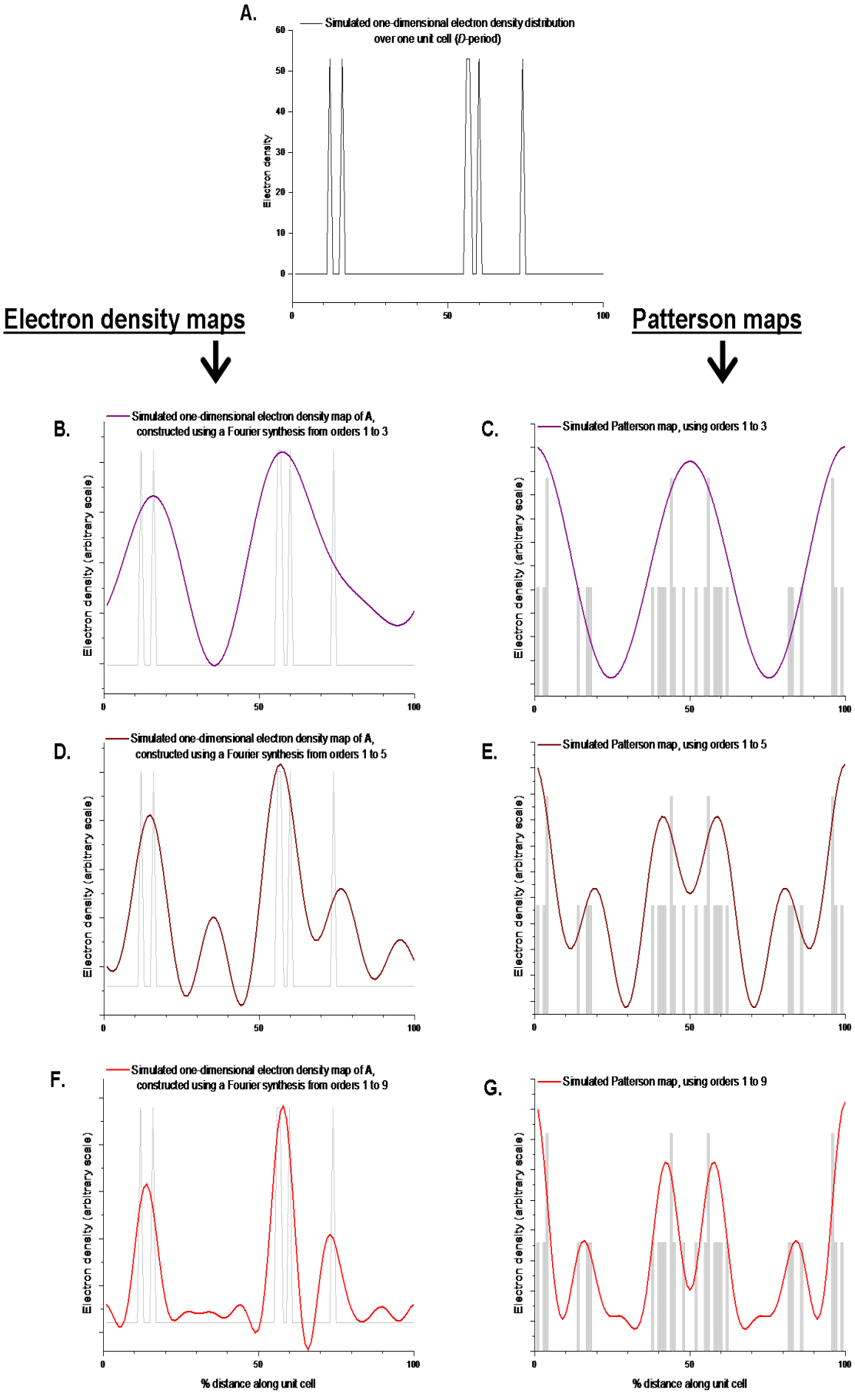

A.
Simulated one-dimensional electron density distribution over one unit cell (D-period)
Electron density
Electron density maps
Patterson maps
B.
Simulated one-dimensional electron density map of A, constructed using a Fourier synthesis from orders 1 to 3
C.
Simulated Patterson map, using orders 1 to 3
D.
Simulated one-dimensional electron density map of A, constructed using a Fourier synthesis from orders 1 to 5
E.
Simulated Patterson map, using orders 1 to 5
F.
Simulated one-dimensional electron density map of A, constructed using a Fourier synthesis from orders 1 to 9
G.
Simulated Patterson map, using orders 1 to 9
Electron density (arbitrary scale)
% distance along unit cell

Figure 5.3 (Previous page) Illustration of construction of Fourier synthesis of experimental one-dimensional electron density maps (**B**, **D**, **F**) of "actual" simulated electron density distribution (**A**), which involves both amplitude and phase information (see Equation 4). The actual centres of electron density are superimposed onto **B**, **D** and **F** in light gray, for comparison. The corresponding Patterson maps (see section 5.1.1.1, Figure **5.1** and Equation 2), constructed without use of phases, are also presented (**C**, **E**, **G**). Plots **B** and **C** are constructed using reflection orders 1→3, **D** and **E** using orders 1→5 and **F** and **G** using orders 1→9. As the number of orders used increases, so does the resolution. The actual map contains six equally electron dense peaks, at positions 12, 16, 56, 57, 60 and 74% along to unit cell. A histogram of all 30 possible spacings between these (excluding each with itself, i.e. 0 and 100%) has been superimposed onto the Patterson plots (**C**, **E**, **G**) in light gray. Therefore their potential usefulness and significance can be appreciated.

5.1.3 X-ray source

All data presented in this Chapter 5, and Chapter 6, were collected on beam line 18-ID run by the Biophysics Collaborative Access Team (BioCAT) at the Advanced Photon Source (APS) synchrotron at the Argonne National Laboratory, Illinois, USA. The beam line's position within APS is shown in Figure 5.4.

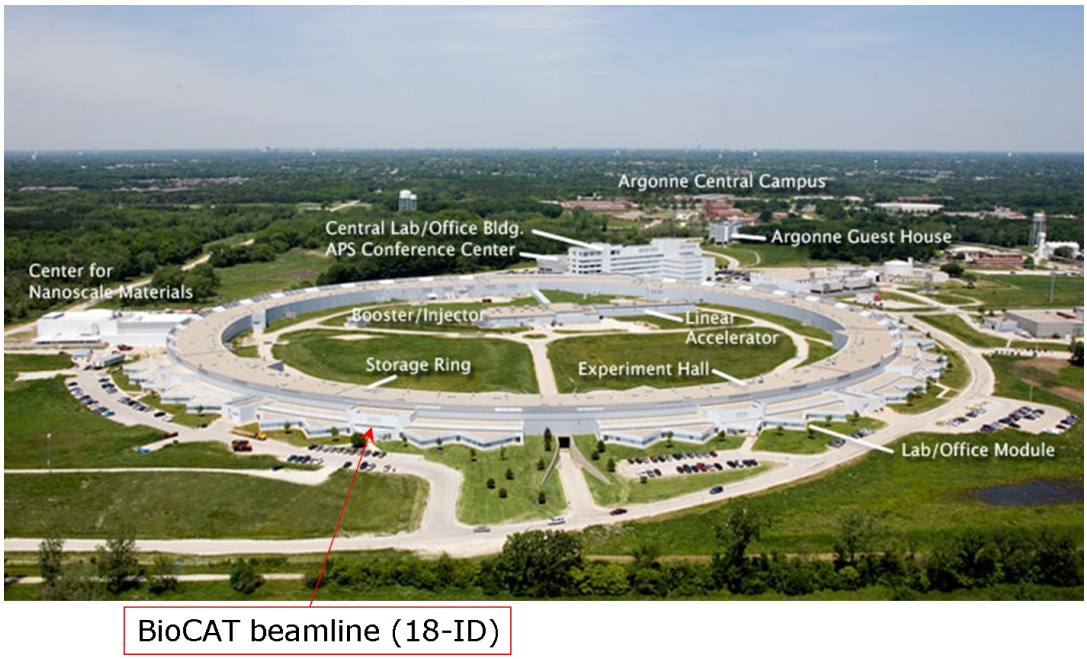


Figure 5.4 Aerial view of APS (from www.aps.anl.gov/About/APS_Overview, as of August 2013). The principle components of the synchrotron are labelled. The position of the BioCAT beam line used for the X-ray data collection portion of this work is indicated by the red arrow. For an idea of scale, cars can be seen on the photo and the circumference of the electron storage ring contained within the circular component of the building is 1104 m.

APS is a third generation “synchrotron” (Moncton, 2002) and is presently the brightest synchrotron source of X-rays for research in the USA and in fact, the Western Hemisphere. More information about its operating parameters can be found at www.aps.anl.gov/About/APS_Overview (As accessed in August 2013). The storage ring uses over 100 electromagnets to contain the stored electrons, which have been accelerated to an energy of 7 GeV and are travelling at a relativistic speed.

The X-rays are produced for the BioCAT beam line in sector 18 by an insertion device (ID) placed inside the storage ring and designed to tap energy from the high energy stored by the electrons. Third generation synchrotrons use undulators for this purpose. Undulators are comprised of a row of paired magnets, each pair being on opposite sides of a gap and polarity-aligned but of the opposite alignment to the neighbouring pair. This arrangement of alternating magnetic field polarity continues down the gap. Stored electrons propagate along the gap and oscillate, losing some energy as they do so, emitting photons. Undulators are capable of producing a comparatively high brilliance of radiation with low energy bandwidth and angular divergence, therefore a high flux density. The beam line can produce X-ray photons equivalent to the energy range 3.5-35 keV with a flux of 2 x 10^{13} photons/sec at an energy of 10 keV (https://beam.aps.anl.gov/pls/apsweb/beamline_display_pkg.display_beamline?p_beamline_num_c=25, 2012). For the interested reader the design specifications and components involved in the production, monochromation and focussing of X-rays on the BioCAT beam line are described in more detail by Fischetti et al. (2004).

The APS synchrotron has enabled high quality-low noise X-ray diffraction data collection in this work over a time period of a matter of seconds per dataset.

5.1.4 Aims of the work in this chapter

This work has been carried out with the aim of producing an axial one-dimensional electron density map of MTT collagen using X-ray diffraction and isomorphous replacement, using two heavy atom derivatives of native MTT collagen for phase determination. It is anticipated that this will be similar to the RTT map presented previously (Orgel et al., 2000; Orgel, 2000) and complementary and comparable to the native MTT TEM data presented in Chapter 4. Within this work, the MTT map will be used as a baseline for the X-ray diffraction study of MTT collagen glycated *in vitro* presented in the following chapter allowing glycation-related changes to be studied. Beyond this work, it is hoped that the map will form the basis of further study to produce a three-dimensional electron density map of MTT collagen.

5.2 Materials and methods

All experiments in this chapter were performed on the BioCAT (18-ID) beamline at the APS synchrotron at Argonne National Laboratory, Illinois, USA.

5.2.1 Materials

The chemicals and consumable materials used were from storage in the BioCAT beam line preparation laboratory, except sodium tetrachloroaurate(III) dihydrate ($NaAuCl_4 \cdot 2H_2O$), cat. number 298174-1G, obtained from Sigma-Aldrich Company Ltd, Gillingham, Dorset, UK.

In addition to that described in Chapter 2, section 2.1, the computer software used in data analysis in this section were FIT2D (Hammersley, 1997), OpenOffice.org v.3.3 and using Linux Ubuntu OS, PGPlot isomorphous phasing software developed at Illinois Institute of Technology (IIT) and kindly provided by Dr Orgel. ImageJ v. 1.46i (Rasband, 2011) was used to rotate and adjust diffraction pattern images, in preparation for inclusion within this work.

5.2.2 Preparation of MTT fibres for X-ray diffraction analysis

The procedure for obtaining mouse tails and dissecting out the tendons was described in the Chapter 2, section 2.3. The tendons used were from mouse tails removed from 3 month and 2 day old male CD-1 mice immediately *post mortem* and then stored at -20^{o}C in the School of Optometry and Vision Sciences at Cardiff University until transport to APS on dry ice. The samples were then stored at -80^{o}C until required for use. All MTTs were dissected into a plastic Petri dish containing PBS (137 mmol/l NaCl, 2.7 mmol/l KCl, 10 mmol/l Na_2HPO_4 and 1.8 mmol/l KH_2PO_4, adjusted to pH 7.4).

X-ray diffraction data were required for:-

1. Native MTT, stored in PBS post-dissection for at least 10 minutes;

2. Native MTT, isomorphically replaced with iodine atoms; and

3. Native MTT, isomorphically replaced with gold atoms.

Heavy atom isomorphous derivatives were produced by placing MTT fibres into a derivatizing solution for at least 1 hour. These were separate 1 mg/ml KI or 1

mg/ml $NaAuCl_4 \cdot 2H_2O$ solutions in PBS, for iodine and gold respectively (Orgel, 2000; Antipova, 2008). Following isomorphous replacement and prior to X-ray diffraction analysis, individual fibres were removed from the derivatizing solution by grasping one end with fine-tipped tweezers and immersed and continuously moved through PBS solution for approximately 10-15 seconds to wash.

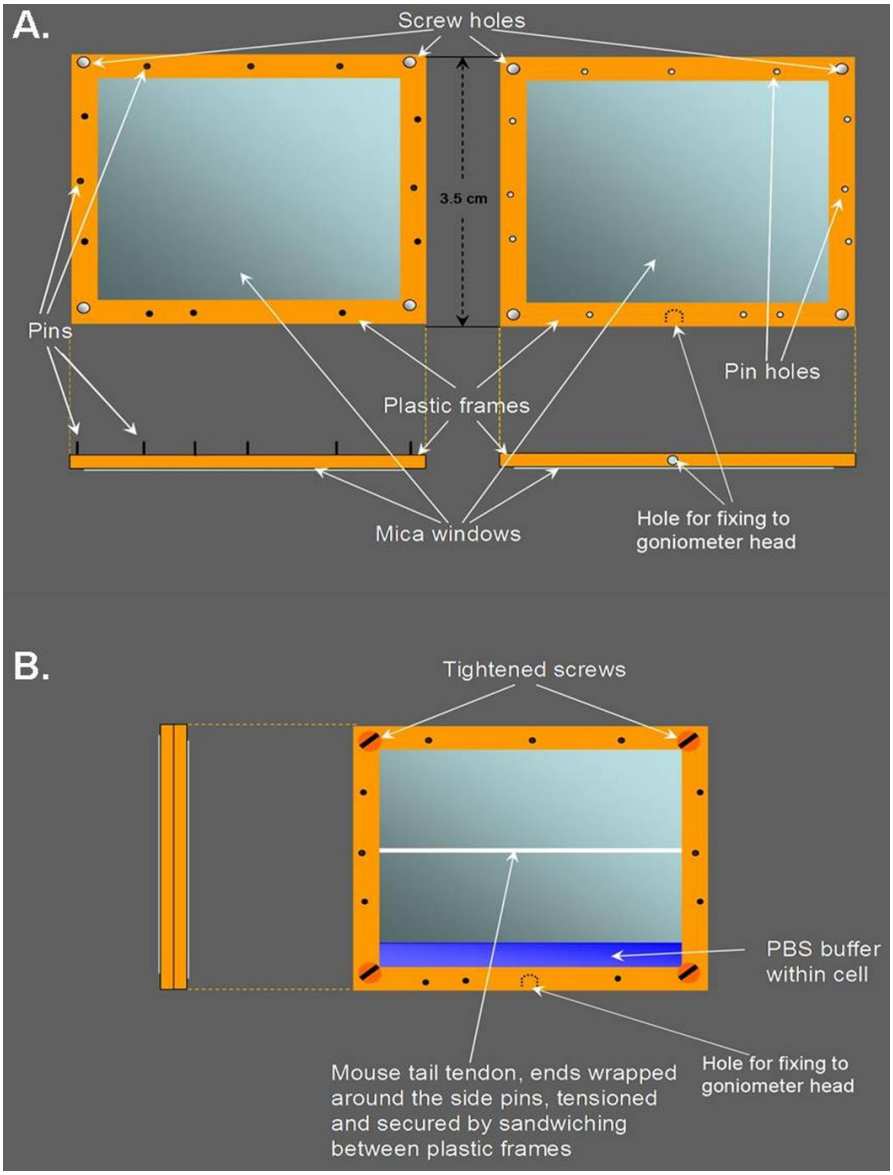


Figure 5.5 Diagram showing sample holder used to support MTT fibres during X-ray diffraction experiments. The holder is assembled from two halves (A), with locator pins on one side fitting into holes in the other half, to ensure accuracy of alignment. The halves are held together tightly by a screw in each corner (B). The MTT fibre's ends are wrapped around a pin on each side until the fibre is taut and then with a small amount of PBS in the chamber, the halves are assembled and sealed with screws.

Immediately prior to data collection, each MTT fibre was installed in a sample holder (Figure 5.5), kindly provided by Dr Orgel. Native non-isomorphically replaced fibres were transferred from the PBS into which they had been dissected and isomorphically-replaced fibres, from the post-derivatization wash. Care was taken to ensure that fibres were handled only by their ends. Installation was effected by wrapping each fibre's ends around pins at opposite edges of the frame, ensuring the fibre was under sufficient tension to remove the tendon crimp. The ends of the fibre were then further secured by sandwiching between the complementary halves of the holder when assembled and closed. The chamber inside was kept humidified by the presence of a small amount of PBS within to ensure the tissue remained hydrated. Each fibre was suspended horizontally, above the buffer.

5.2.3 X-ray diffraction experimental setup

Following installation of the subject MTT fibre in the sample chamber, the holder was transported to the experimental hutch. A goniometer head was attached to the sample holder by means of a metal prong on top, which was inserted into the hole at the base of the sample holder. Its magnetic base was then fixed firmly in place on top of the moveable goniometer platform. The sample holder was situated between the end of the source of the focused monochromatic X-ray beam and the mica window at the entrance to the evacuated flight path chamber. During setup of the beam-line for these experiments, distances between these components were minimised to avoid significant attenuation of the X-ray beams by air due to scattering.

The experimental arrangement is shown in Figure 5.6. This was an 8 keV energy setup for low to medium angle diffraction data collection, therefore the X-ray beam was of monochromatic wavelength 1.54 Å. The beam was restricted to 100 µm across by means of beam-defining slits immediately upstream of the beam's egress point from the source assembly. The windows in the sample holder and the front of the flight path chamber were made of mica, as this material produces minimal interference with, or scattering of the X-ray beams. The flight path chamber was kept evacuated. With most of the distance between the sample and the detector being a vacuum, there was no means of removal of X-rays within this space.

At the rear of the detector was a window made of Kapton film, upon which was supported a beam stop to prevent the highly intense primary X-ray beam from reaching and destroying the detector. Behind this was the MAR 165 CCD detector itself. The detection surface was a 165 mm circular array of 4096 x 4096 X-ray-sensitive pixels,

mapped onto a single CCD chip via fibre optic tapers. Resolution was adjustable. For these experiments, the resolution used was 1024 x 1024 pixels, i.e. 4 x 4 binning, giving an effective pixel size of 160 μm. The diffraction set-up had been calibrated using the 5.8380 nm reflection in powder diffraction patterns from silver behenate (Huang et al., 1993) and the 67 nm meridional spacing of RTT.

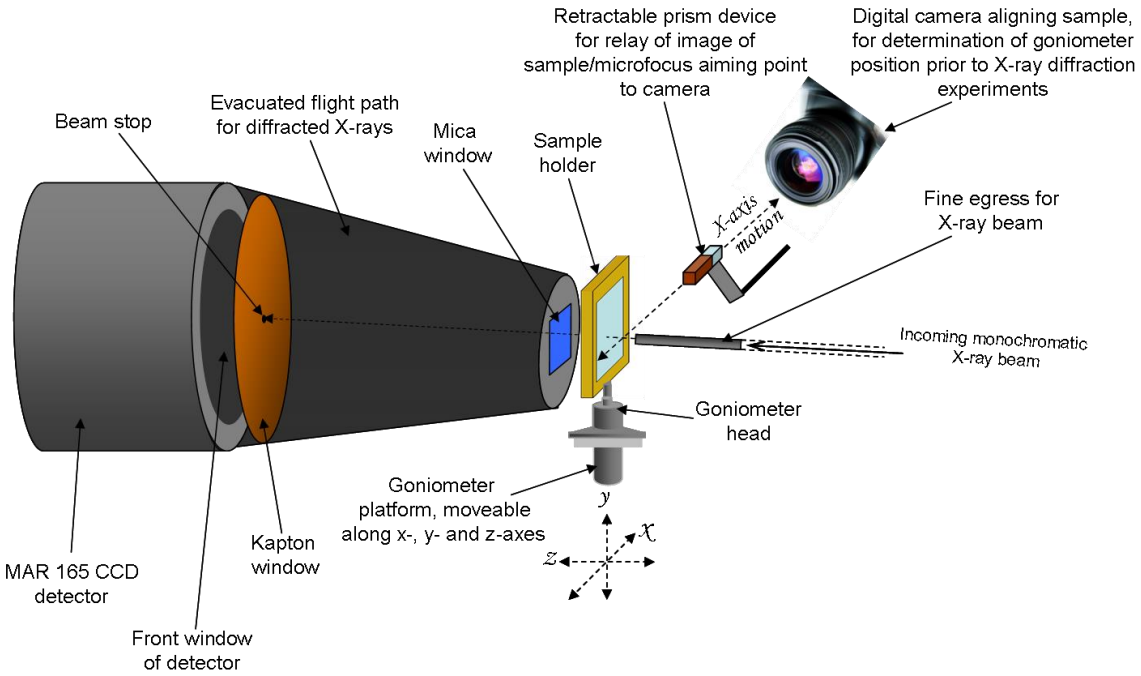


Figure 5.6 Positioning of sample holder relative to the other proximal active components of the X-ray diffraction data collection experiments carried out in this study. The horizontal dashed line-arrow through the sample holder through to the beam stop represents the path of the primary X-ray beam. Note this is not represented to scale but designed for clarity. Distance between the sample holder windows and adjacent components along the primary beam path were minimised. Sample to detector distance was 470 mm.

Within the arrangement was incorporated a small prism assembly, aligned with an electronic digital moving image camera. By means of an electric motor attached to the arm supporting the prism, this could be moved remotely into a position between the sample holder and the X-ray micro-source. A magnified image of the region of the sample/holder through which the X-ray beam would pass could then be relayed to video screens both within the hutch and on the beam line's workstation outside the hutch. Crosshairs drawn on these screens allowed the precise location of the beam's strike point to be ascertained. Using this, by adjusting the *x*- and *y*-positions of the goniometer head via electric servo motors operable from both within the hutch and at

the workstation, the MTT sample was located. The coordinates of regions of the tissue for study were then identified.

5.2.4 Data collection

All control of the experiment and data collection was achieved at the workstation outside the experimental hutch of beam line 18-ID, using the BluIce-EPICS suite of software (see: www.gmca.anl.gov/ccp4_2009_06/ccp4_bluice_overview.pdf, accessed August 2013). The coordinates of regions of the sample to be "shot" with the beam as defined by relative horizontal and vertical (*x* and *y*) goniometer position in millimetres, was determined using the prism and camera arrangement described in section 5.2.3. Initially, test shots were carried out by aiming the beam at sites roughly along the centre of the fibre under study. Before any shots were carried out, the prism was retracted fully from between the sample holder and the X-ray source tube to avoid blocking of the beam. Diffraction patterns from the test shots were displayed on the workstation screen immediately after detection. Where pixels within the detector were over-saturated with incident scattered X-ray photons, two variables could be adjusted to rectify this and avoid pixel over-saturation and potentially even, detector damage. These were exposure time and also the positioning of attenuating filters within the flight chamber (not shown in Figure 5.6, for clarity) to reduce the amount of scattered radiation from the sample reaching the detector. For the former, exposure times between 1 and 5 seconds were used. The attenuating filters were sheets of aluminium 1, 2, 4, 8 and 16 μm in thickness. Each was independently adjustable via switches on a control box.

For each sample, using the predetermined start and end coordinates along the axis of the fibre, an automated scan was set up using discrete step-distances between shooting sites. Step distances were at least 0.1 mm, chosen with regard to beam diameter because of the localised tissue damage that can occur due to the irradiation at the point of incidence. Prior to the start of each scan, a dark background reading was carried out, and this was automatically used to correct the sample data. For each scan, the per-shot exposure time remained constant.

All patterns detected were saved under a uniquely identifiable sequentially numbered file name, using the *TIFF* computer image format.

5.2.5 Data extraction and analysis

5.2.5.1 Extraction of numerical data

All MTT meridional reflections from diffraction patterns were translated into peak height plots using the FIT2D software (Hammersley, 1997). The transference is described more graphically in section 5.3.1. In brief, it was performed under the "POWDER DIFFRACTION (2D)" option. The "CAKE" choice on the menu allowed data to be extracted through use of sectors. The sector origin was defined as the exact centre of the diffraction pattern, where the primary beam would have struck the detector in the absence of the beam stop. The sector arc was placed as far as was possible from the origin without any portion of it being outside the diffraction image. The radii were positioned at as acute an angle as possible relative to each other, whilst still including all the pattern components required. Any random, anomalously bright pixels were masked. Integration of the sector was carried out along a curved line parallel to the outer arc with azimuthal binning set to 1. Therefore all reflections were integrated into normal distribution Gaussian peaks on a 1D plot of reflection intensity v distance from centre.

All plots were exported as numerical data in the form of *.CHI* files and were converted to spreadsheet format using the OpenOffice.org v. 3.3 spreadsheet software.

5.2.5.2 Peak height and amplitude extraction

The extracted data for individual diffraction patterns were transferred to spreadsheets in OriginPro v.8.5 graphing software (OriginLab Corportation, Northampton, MA, USA). For each group, data from all individual traces were plotted on the same axes and inspected for anomalies and spurious noise spikes, the data for which were removed on the spreadsheet. An average trace was then produced. As each pattern normally produced two useable sets of meridional reflections on opposites sides of the centre, two average traces could be produced for each scan.

For each average trace, the positions of the reflection peaks to be used were determined, for as many consecutive orders as was deemed measurable for all three groups, i.e. native and the two derivatives. The data for each peak were isolated and plotted in turn and the baseline was levelled by straight line subtraction. Appropriate marker points defining the peak data were placed on either side of the peak with due regard to baseline quality and a Gaussian curve fitting algorithm within the OriginPro

software, applied. This gave a Gaussian maximum peak height, which was recorded. Structure factor amplitude, or the square root of the intensity maxima or peak heights, is the accepted measure of reflection intensity (Glusker et al., 1994) and was calculated for all datasets.

In order to produce an average set of amplitudes for each group, datasets of amplitudes were $\log_{10}$-scaled against one another using linear least squares analysis and then the means were calculated. The means for each of the two derivatives' amplitude datasets were then $\log_{10}$-scaled relative to the untreated MTT data to compensate for differences in sample thickness and variations in the experimental conditions employed to allow for this.

Certain corrections can be applied which improve the comparability of the peak amplitudes. Background correction was performed automatically during data collection. The other correction that was applied was the Lorentz correction. This corrects for increasing loss of peak intensity at higher orders due to the diffracted X-rays remaining longer in the diffracted condition prior to detection, as well as peak spreading and loss of peak shape due to the increasing angle of incidence of the diffracted radiation with the detector surface. A series of correcting "Lorentz factors", as previously applied for both types I (Orgel et al., 2000; Orgel, 2000) and II (Antipova, 2008) collagens was used to effect this correction. These had been calculated previously as described by Orgel (2000) using a computer simulation of a series of meridional orders, each with actual intensities set to unity and distributed as a Gaussian function. The resulting factors are based on the predicted diminution of the meridional intensities as sampled and these are presented in Table 5.1.

Meridional reflection order	Lorentz correction factor
1	1
2	1.027869
3	1.057334
4	1.08854
5	1.126193
6	1.163986
7	1.204403
8	1.247729
9	1.294288
10	1.352642
11	1.38331
12	1.4154
13	1.449014
14	1.484264
15	1.531761
16	1.570653
17	1.61157
18	1.654677
19	1.700152
20	1.762065
21	1.774901
22	1.787926
23	1.801143
24	1.814557
25	1.843342
26	1.859032
27	1.874991
28	1.891227
29	1.907746
30	1.941376

Table 5.1 Lorentz correction factors used to correct amplitude values.

5.2.5.3 Patterson plots

Patterson plots were produced according to the one-dimensional Patterson function described in section 5.1.1.1 using a specially set up template in Excel, kindly provided by Dr Orgel. All Patterson plots were in one dimension along the MTT fibrillar axis. The amplitudes of the cosine wave functions for each reflection or order, all initially equal, were multiplied by the squared amplitude (intensity) for their respective orders (Bradshaw et al., 1989; Orgel, 2000), as determined above. The wave functions were then added and plotted using Excel, giving the Patterson function plot (Glusker et al., 1994). All Patterson maps were plotted over 100 points along the abscissa.

Iodine and gold difference Patterson plots, were produced based on the rationale expressed in Equation 3 given in section 5.1.1.2. Prior to the calculation, it was necessary to correct for the non-comparable arbitrary nature of the amplitude units due to differences in the separate samples used. This was effected using a logarithmic scaling function similar to that used previously (Antipova, 2008), employing the following formulae:-

$$|F_{l\,Deriv,\,scaled}| = (10^{(((\mathrm{Log}_{10}|F_{l\,Deriv.}|).a)+b)}).c \quad - \quad \textit{Equation 5}$$

$$|F_{l\,Native\,scaled}| = (10^{(((\mathrm{Log}_{10}|F_{l\,Native.}|).a)+b)}) \quad - \quad \textit{Equation 6}$$

where $|F_{l\ Deriv.}|$ and $|F_{l\ Native}|$ were the structure factor amplitudes for order l for the isomorphically replaced derivative collagen and native, non-isomorphically replaced collagen, respectively and a, b and c are scaling factors. The c factor was necessary to achieve linear scaling to equivalence during the logarithmic scaling process, and was applied solely to the derivative amplitudes. The other scaling factors were each the same value for the derivative and native, untreated collagen and for all reflections. By adjusting the factors, the shape of the difference Patterson plots could be adjusted. The aim was for them to resemble as closely as possible, the predicted Patterson plots for iodine and gold based on the relative locations of the binding sites of the two atoms modelled on the simulated MTT fibril, synthesised as described in Chapter 2, section 2.5.

The predicted Patterson plots were produced from the simulated fibril for both derivatives as follows: First, the positions of the amino acids to which iodine and gold bind – tyrosine and histidine, and methionine and histidine, respectively - were located along the simulated MTT fibril using a logical algorithm in Excel. Every amino acid

position that contained such a binding site anywhere across the fibril was given a numerical value corresponding to the number of these. Using the OriginPro software, separate histograms were plotted representing one *D*-period showing all the anticipated heavy atom electron scattering sites for each element assuming unvarying amino acid position translation distance along collagen molecules. Taking into account the resolution likely to be achieved with the number of orders that were used, from the above histograms, every possible vector that could exist between the electron dense heavy atom-containing positions or regions were calculated. For each, the mean electron density or scattering power between the two components was calculated and the *x*-axis vector distance for the Patterson, determined. These were plotted as a bar graph, and then as a histogram, which was conducive to smoothing using the Savitsky-Golay algorithm. The predicted Patterson function was used as a model for comparison for the experimental plot.

A second means of producing a predicted heavy atom Patterson was employed, using a Fourier algorithm to convert predicted electron density data in real space, into structure factors, i.e. amplitude and phase. This was performed using an Excel spreadsheet, set up and kindly provided by Dr Orgel (See Orgel (2000)), which was adapted for use as required. Briefly, it allowed points of modelled relative electron density values along a unit cell (*D*-period) to be input and superimposed onto real and imaginary space (cosine and sine wave) functions for each order, l. The squares of these resultant predicted real and imaginary amplitudes were added to give the predicted intensities for each order and then the square root calculated, thus:-

$$|Fh_{l\,pred}| = [|Fh_{l\,pred,\,Real}|^2 + |Fh_{l\,pred,\,Im.}|^2]^{½} \quad - \quad \textit{Equation 7}$$

where $|Fh_{l\ pred}|$ is the predicted amplitude of order l due to the heavy atom, h's distribution, with $|Fh_{l\,pred,\,Real}|$ and $|Fh_{l\,pred,\,Im.}|$ being the predicted amplitudes for the real and imaginary components. For the predicted heavy atom difference Patterson plots produced using this method, numbers corresponding to the respective heavy atom electron density, taken as being the atomic numbers of the elements (53 for iodine, 79 for gold) were placed in modelled binding positions within the real space input component of the spreadsheet. A negligible number was entered in non-binding positions. The value 0.01 was considered reasonable. The $|Fh_{l\ pred,\ Real}|$ values generated were squared and used as with experimental data, described above, to produce a predicted Patterson based on a Fourier synthesis.

5.2.5.4 Phase determination

Using the methodology in section 5.2.5.3 gave a narrow range of scaling factors that produced difference Patterson plots for iodine and gold, most closely resembling those predicted. The $|F|$ values as used in these plots were first adjusted to be equivalent, by multiplying the gold isomorphically-replaced data appropriately by the ratios of the relatively-scaled, native, non-derivatized MTT values used in the iodine and gold plots, for each order. The data were then transferred into the relevant (native, iodine- and gold-derivatized) data fields of the PGPlot software that had been customised for phase determination.

The software was used to produce Harker diagrams (see section 5.1.1.2 and Figure 5.2) for each order. To determine the phase suggested by each successive diagram, where the three circles intersected at or close to a point, by moving the cursor over the point and clicking with the left mouse key, the phase angle in radians relative to the 0 or 2π radians position in real space for the native, non-isomorphically replaced data circle, was calculated.

As no set of scaling factors used produced a usable three-way intersect for all orders, different combinations of the "best" scaling factors were used from the difference Pattersons for iodine and gold. For each, three values for scaling factor a and two values for factor c were used, with b remaining constant. The resultant six datasets for iodine and gold were combined in every possible combination, resulting in 36 sets of phase values being produced. These phase values were scored according to the closeness of intersects:

1. Perfect or near-perfect.

2. Strong suggestion of a narrow range of phase value, but with less certainty.

3. No three-way intersect, but still an apparent probability of phases within a relatively small range.

There were Harker diagrams which implied two possible phases and these were noted and scored as above. Finally, there were diagrams that gave no indication of any phase probability, and these were disregarded.

The phases were tabulated using an Excel spreadsheet, colour coding the scores for clarity. For each order in turn, the phase data were transferred to a spreadsheet in OriginPro and histograms plotted showing the distribution of the phase probability values on an *x*-axis scale of 0 to 2π radians. For most, three histograms were created

containing data from **1**, **1**+**2** and **1**+**2**+**3** above. These were compared and the most likely phase probability determined. Some orders were ambiguous or equivocal between two or more possibilities. In these cases, the Harker data that had produced two possibilities of phases were taken into account to further try to narrow the most probable phase down. Where there remained ambiguity, both phases were recorded.

5.2.5.5 Electron density model building

An Excel spreadsheet, previously designed and set up to produce electron density plots from amplitude and phase data determined for the meridional reflections, was kindly donated by Dr Orgel. Some adjustments were made to render it suitable for calculating a one-dimensional electron density plot for MTT to the requisite number of orders' resolution. The electron density plot was created using a Fourier synthesis, carried out as described by Equation 4 (section 5.1.2) and in real time, such that changing amplitude or phase data for any order would produce an immediate update of the trace. Each structure factor's data were entered into the spreadsheet in turn. As mouse tail type I collagen structure was anticipated as being very similar to that in RTT, a RTT collagen electron density map was constructed to the same number resolution using data used in previously published work (Orgel et al., 2000) and used as a reference. During construction, the rat collagen map was aligned with the mouse data to best correlation using the Pearson r linear fit, and the *r*-value monitored with the addition of amplitude and phase data for each successive order. Where the choice of phase was less clear, the correlation influenced the choice of phase used. The effect on the mouse collagen map was also observed: the phase that contributed to the improved resolution of the map already in place, e.g. due to existing peak sharpening and augmentation, was favoured.

A predicted electron density map for MTT collagen was produced using the simulated fibril described in Chapter 2, section 2.5. Numbers of electrons per amino acid residue in an aqueous environment (Hulmes et al., 1977) and amino acid residue volumes (Goldsack and Chalifoux, 1973; Chothia, 1975) were used to calculate electron density for each amino acid and the total electron density due to the collagen across the fibril was plotted for each amino acid position. The amino acid translation distances along the molecules was taken as being constant. Using this, the similarities and correlation between experimental and simulated *D*-periods could be assessed and measured.

5.3 Results

5.3.1 Diffraction patterns

A diffraction pattern collected for non-isomorphically replaced MTT collagen is shown in Figure 5.7. It can be seen clearly that the meridional reflections are sharp and clear and well defined against the background. Studying the first 30 orders, the stronger reflections – i.e. orders that are notably stronger than those surrounding them, so 1, 3, 5, 9, 12, 20, 21, (22), 25, 27 and 30 – were found to be the same as identified previously in mouse (McBride Jr et al., 1997) and rat (Brodsky et al., 1982). The equatorial reflections are much more diffuse and indistinct. Throughout these experiments, the equatorial reflections were either as shown or more often, absent altogether.

Meridional diffraction patterns for non-isomorphically replaced, iodine- and gold-isomorphically replaced native MTT are shown in Figure 5.8. Comparing the heavy atom derivative tendon fibre sample patterns with those of the untreated samples, differences between the gold-treated collagen pattern and the native are evident with respect to their relative intensities. In this case also, arc widening of each reflection can also be seen. This was a manifestation of changes in the mechanical behaviour of gold-treated collagen fibres noted during the experiments, i.e. they became “springier” along their length, less likely to hang straight under the force of gravity and trickier to reliably hold under tension in the sample holders. As the arcs were integrated along their curved lengths (section 5.2.5.1), this was not considered to be a problem in this respect. In contrast, the iodine treated tendon fibre patterns showed little difference from the native ones, neither with respect to relative intensities of the reflections nor the lengths of the reflection arcs.

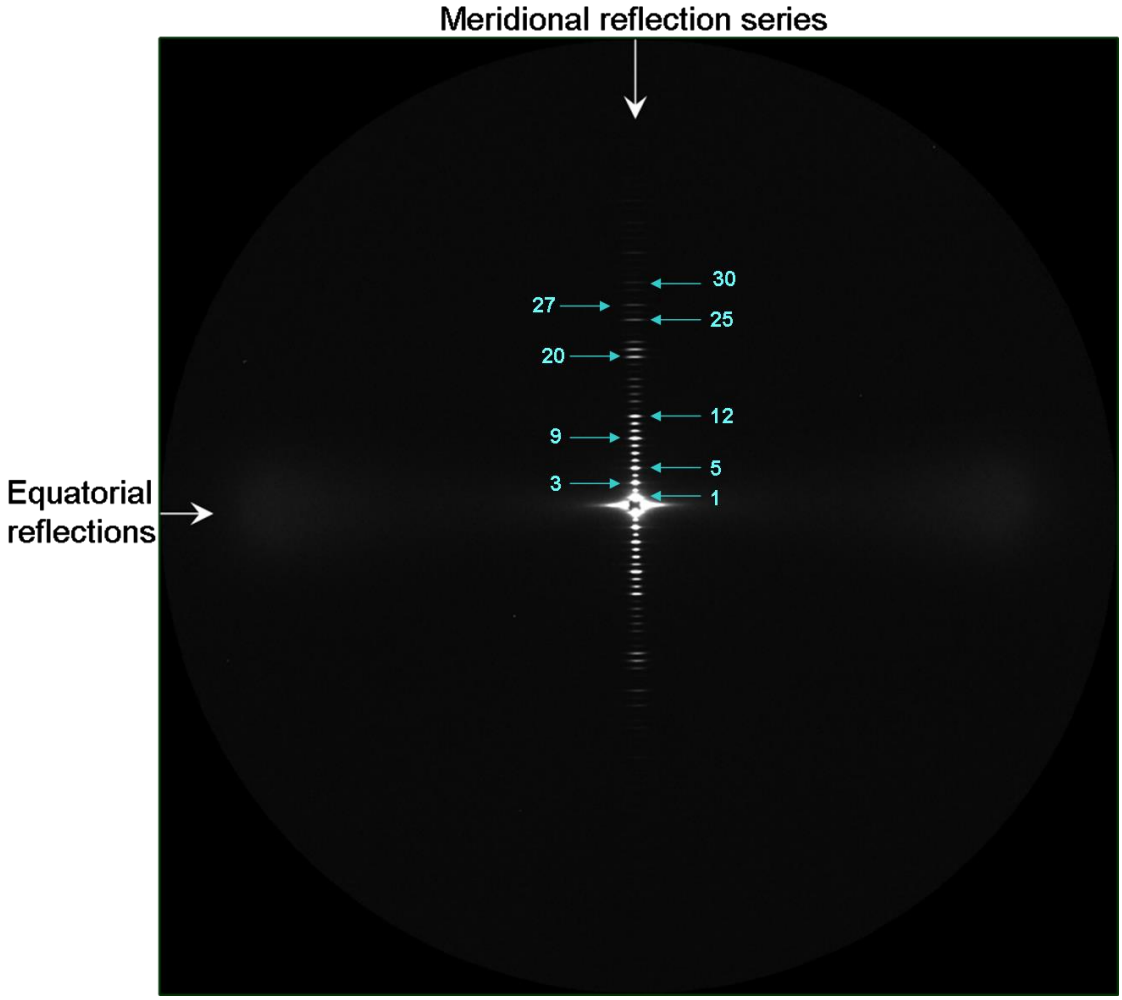


Figure 5.7 Example of an entire native, non-isomorphically replaced X-ray diffraction pattern collected at APS. The axes along which the meridional and equatorial reflections are located are indicated. The meridional reflections lie along fibrillar axis. Most of the stronger reflections are labelled with their numbers and arrows, in the upper half. In reality, the pattern was collected using the arrangements described in sections 5.2.2 and 5.2.3, with the fibres held horizontally. The pattern shown has been rotated 90^{o}, to appear as per generally accepted convention. The circular, dark grey area represents the entire detector surface.

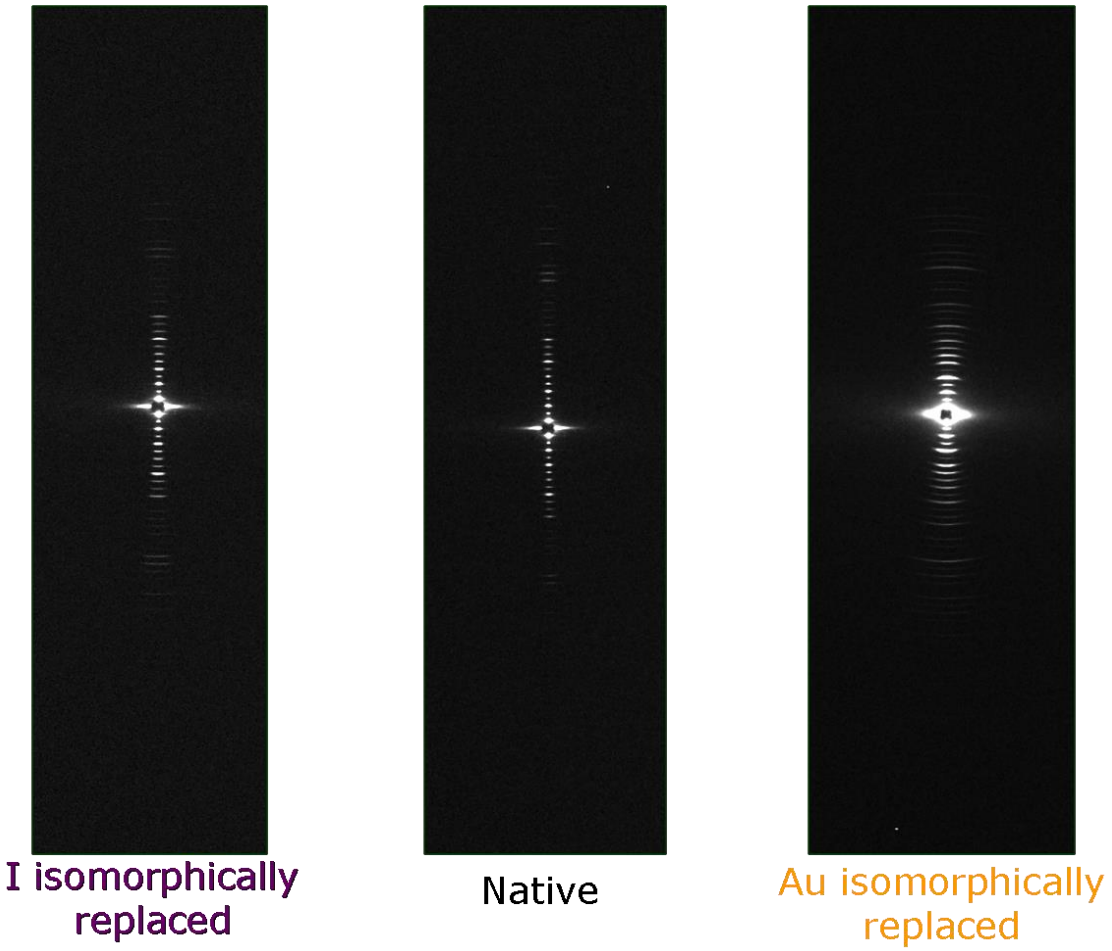


Figure 5.8 Meriodional X-ray diffraction patterns for iodine- and gold-isomorphically replaced MTT collagen, with the untreated native pattern in the centre for comparison. The equatorial reflections and much of the extraneous detector surface area has been excluded from each image in order to emphasise the meridional reflections.

5.3.2 Extracted peak heights and amplitudes

Examples of FIT2D-extracted data from the native MTT diffraction patterns and the iodine- and gold-isomorphically replaced ones, are given in Figure 5.9 and Figure 5.10. In Figure 5.9, the linear and semi-log plots of the extracted data are shown. On a linear scale, due to the falloff of the intensity, when presented on the same scale as the first order, the orders are barely visually discernable beyond order 9. On the semi-log plot, reflection peaks are discernable on the same axes as the 1st order, out to the 30th and beyond and provides a clearer means of comparison between this and other sets of data (Antipova, 2008). Averaging was effective in reducing the remaining background noise level.

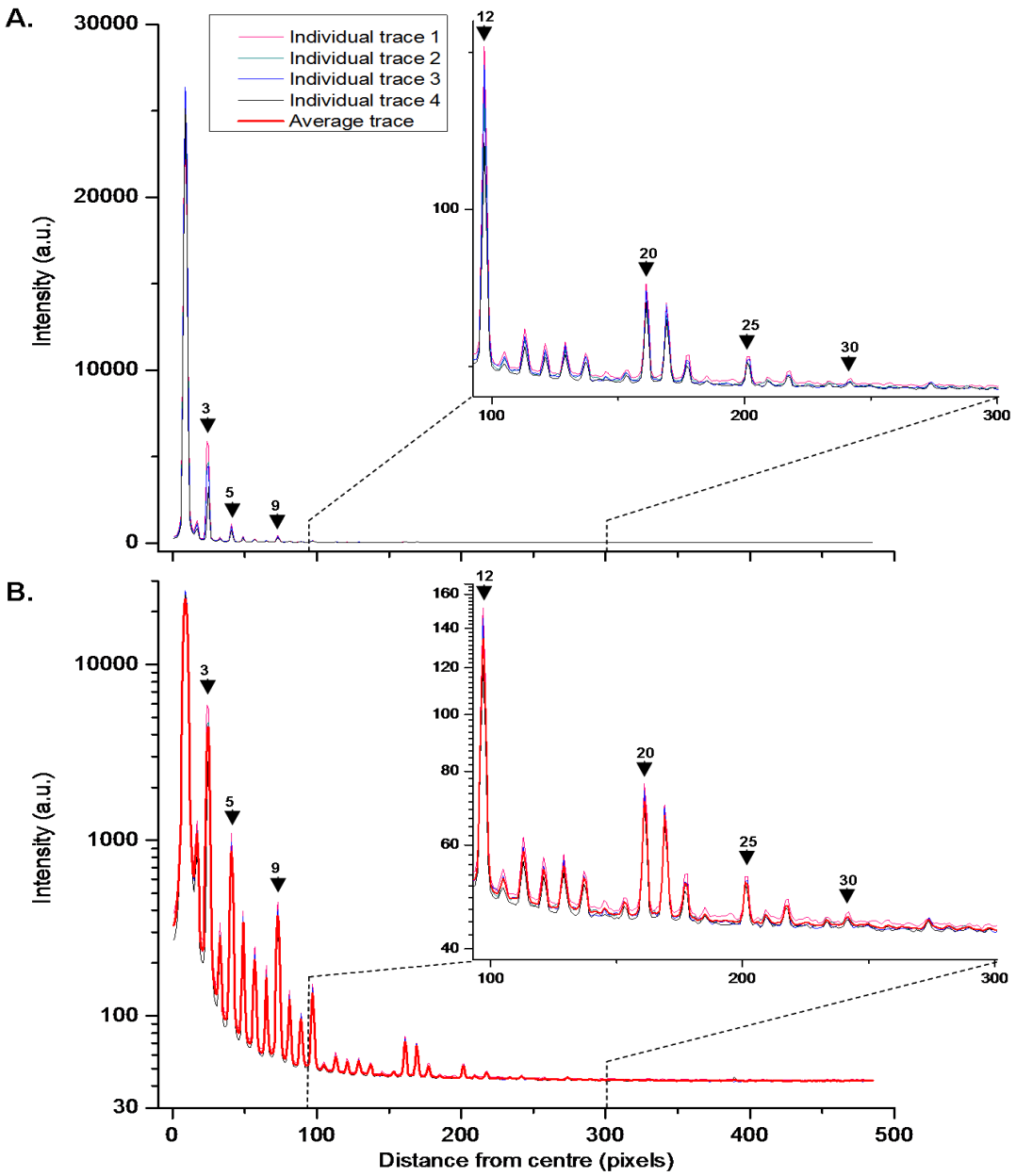


Figure 5.9 Data plots of reflection intensity data for untreated native MTT. Plot A shows the data for individual traces only (n=4), on a linear scale. Plot B shows the individual traces again on a semi-log plot for better representation of the smaller higher orders, with the average represented as a thicker red line. The inset graphs show magnifications of orders 12 to 30, and beyond. Black arrows with numbers identify the more prominent orders. The data was taken from reflections recorded on the left side of the detector surface, no filters, with 1 second data collection time for each individual trace.

In Figure 5.10, average intensity data plots extracted from X-ray diffraction patterns from untreated MTT are compared with plots from iodine- and gold-treated MTT. Visual inspection reveals that the data plots from both sides of the detector produced plots and relative peak heights that were close. The native MTT plot cannot easily be distinguished from the iodine treated fibres on the basis of relative peak heights. Conversely, the differences between the untreated and Au-treated MTT plots are obvious. Compared to the untreated, Au isomorphous replacement appeared to have resulted in a relative increase in the intensities of 15 orders (5, 7, 8, 11, 13, 15, 17, 18, 20, 22, 23, 26, 27, 28, 30), a decrease in 6 orders (6, 9, 10, 16, 21, 29), with 8 appearing to remain the same (2, 3, 4, 12, 14, 19, 24, 25).

Using OriginPro, peak heights were calculated from a fitted Gaussian curve. Each fit was checked “by eye”, to ensure that it was sensible and that any interference from background noise was minimised. This was most important for the smallest reflections. It was determined that peaks could be measured with reasonable accuracy between orders 1 and 30. Beyond this, although some peaks could be located, the noise:peak ratio was considered to be too high.

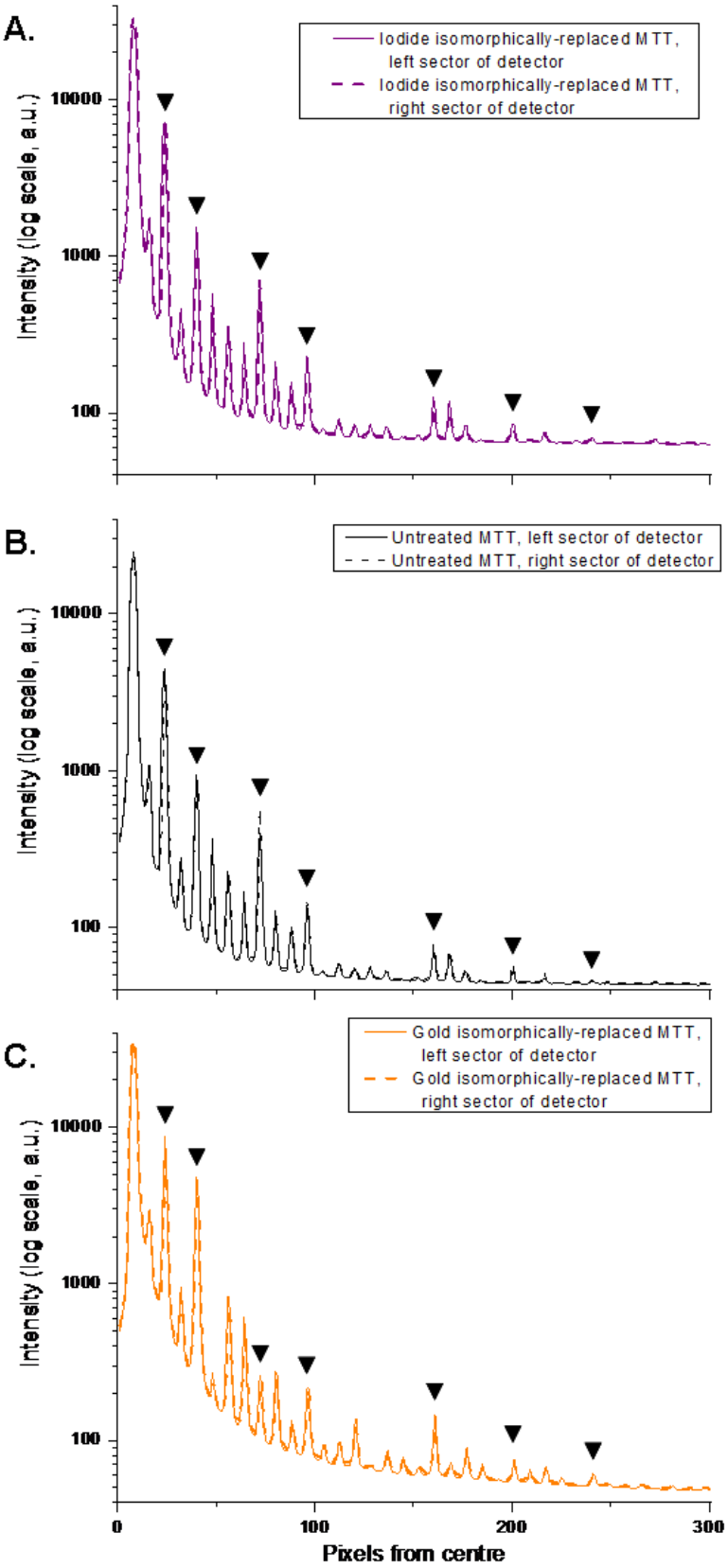

A.
Iodide isomorphically-replaced MTT, left sector of detector
Iodide isomorphically-replaced MTT, right sector of detector
Intensity (log scale, a.u.)
10000
1000
100
B.
Untreated MTT, left sector of detector
Untreated MTT, right sector of detector
Intensity (log scale, a.u.)
10000
1000
100
C.
Gold isomorphically-replaced MTT, left sector of detector
Gold isomorphically-replaced MTT, right sector of detector
Intensity (log scale, a.u.)
10000
1000
100
0
100
200
300
Pixels from centre

Figure 5.10 (Previous page) Comparison of native, untreated MTT semi-$\log_{10}$ intensity plot (**B**) with those isomorphically replaced by treatment with 1 mg/ml KI (**A**) and 1 mg/ml $NaAuCl_4$ (**C**). The untreated MTT data was placed centrally, for easy comparison with both isomorphous derivative plots. For each, two average plot lines are shown, one (solid) for the left side of the detector and the other (dashed) for the right. For each line on (**A**), n = 4; on (**B**), n = 5; on (**C**), n = 5. Visually, it is difficult to discern an appreciable difference between the untreated and iodinated, but there are clear differences in relative peak heights between non-treated and Au-treated. For ease of comparison, the black arrowheads point to the peaks for orders 3, 5, 9, 12, 20, 25 and 30, left to right. All data recorded using no filters, with a 1 second data collection time for each pattern used.

For each scan, average plots were taken for the left and right side of the detector and for these, peak heights for orders 1 to 30 were determined. For all three groups, data sets with and without filters were used. Peak heights were measured for orders 1 to 12 from the data obtained with the appropriate aluminium filters in place, and for orders 6 to 30 without. The overlap – i.e. orders 6 to 12 – were plotted in each case to produce a linear scaling factor, which was then used to scale the filtered data relative to the unfiltered higher orders. Orders 1 to 5 were derived from the filtered data in this way and the more sensitive unfiltered data was used for the measurement of orders 6 to 30. For this portion of the study, the two sets of peak heights were converted to $\log_{10}$ values, plotted against one another and then $\log_{10}$-scaled against each other on the basis of a linear fit achieved using a least squares algorithm. The duplicate averages were square-rooted to produce the peak amplitudes and multiplied by the factors given in Table 5.1, for Lorentz correction. The averages of these were plotted with error bars based on the duplicates, to give an indication of the relative errors associated with peak height extraction from each pattern (Figure 5.11).

The total of the coefficients of variation (CVs) (SD/mean) were 2.025 for untreated MTT, 1.340 for iodinated MTT and 0.779 for the gold treated, suggesting that peak height extraction from the heavy atom treatment patterns is more precise than for the untreated tissue. This agreed with the experience obtained from performing this step and is evident from examination and comparison of the plots.

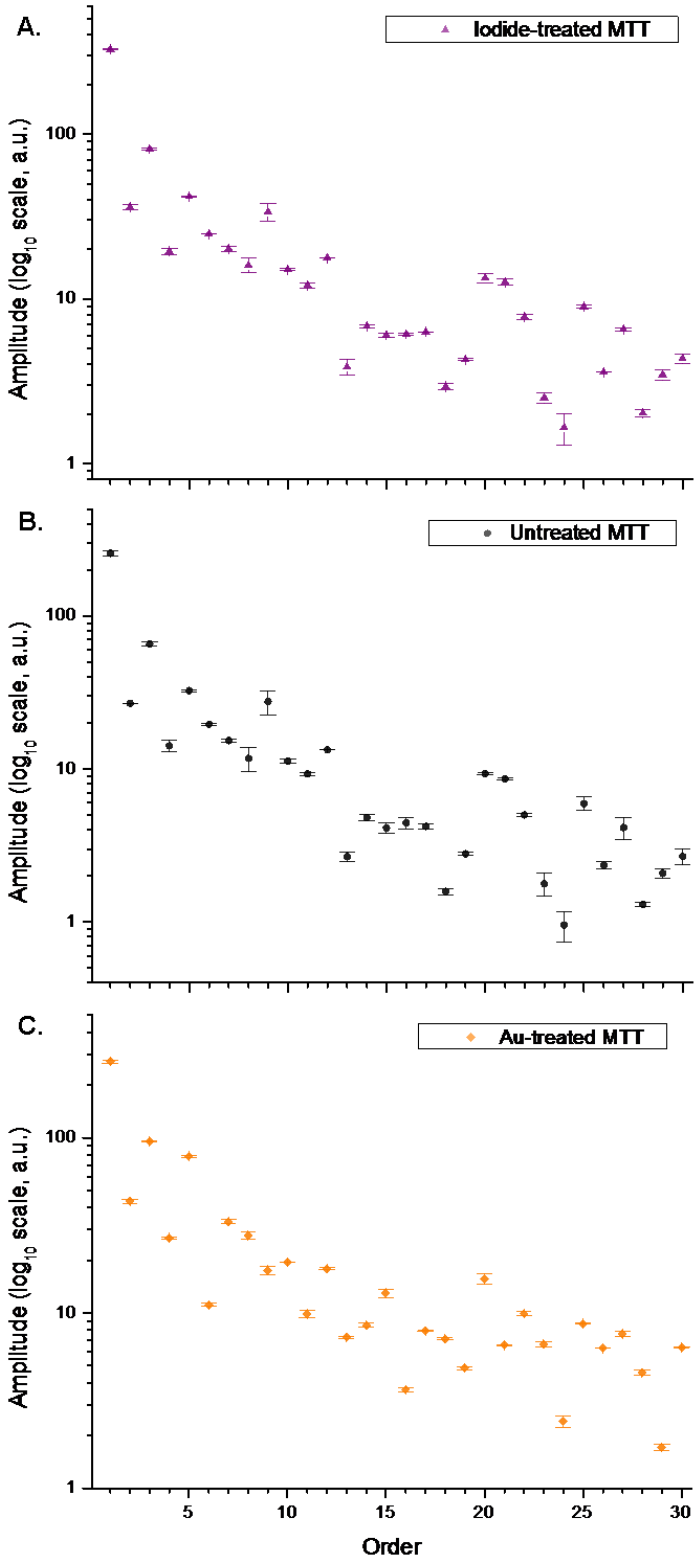


Figure 5.11 Plots of the mean Lorentz corrected amplitudes for native MTT for orders 1 to 30: **A.** iodine treated; **B.** non-isomorphically replaced; **C.** gold treated. Error bars represent ±1 SD based on left and right side of detector data extraction comparison.

Order	Untreated MTT amplitudes - Lorentz corrected	Iodine-treated MTT amplitudes - log-scaled and Lorentz corrected	Au-treated MTT amplitudes - log-scaled and Lorentz corrected
1	257.8	288.9	184.5
2	26.8	27.7	27.1
3	65.4	65.5	61.5
4	14.2	14.3	16.3
5	32.5	32.1	50.1
6	19.5	18.5	6.50
7	15.3	14.7	20.4
8	11.7	11.6	16.8
9	27.5	25.5	10.4
10	11.3	10.8	11.7
11	9.28	8.51	5.70
12	13.4	12.8	10.6
13	2.67	2.53	4.13
14	4.80	4.63	4.87
15	4.12	4.05	7.57
16	4.44	4.11	2.01
17	4.20	4.23	4.50
18	1.58	1.88	4.02
19	2.78	2.80	2.69
20	9.30	9.38	9.16
21	8.62	8.82	3.68
22	4.99	5.27	5.68
23	1.78	1.59	3.73
24	0.95	1.02	1.29
25	5.93	6.10	4.93
26	2.34	2.31	3.51
27	4.14	4.35	4.29
28	1.30	1.26	2.52
29	2.08	2.22	0.90
30	2.68	2.82	3.55

Table 5.2 Lorentz-corrected amplitudes for untreated, native MTT, plus MTT fibres isomorphically replaced with iodine and gold. The iodine and gold data were $\log_{10}$-scaled against the untreated data.

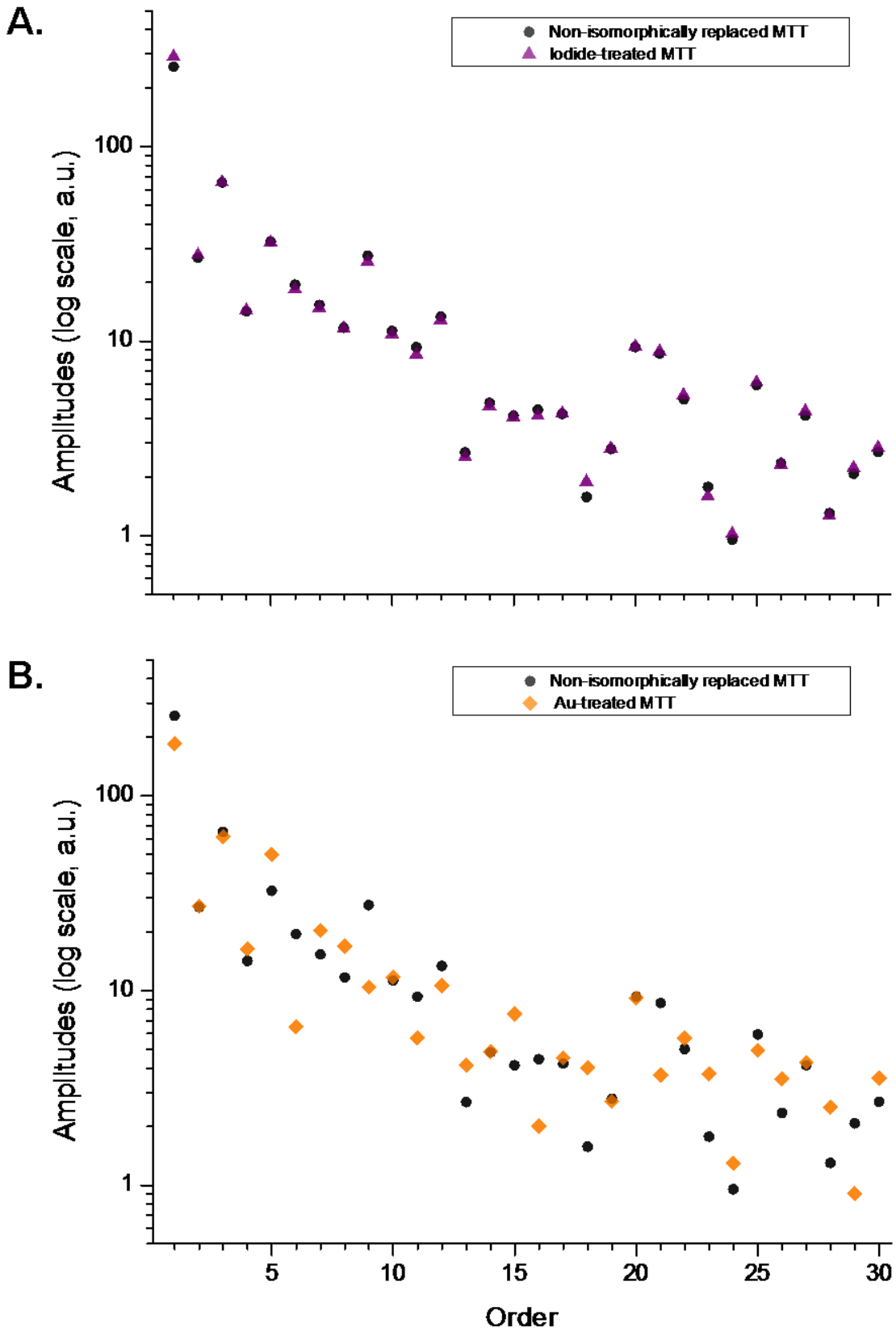


Figure 5.12 Semi-log plots comparing **A.** untreated and iodine- and **B.** untreated and gold-isomorphically replaced MTT Lorentz corrected mean amplitudes for the 30 meridional orders measured. The iodine and gold data were $\log_{10}$-scaled against the untreated MTT data prior to Lorentz correction.

The uncorrected average amplitudes for the iodine and gold treatment groups were log_{10}-scaled against the native to correct for variations between datasets due mainly to non-uniformity in the sample thickness. The resultant values are given in Table 5.2. The heavy atom treatment group amplitudes could then be plotted in turn on the same axes as the untreated MTT data and compared directly (Figure 5.12). As in Figure 5.10, the amplitude values for iodine-treated MTT shows little obvious difference to the untreated MTT group (Log_{10} native v Log_{10} iodine derivative peak heights correlation: Pearson's r = 0.999, R^2 = 0.998). However, differences between the Au-treated and the untreated tissue data are evident (Log_{10} native v Log_{10} gold derivative peak heights correlation: Pearson's r = 0.931, R^2 = 0.863).

5.3.3 Patterson plots

From these intensity (amplitude squared) data, Patterson plots could be constructed as described in section 5.2.5.3 and these are shown in Figure 5.13. Two versions are shown. One set of Patterson plots is constructed from the intensities of all 30 orders. The first order represents vectors between sites of electron density of 0%/100% apart along the unit cell - i.e. distance of any given centre of electron density from itself – which is known information. The first order intensity is also at least two orders of magnitude higher than those for most of the other orders. Therefore the second set of traces shows the Patterson plots excluding the first order to avoid first order "washout" and to show more detail between electron dense centres of interest. Predictably, the untreated and iodine Patterson plots are almost identical, whereas the gold-treated MTT plots both show appreciable differences. These appear to reflect not only changes in relative vector peak heights and therefore electron densities themselves, but also shifts in some electron dense area's spacings. The large untreated peak at 34% (or 66%) appears to have shifted to 39% (or 61%) and conversely, the trough at 16% (or 84%) appears to have shifted to 12% (or 88%) as a result of the gold treatment.

Iodine and gold difference Patterson plots were produced as outlined in section 5.2.5.3, by means of use of logarithmic scaling factors (Antipova, 2008). The locations of the heavy atom scattering sites were determined for iodine and gold, also as described (section 5.2.5.3). All the possible vector spacings reasonably resolvable between these sites were calculated and incorporated into a separate histogram for each of the two elements. The peak heights represent the average number of atoms bound at each locus involved in the vector. Where two or more vectors have the same spacing value, they have been totalled, as the experimental Patterson would also be additive in this

way. These maps and plots are shown in Figure 5.14 and Figure 5.15, for iodine and gold respectively. The predicted Patterson plots, produced from the histogram values using a Savitsky-Golay algorithm, are represented by a black line trace on the histogram.

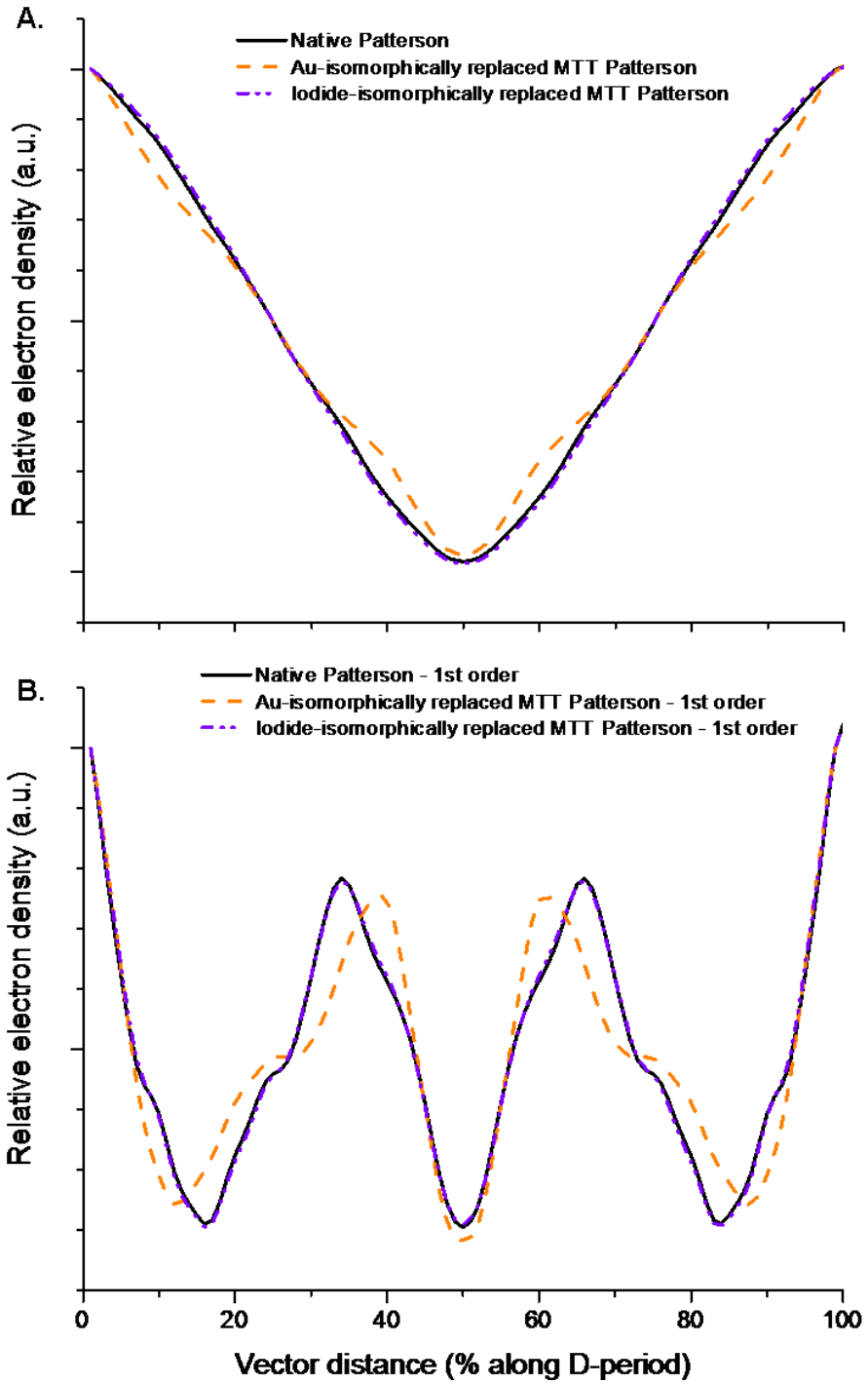


Figure 5.13 Patterson maps for untreated, iodine- and Au-isomorphically replaced MTT collagen. **A.** shows Pattersons constructed using all 30 scaled, corrected amplitudes and **B.** shows the same, with the exclusion of the first order in the Fourier synthesis procedure, to enable more detail to be seen without first order "washout".

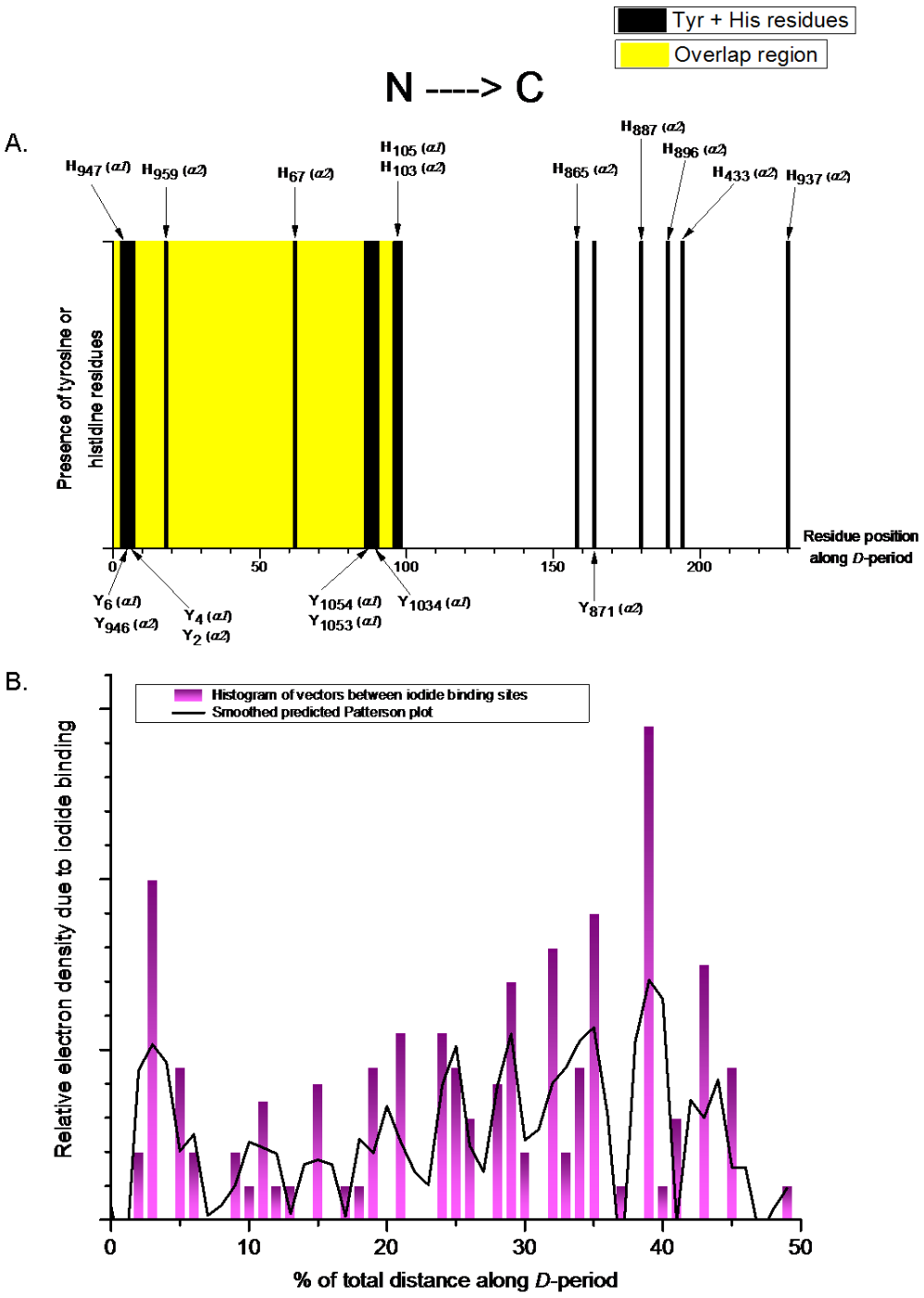


Figure 5.14 **A.** Representation of relative positions of iodine binding sites (tyrosine and histidine residues) along one *D*-period of MTT type I collagen. The identity of each amino acid from the primary sequence is indicated; **B.** histogram of all possible vectors between iodine binging sites, excepting 0% (or 100%), discernable at the resolution achievable using 30 orders of diffraction data (2.3nm). The black line shows a smoothed plot of the histogram using the Savitsky-Golay smoothing algorithm (4 points in smoothing window, 3rd order polynomial) and represents a predicted iodine Patterson plot. As the Patterson is centrosymmetric, only the lower 50% is shown for clarity.

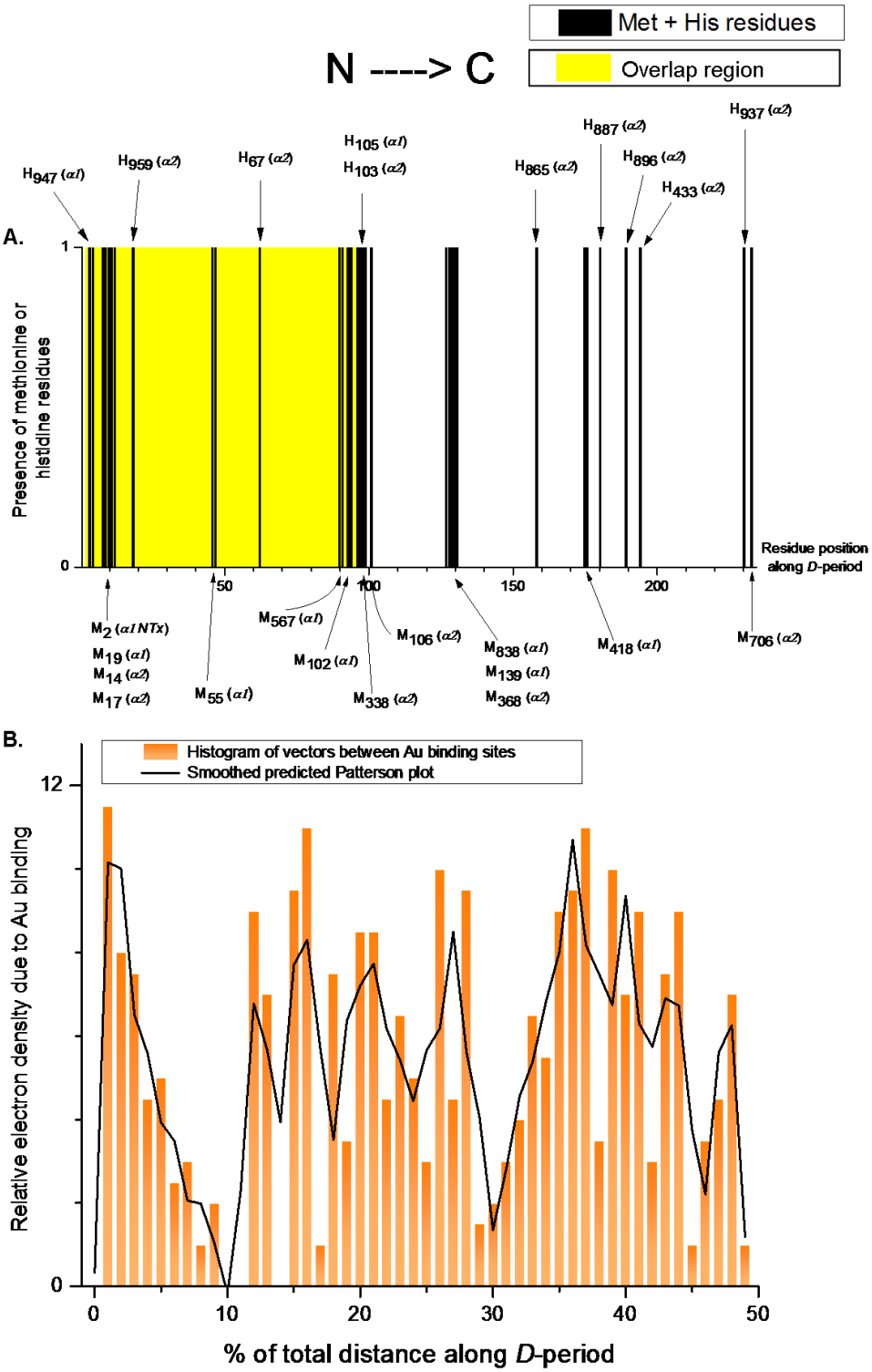


Figure 5.15 **A.** Representation of relative positions of gold binding sites (methionine and histidine residues) along one *D*-period of MTT type I collagen, in similar fashion to Figure **5.14**A; **B.** as with Figure **5.14**B, for the gold binding sites. Smoothing was performed using the same algorithm and parameters as for the predicted iodine Patterson plot.

In addition, a reverse Fourier method previously described in section 5.2.5.3 was used to produce amplitudes (and phases) for the difference Pattersons for each atom, by using the predicted binding sites along the 234 amino acid distance of a *D*-period. The intensities (squared amplitudes) were used to produce alternative predicted Patterson difference plots for each atom, and these are compared with the smoothed histogram Pattersons in Figure 5.16. Broadly, for both atoms, both methods produced very similar plots. Visually, the smoothed histogram method appeared to contain more detail.

For the iodine predicted Patterson plots, there did appear to be some peak position discrepancies between the two methods, especially regarding the peaks at positions 11% and 16% (Fourier method) along the abscissa. However, these only represent discrepancies of 1%, and are down to the effects of using large scaling, i.e. each division represents a full 1% and only half of the plot is shown for clarity to avoid centrosymmetric duplication of the same data. There are also minor discrepancies in relative peak heights between the predictions. As shown, for the iodine, the most notable of these was for the peak at the 36% position (Fourier method). However, the number of peaks and general shape of the predicted plots appeared to be the same for both methods.

For the gold predicted plots, the similarities were also obvious for both methods. In terms of peak-positional discrepancies, the large peak at 16% along the abscissa using the Fourier method, is split into two smaller peaks by using the smoothed histogram method. The two peaks at 42% and 45% using the histogram method are represented by little more than a shoulder on the Fourier 37% peak. Notwithstanding these discrepancies, the most striking peak height difference for the predicted gold difference Patterson plots, is for the peak at 29% (both methods). The peak here is higher and sharper using the histogram method as compared with the Fourier method.

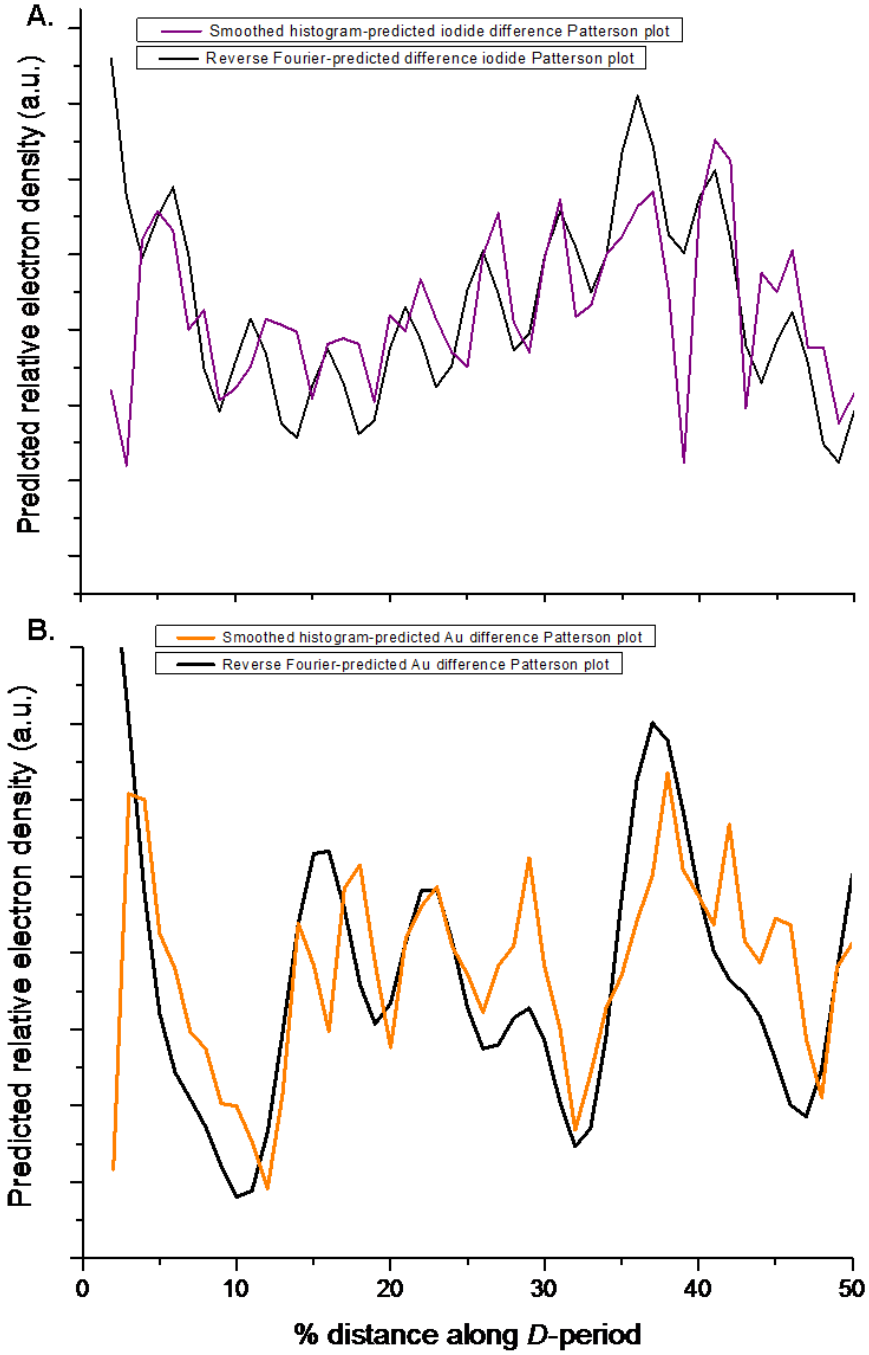


Figure 5.16 Predicted heavy atom difference Patterson plots for **A.** iodine and **B.** gold binding sites along the MTT collagen D-period. The coloured lines – purple for iodine and orange for gold – represent the plots produced using the histogram method and shown in Figures 5.14 and 5.15 respectively, whereas the black lines show the corresponding reverse Fourier plots for each heavy atom.

Finding the correct magnitude of the scaling factors for the native and isomorphically replaced experimental amplitudes was trial and error. Fine tuning was then used to find the experimental heavy atom Patterson plots that most closely resembled the predicted plots using live, scaled difference plots in Excel and comparing to the predicted traces. The comparisons are shown in Figure 5.17 and Figure 5.18 for iodine and gold, respectively.

Despite the extracted iodine-treated MTT amplitude data being difficult to distinguish from the native, (Figure 5.12, Figure 5.13), the iodine difference plot produced shows consistency with the predicted Patterson plots. The iodine Patterson's experimental peaks broadly show better alignment with the Fourier-predicted Patterson plot than with the histogram-predicted trace, but the discrepancy between the predicted traces has been considered and would account for this. All peaks on both predicted traces are directly attributable to experimental traces (Figure 5.17), with the exception of the peak at 31%. In Figure 5.17A, the peak shoulder at 34% seems to exactly align with the experimental left-double peak, so the 31% peak could be taken to represent the smaller peak at 29%. In Figure 5.17B, the Fourier-predicted Patterson lacks this shoulder, however.

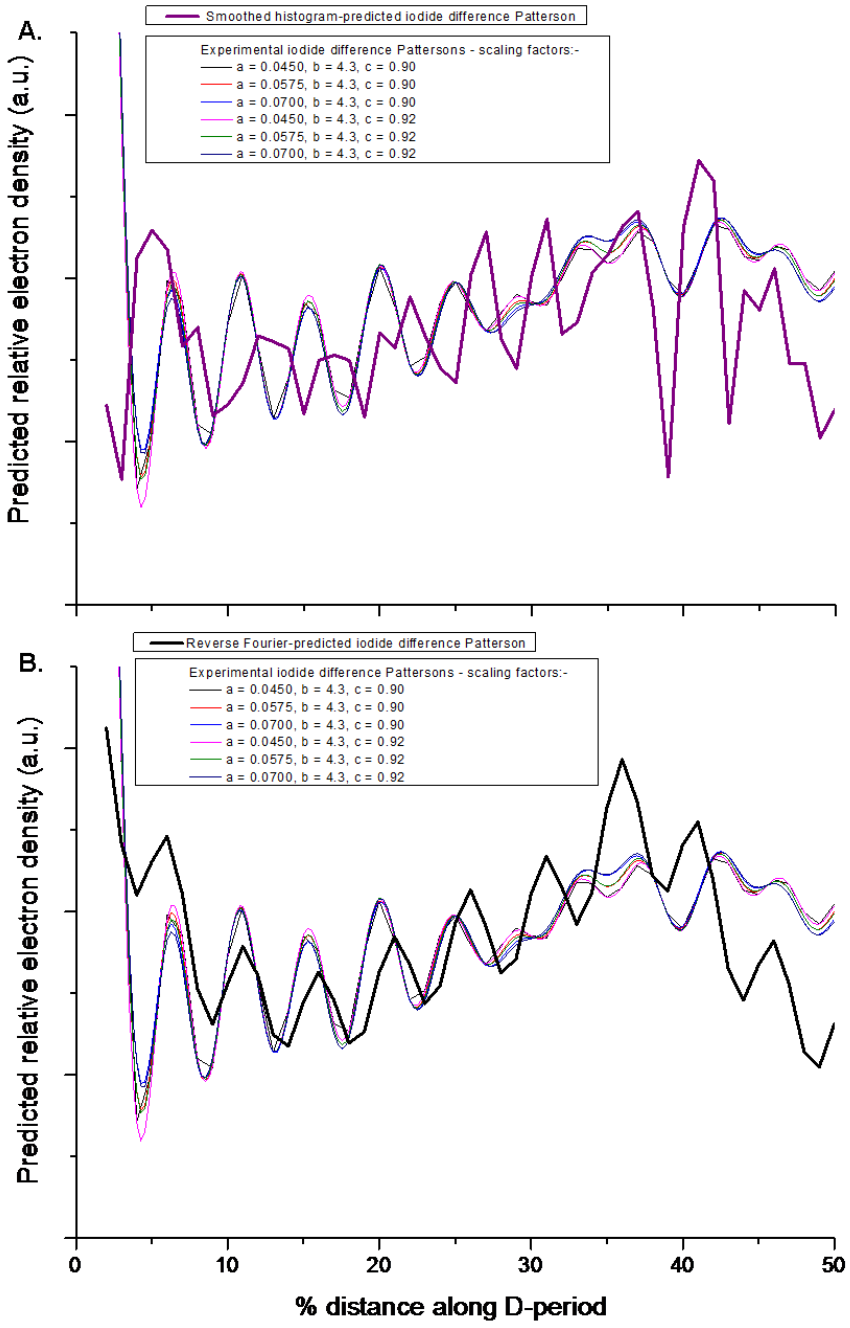


Figure 5.17 Experimental iodine difference Patterson plots, based on subtractions of untreated MTT amplitude data from iodine isomorphically replaced native MTT amplitudes after scaling of both using the scaling factors (a,b,c) given. By comparison with the predicted Patterson plots, the best six sets of scaling factors (insets) for the experimental data were selected. **A.** Scaled experimental traces overlaid with smoothed histogram-predicted Patterson plot, and **B.** the same experimental traces compared with the Fourier-predicted Patterson.

From Figure 5.18, striking similarities can be seen between the best-scaled experimental gold difference Patterson plots and the predicted. However, in order to best illustrate this, both predicted Pattersons had to be shifted 2.8% to the left relative to the experimental traces to avoid a general peak position discrepancy and achieve a good visual alignment. In contrast to the iodine plots, the gold experimental data best seems to best match the histogram-predicted Patterson plot in Figure 5.18A. With the exceptions of the peak at 45% and the half-peak at 50%, all the experimental peaks have a corresponding predicted peak. Interestingly, the 45% and 50% peaks exactly correspond to the positions of predicted peaks here, so are the only ones for which the predicted Patterson does not need to be shifted left, to achieve alignment. These correspond with an Au-native crossover coordinate seen in Figure 5.13B. A further crossover point at 27% might be consistent with a possible peak split discrepancy between the experimental data, explaining why the predicted peak at the 29% position is higher than the experimental peak, i.e. one could assert that the experimental shoulder at the 29% position is a non-shifted component of that predicted peak, and that the shift of the remainder has caused a lowering of the peak height.

The differences between the predicted gold difference Pattersons in Figure 5.18A and B have already been mentioned and unlike for the iodine, the histogram-predicted Patterson plot was unequivocally superior when compared with the Fourier-predicted Patterson, in acting as a model for the experimental scaling factors. Certainly, during the scaling process, there was little indication that a different set of factors could be used to produce a set of traces more closely resembling the Fourier-predicted Patterson.

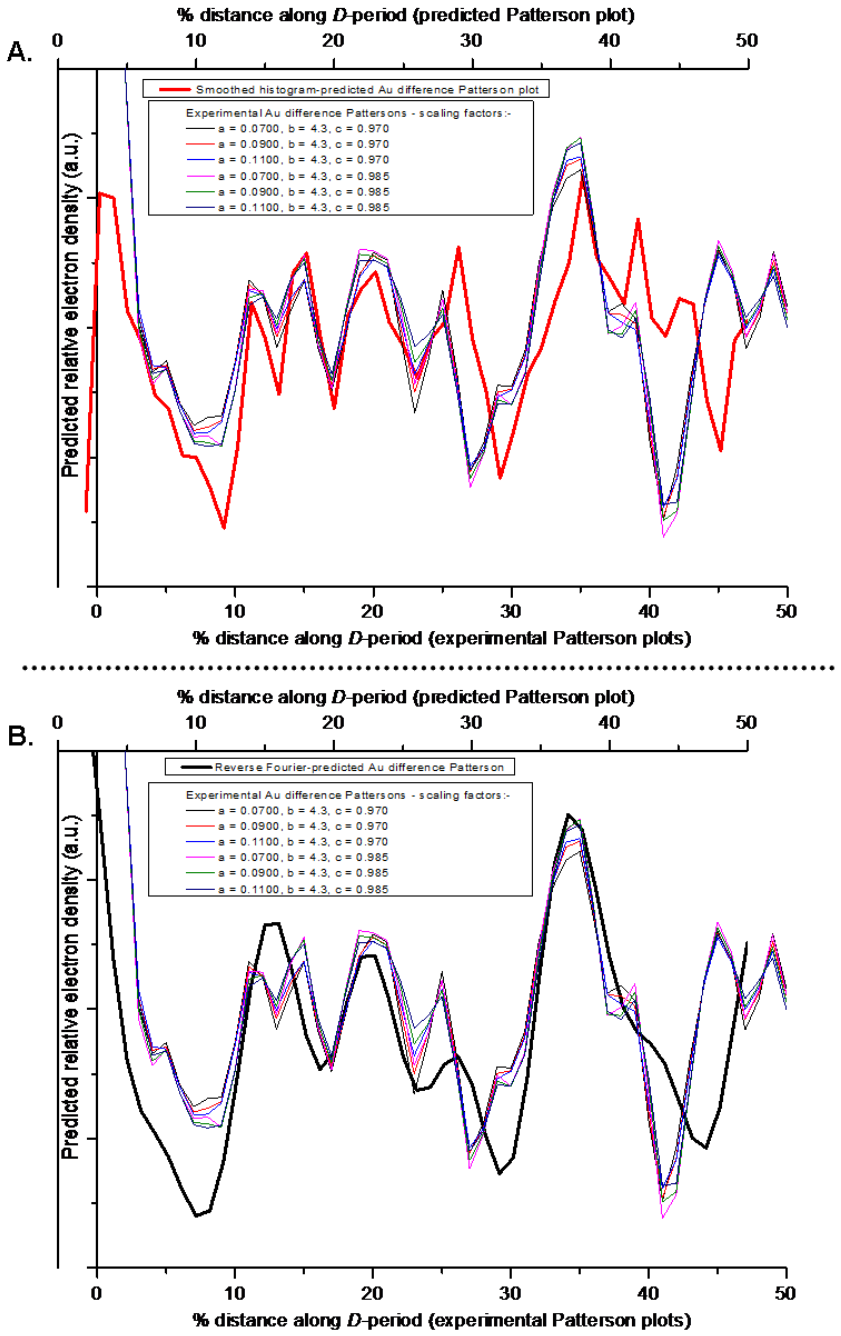


Figure 5.18 Experimental gold difference Patterson plots, analogous to iodine difference data shown in Figure **5.17**. The experimental traces shown were produced using the best sets of scaling factors (insets). **A.** Experimental data compared with smoothed histogram-predicted Patterson plot, and **B.** compared with Fourier-predicted Patterson. As there seemed to be an overall left shift of the experimental data as compared with the predicted in both cases, the predicted Patterson plots have been shifted left by 2.8% to achieve a visual best fit.

5.3.4 Phase determination

As described in section 5.2.5.4, using the Harker plot method programmed into the PGPlot software, the six sets of scaling factors determined for each derivative, above, were used in every possible combination to produce 36 sets of scaled amplitudes and then phase probabilities for the 30 orders. No one combination of scaled amplitudes produced phase probabilities, i.e. close three way intersects of the three circles, for all 30 orders and the degree of success varied with different combinations. All of the *individual* phases determined are shown in Appendix 2.

All of the individual phase probabilities were plotted as bar graphs, to represent phase probability distributions over the range 0 to $\pi/2$ radians. Any individual phases that lay within 0.3 radians (<5% of 2π radians) of one or more others, was considered part of a cluster. Examples of these are presented in Figure 5.19, which shows the plots for orders 1, 12 and 22. Orders 1 and 12 were considered to have strong phase probabilities for single phase means. On consideration of its graphical data, order 22 was considered equivocal between two phase probability means. The phase probability mean that was used was eventually selected, based on producing a better correlation of the in-progress MTT electron density map with the RTT map comparator, as well as it clearly contributing to enhancing and sharpening peaks and features already produced during the progressing map building MTT process. The phase probabilities determined by this method, together with possible alternatives, are given in Table 5.3. The selection of the phase to be used for each order was inextricably linked with the Fourier synthesis of the MTT electron density map (section 5.3.5). However, out of the 30 orders, 23 of the phase probability means chosen for use in the MTT map were selected from the first choice phase probability means (second column in the table). Of these orders, 12 had yielded just one mean value.

Order 11 was a choice between two individual values, only seen on the second tier bar plot, i.e. there were no precise three-way Harker circle intersects. The third plot showed little additional useful information. Therefore the choice of which of the individual values to use had to be made at the map building stage, according to best correlation with the RTT model.

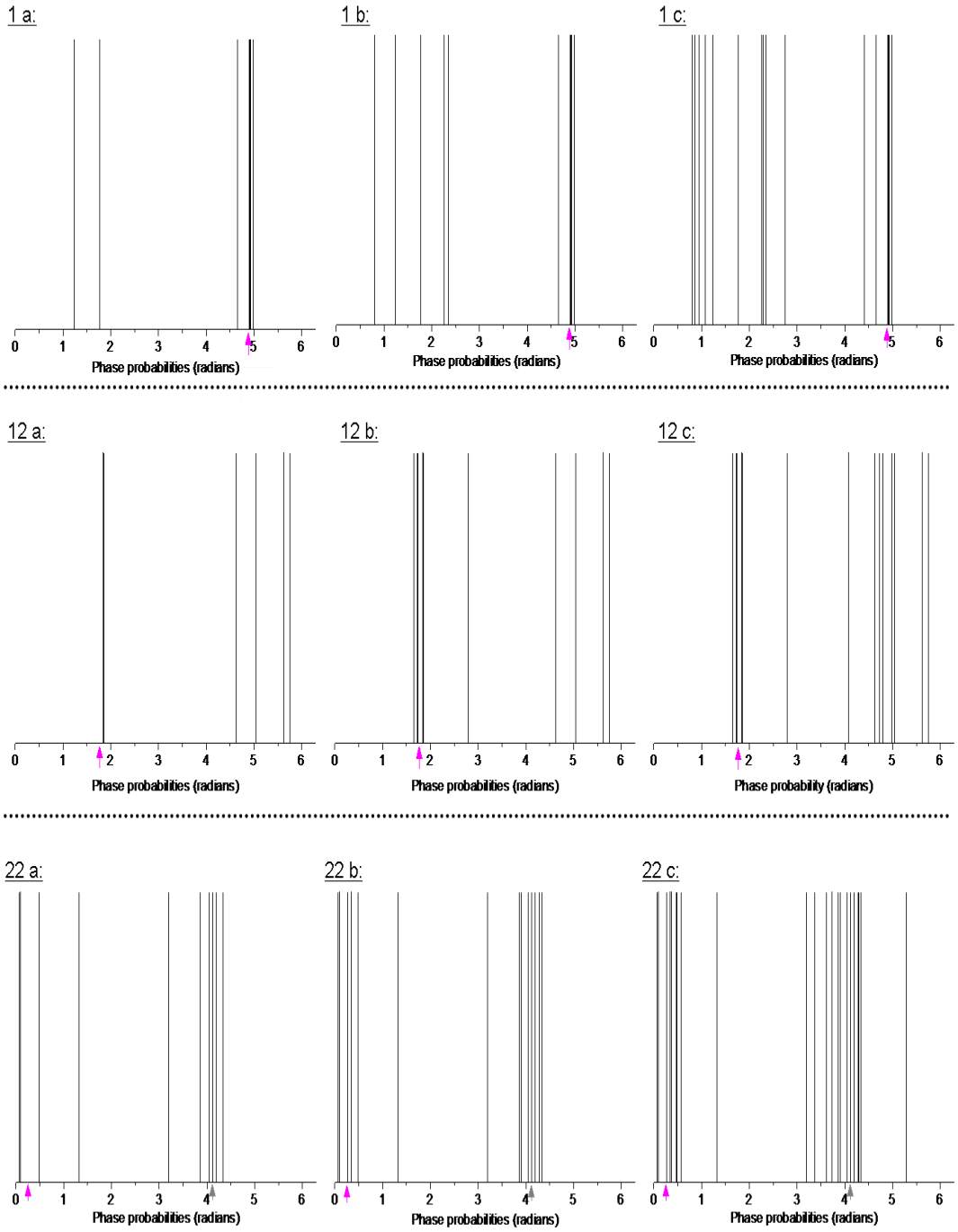


Figure 5.19 Examples of phase probability bar graphs for orders 1, 12 and 22. For each order, up to three plots were produced: (a) only bars representing a precise three-way Harker circle intersect included; (b) all bars from (a) plus bars representing close-to-three-way intersects; and (c) all bars from (a) and (b), plus bars representing non-three-way-intersects that nonetheless were close enough to indicate a probability to within $\pi/4$ radians. The magenta arrow along the *x*-axis shows the phase that was used for electron density map building. For orders 1 and 12, these were considered unequivocal. For order 22, the gray arrow indicates a second possible phase (see text).

Order	First preference phase (mean, in radians)	Number of individual phases values	% error	Second preference phase (mean, in radians)	Number of individual phase values	% error
1	4.8809 †	4	2.1	N	-	-
2	4.5677	3	2.2	1.0622* †	1	1 sample
3	0.4591	2	0.2	4.0230 †	2	6.1
4	0.1098 †	6	5.7	N	-	-
5	4.2365 †	2	0.9	N	-	-
6	1.8899 †	3	1.6	2.4829	5	1.8
7	2.2642	3	0.1	3.4533 †	3	2.7
8	4.4808 †	3	0.8	N	-	-
9	3.0875 †	3	0.4	3.8024	4	3.2
10	5.9516 †	3	5.0	N	-	-
11	1.8613	1	1 sample	5.7125 †	1	1 sample
12	1.7656 †	5	1.3	N	-	-
13	4.1526 †	2	0.2	0.9445	2	0.6
14	0.2783 †	2	3.4	3.5063	2	7.2
15	2.8377	4	3.4	4.7581 †	6	6.7
16	1.7029 †	3	1.5	0.6013	3	2.6
17	4.4974 †	1	1 sample	N	-	-
18	2.2302 †	3	0.1	N	-	-
19	4.5044 †	2	0.8	N	-	-
20	1.7770 †	3	1.4	5.2180	2	4.2
21	5.4280 †	4	0.9	2.4015	3	1.0
22	4.1134	5	2.8	0.2498 †	5	2.8
23	0.3526 †	5	3.9	N	-	-
24	5.1331 †	5	4.4	1.5082	4	3.6
25	0.6833 †	5	1.0	4.6140	6	2.1
26	4.1613 †	8	5.0	N	-	-
27	1.0180 †	3	3.2	0.6190	13	7.7
28	1.6016 †	4	2.2	4.7821	4	1.6
29	6.2385 †	3	1.7	N	-	-
30	5.2723	4	2.4	4.5485 †	4	1.7

Table 5.3 Phase probabilities obtained from the Harker diagram phase determination procedure. From the phase determination bar plots, the most likely phase was determined on the weight of visible evidence and calculated as a mean of the individual phases that suggested a particular probability mean. These are given on the left half of the table. For 18 of the 30 orders, it was determined that there were more than one possible phases and so the alternatives are given on the right half of the table. The "% error" for each phase is a sample-based SD divided by 2π and indicates precision and not accuracy. † The phases that were actually used in the eventual electron density map of MTT. Of these, 23 were first choice phases. * 2nd order second choice phase, based on a single sample, was used in preference to the first choice value, as it correlated better with the rat tail model and better resolved the "shoulder" problem. See text for further explanation.

Although order 17 was based on the one precise intersect phase probability value from the first plot, i.e. 4.4974, the third plot contained six values, four of which (including the value used) fell within the range 4.1445 to 4.6931. However, the other three values were less precise guide values and so the single value was used.

Orders 3, 7 and 15 were also equivocal with alternative phase probability values having been produced. Of these, the means for order 15 were clearly the least precise. Order 30 had also been determined as having two possible mean phase probability values, but these were sufficiently close together, that any effects on the correlation of the MTT electron density map with the RTT model (Pearson r-value difference of <0.0001) or the visible shape were small.

The use of the 2nd order value based on a single, second choice value was not only due to a spread of individual phase probability values across the bar plot and related to improvement of correlation with the model (Pearson r-value difference of 0.0167), but also to the shape of the MTT electron density map as compared with the models (RTT map, plus simulated electron density map). This latter will be discussed below. In addition to the preferred phase probability presented, there was a closely-spaced pair of individual phases (5.4925 and 5.5405, mean = 5.5165) and four other individual phase values spaced relatively evenly between 0.2296 and 4.0357. Because of this and the correlation and shape ramifications, the 2nd order phase probability was considered the most difficult to determine for the untreated native MTT.

5.3.5 Electron density plot for one MTT D-period using Fourier synthesis

The electron density map was gradually built up, using the Lorentz corrected amplitudes for native, untreated MTT presented in Table 5.2. Starting with the first order, addition of each new structure factor (amplitude and phase) had some effect on the map adding progressively higher resolution. Monitoring the changes in the nascent MTT map in terms of improved correlation with the optimally-aligned model (the Orgel RTT map) and the effect of the addition of each new structure factor on pre-existing features, proved to be a successful approach. In this way the relevance and suitability of each successive phase where there were alternatives, could be monitored.

The only order which proved problematic was the 2nd order. Of seven individual phase values derived from precise intersects, two closely-spaced pairs were seen on the bar plot indicating alternative phases of 4.4896 and 5.5165. Two additional bars in the

second tier plot added weight of evidence that the first of these was the most likely phase probability. However, use of either of these exacerbated the presence of a "shoulder" at the N-terminus of the overlap region. Use of one of the single phase probabilities, 1.0622, resulted in both an amelioration of this unexpected feature and an improved correlation with the rat tail model as compared with the "first choice" value given in Table 5.3, in the otherwise-complete map (Pearson r value: 0.960, compared with 0.943).

The MTT electron density map is presented in Figure 5.20. The complete list of structure factors used in its construction is shown in Table 5.4. Reasonable representations of the gap and overlap regions of the *D*-period are marked and their relative values, 0.52:0.48, are consistent with values previously suggested for RTT collagen, i.e. 0.52:0.48 (Bradshaw et al., 1989), 0.53:0.47 (Brodsky et al., 1982; Fraser et al., 1987; Antipova, 2008) and 0.54:0.46 (Orgel et al., 2000). The map is also overlaid with the Orgel RTT model and predicted MTT electron density model (Chapter 2, section 2.5) maps in Figure 5.21, with the correlations given. The correlation of the MTT map with the RTT is strong (Pearson's r value of 0.960, R^2 = 0.921), with obvious similarities between the two traces. The correlation of the experimental MTT map with the predicted map is weaker (Pearson's r value of 0.800, R^2 = 0.640). Whilst there are clear similarities between the experimental and predicted MTT maps, there are notable differences and peak shifts. The biggest overall discrepancy between the two is the ratio of the length of overlap and gap regions. The MTT experimental map clearly has a longer overlap region than this predicted model. Within the model, taken as an axial region where, for each given amino acid translational position there are fewer than 14 amino acids occupying loci as counted across five molecules in the fibril, the gap region is 0.58*D* and the overlap region is 0.42*D*. This is discrepant with the previously reported values and those given for the experimental map.

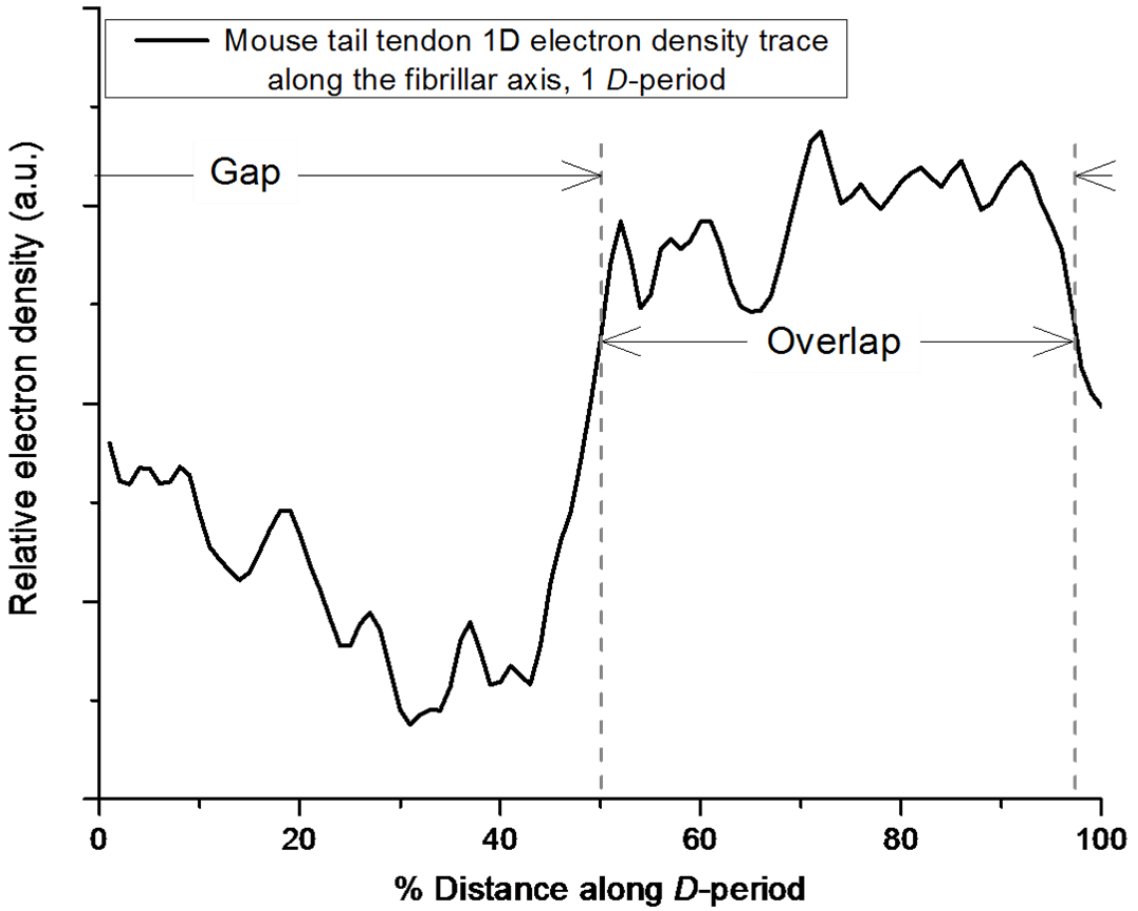


Figure 5.20 Electron density map of MTT type I collagen, representing one D-period. The gap and overlap regions are marked, and have a length of 0.52 and 0.48 the lengths of the *D*-period. Peaks at % distance positions 52, 57 and 61, to the left of the trough at position 65%, are in the "shoulder". This was improved by careful selection of the 2nd order phase. Structure factors for orders 1 to 30 were used in its synthesis, resulting in a theoretical resolution of 2.23 nm, equivalent to a 7.8 amino acid translation distance. Orientation of trace: N→C, left to right.

Order	Structure factor *amplitude*	Structure factor *phase*
1	257.8	4.8809
2	26.8	1.0622
3	65.4	4.0230
4	14.2	0.1098
5	32.5	4.2365
6	19.5	1.8899
7	15.3	3.4533
8	11.7	4.4808
9	27.5	3.0875
10	11.3	5.9516
11	9.28	5.7125
12	13.4	1.7656
13	2.67	4.1526
14	4.80	0.2783
15	4.12	4.7581
16	4.44	1.7029
17	4.20	4.4974
18	1.58	2.2302
19	2.78	4.5044
20	9.30	1.7770
21	8.62	5.4280
22	4.99	0.2498
23	1.78	0.3526
24	0.95	5.1331
25	5.93	0.6833
26	2.34	4.1613
27	4.14	1.0180
28	1.30	1.6016
29	2.08	6.2385
30	2.68	4.5485

Table 5.4 Structure factor amplitudes and phases used in construction of the one-dimensional untreated, native MTT electron density map. The amplitude values are Lorentz-corrected arbitrary units and the phases are in radians.

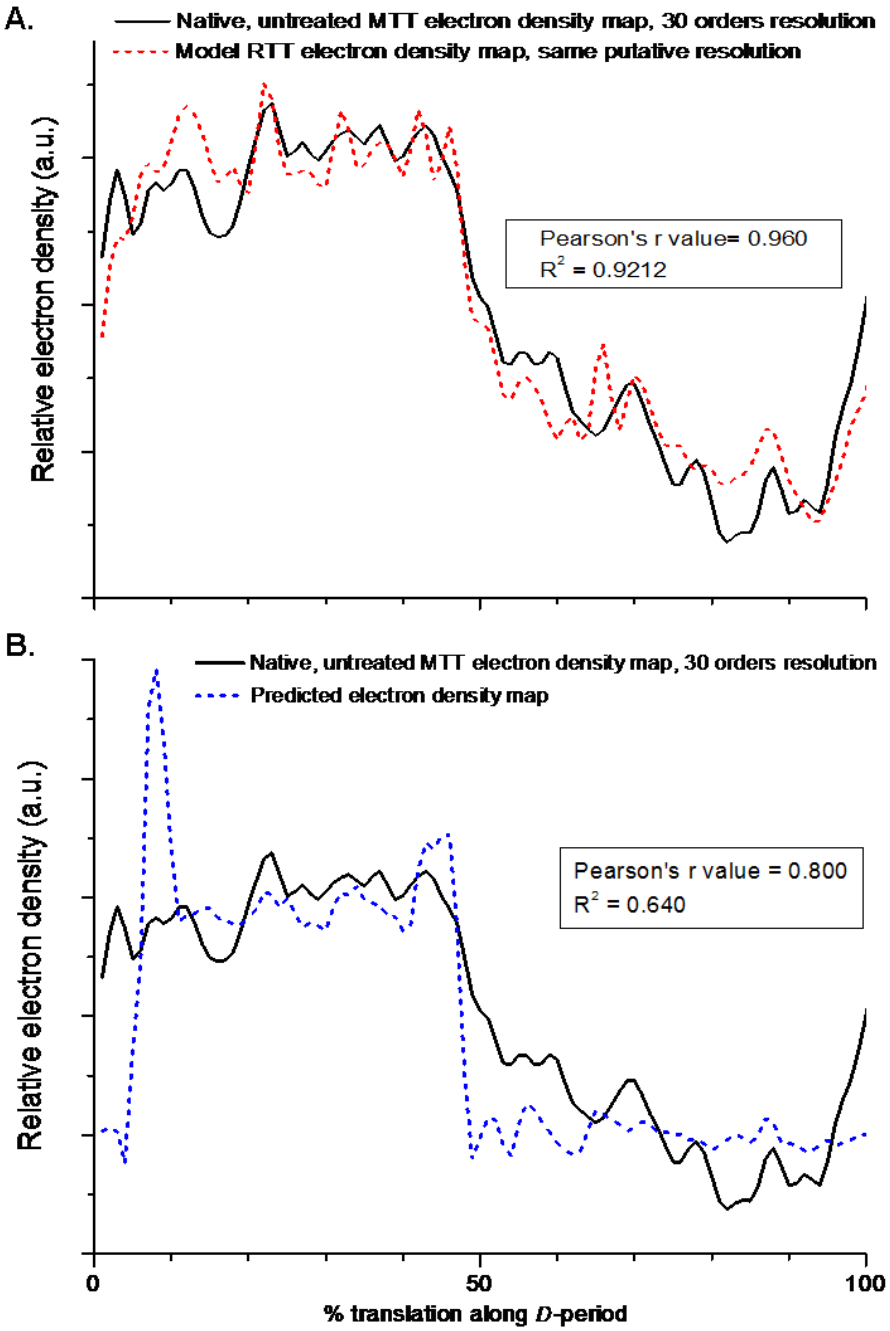


Figure 5.21 Native MTT electron density map of one *D*-period, shifted along the abscissa to align optimally (i.e. to produce the best correlation) with **A.** the Orgel RTT experimental electron density plot and **B.** the predicted electron density map derived from the simulated fibril *D*-period (Chapter 2, section 2.5). The original *D*-period of axial length 234 amino acids was smoothed using the Savitsky-Golay algorithm (8 point window, 3rd order polynomial) and interpolated to 100 points. Pearson's r and R^2 values are given. Orientation of maps: N→C, left to right.

The original simulated *D*-period model brought the primary sequence lysines along the N- and most especially, the C-telopeptides into the proximity of their corresponding triple helical lys-hyl cross-linking sites suggested previously (Orgel et al., 2000; Brady and Robins, 2001; Malone et al., 2004; Malone and Veis, 2004) by folding the telopeptides in a hairpin loop manner, assuming a constant translation distance for each amino acid. Adjustment of the telopeptide conformations in the model produced improvements in the correlation between the model electron density map and the MTT map after optimal alignment. Electron density maps for a number of combinations of straight and folded or bent N- and C-telopeptides were produced and assessed in this way. Table 5.5 shows to what extent these variations affected the correlation between the predicted and experimental maps.

Alteration of original predicted MTT electron density model	Correlation of modified model on optimal alignment with experimental MTT electron density map (Pearson's r value)
Original model	0.800 *
N-telopeptides all straight	0.901†
C-telopeptides all straight	0.785
One α1 N-telopeptide straightened	0.860
One α1 and α2 N-telopeptide straightened	0.878
α2 N-telopeptide straightened	0.836
One α1 C-telopeptide straightened	0.808 ‡
N-telopeptides all straight, one α1 C-telopeptide straightened	0.930 **

Table 5.5 Correlations of variations of the original MTT electron density model with the experimental MTT electron density map. * - Original model (Figure **5.21**B); † - best correlation achieved by straightening a combination of N-telopeptides in the original predicted map; ‡ - best correlation achieved by straightening a combination of C-telopeptides in the original predicted map. ** - combination of the best straightening adaptations of the N- and C-telopeptide ends of the collagen molecules.

Clearly, the model containing all N-telopeptides in the straight conformation and one of the α1 C-telopeptides in the straightened conformation, the other remaining in the originally-predicted folded conformation, produced the best fit with the experimental map (Pearson r value = 0.930, R^2 value = 0.864). The comparison of the improved model with the experimental map is shown in Figure 5.22. From visual inspection of this and the correlation value, the similarities are more obvious.

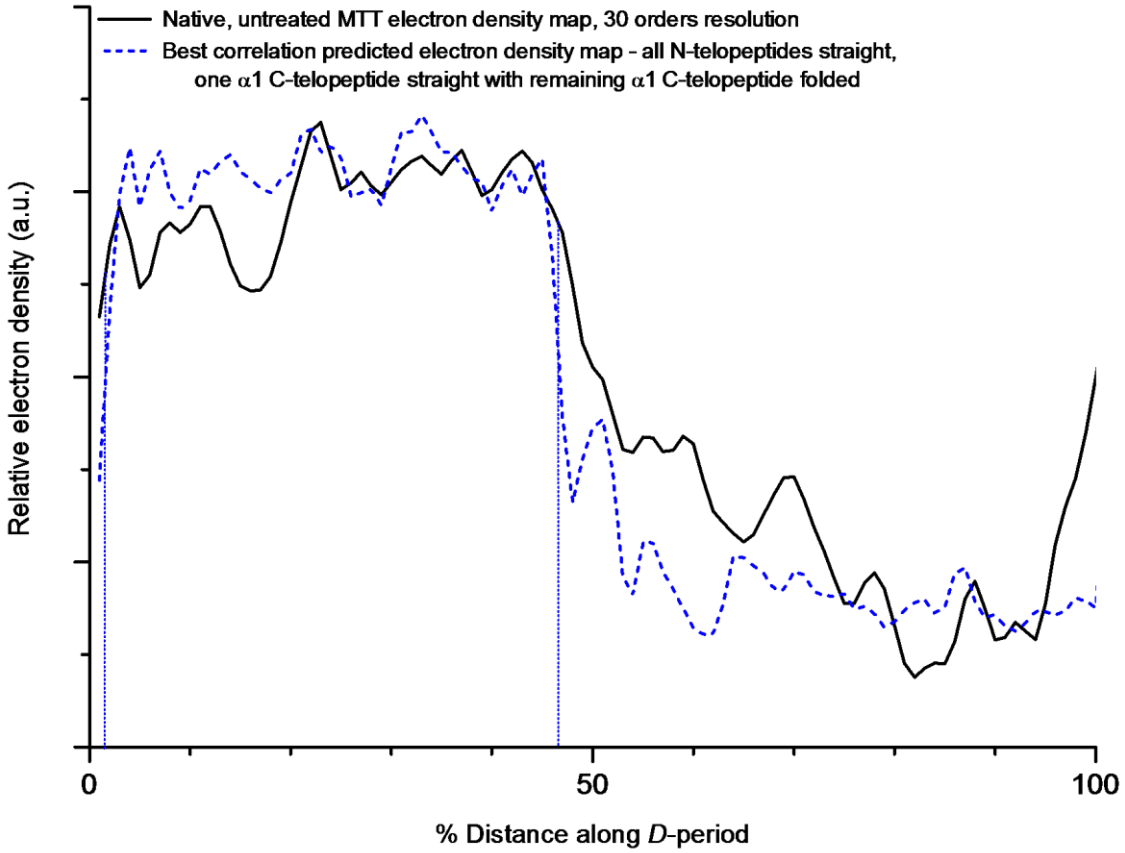


Figure 5.22 Visual comparison of experimental MTT electron density map with the predicted model containing extended N-telopeptides and one extended α1 C-telopeptide with the other in the originally-modelled folded conformation. The model was smoothed and interpolated in the same manner as the original (Figure **5.21**B). The vertical dotted blue lines represent the demarcation between the model's overlap and gap regions within the *D*-period, the definition of the overlap being the presence of 14 or more amino acids axially across the 5 simulated collagen molecules (15 α-chains) of the model fibril, at any given locus. From this model, the gap:overlap ratio is 0.55:0.45. Orientation of map: N→C, left to right.

Using the iodine and gold difference amplitudes to produce the difference Patterson plots during the scaling, prior to phase determinations (section 5.3.3), with the native phases (Orgel, 2000), iodine and gold difference Fourier plots were produced showing the binding positions of both heavy atoms respectively, along one *D*-period. These are compared with the positions of their respective binding site amino acids in Figure 5.23 and Figure 5.24 respectively, in both the original simulated fibril model (**A**) and the improved-fit model described (**B**).

In the case of the iodine plot (Figure 5.23), the peaks at 2% and 37% appear neatly to fit with the tyrosine residues folded into the overlap region in the original model, with a

smaller peak at 41% coinciding with a cluster of histidines. The five strong peaks at 66, 69, 74, 78 and 82% appear to be associated with the four single histidines and the lone tyrosine in the gap region, though the positions do not precisely match those modelled. A cluster of smaller peaks in the overlap region, plus peaks at 46, 51, 59 and 89% in the gap region appear not to have any association with an identifiable binding site as originally modelled. For the improved model, the extended C-telopeptide α1-chain has resulted in a further peak – at 46% - being matched to two tyrosine residues pushed out into the gap region. In addition, all the N-telopeptide tyrosines being extended outwards produce stronger justification for the existence of the peak at 99%, which appeared to be associated with only a single more weakly binding histidine residue in the original model.

For the gold plot (Figure 5.24), the large peaks at 41 and 55% fit in with sizeable clusters of gold-binding amino acids. The broadened peak at 73% appears to be associated with a smaller cluster of two methionines and a histidine and the peak at 21% appears to be associated with the pair of methionines immediately adjacent.

The broad peak at ~10% is offset from the cluster of methionines to its left, but that does not rule out the possibility of there being an association and the spacing in model and experimental plots being discrepant. As with the iodine plot, there are a few peaks which do not appear to be accounted for by methionine or histidine gold binding. Most notable of these is the double peak at 86 and 90%.

With regard to the models, the extending of the N-telopeptides has moved a couple of methionines outwards, to correspond with the gold difference Fourier peak at 97%.

In the cases of both the iodine and gold difference Fouriers, the improved simulated *D*-period fibril model has resulted in better fitting of more of the peaks, adding some evidence favouring the extended N-telopeptides and one straight plus one folded C-telopeptide over the original all-folded configuration.

Remaining discrepancies and unattributed peaks have several possible explanations, which will be discussed below.

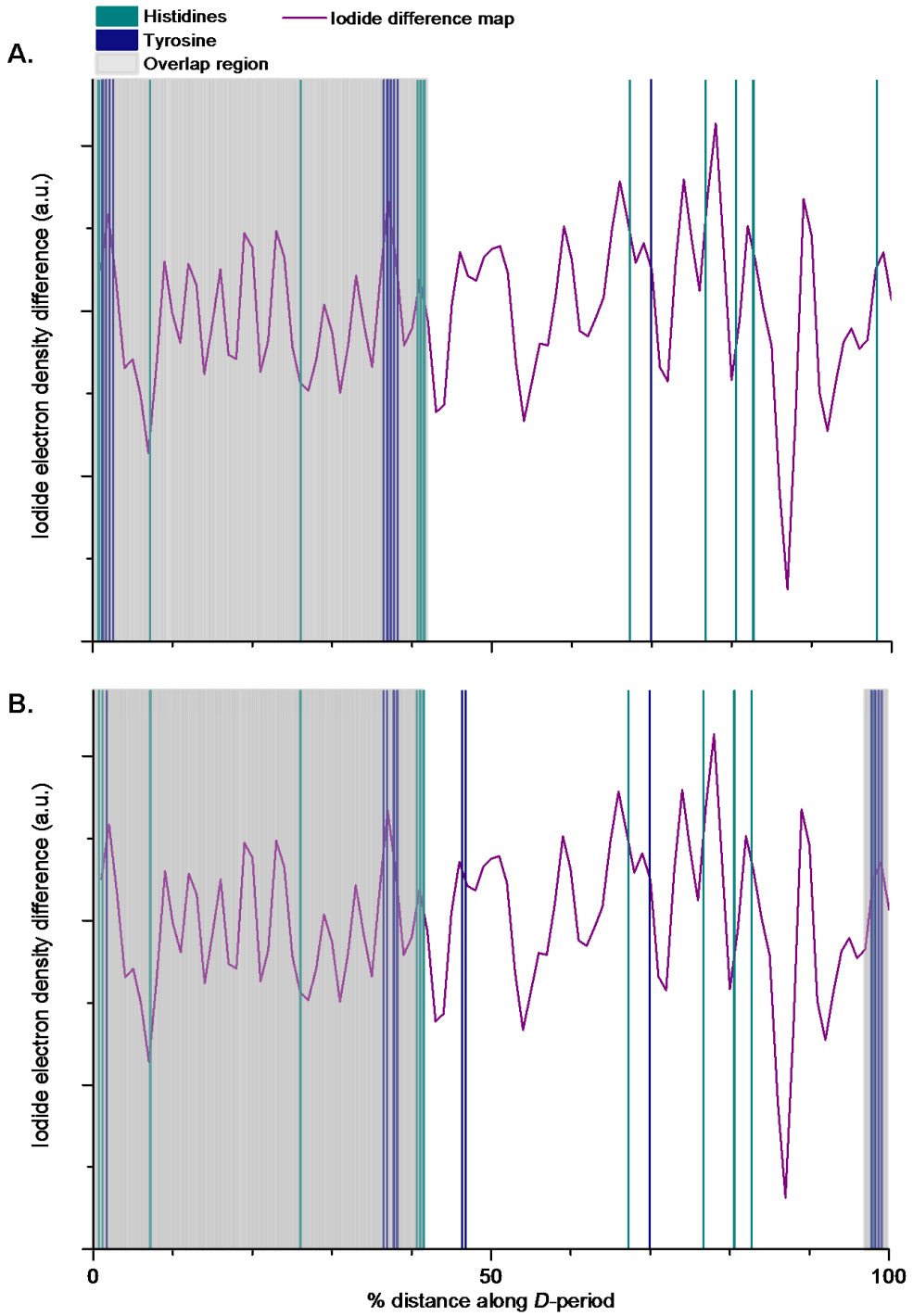


Figure 5.23 Iodine difference Fourier plot overlaid with **A.** original simulated *D*-period model and **B.** improved fit model with all N-telopeptides straightened and one C-telopeptide α1-chain straightened and the other retaining its original fold. The grey background denotes the overlap region, the white corresponding to the gap region. Histidine and tyrosine loci (iodine binding sites) are shown as cyan and blue vertical lines, respectively. Left to right represents N- → C-terminal direction.

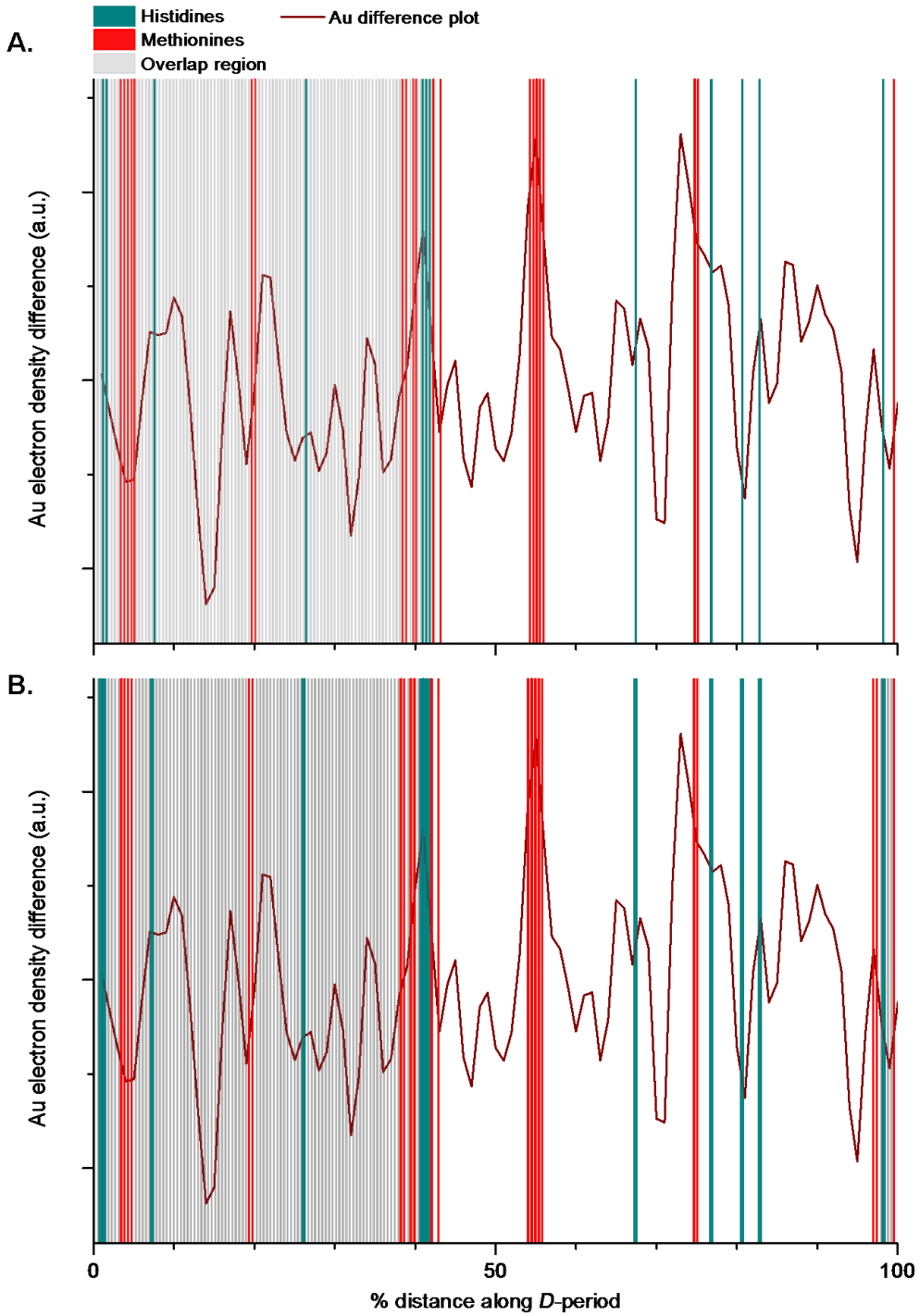


Figure 5.24 Gold difference Fourier plot overlaid with **A.** original simulated *D*-period model and **B.** improved fit model, as in Figure **5.23**. Histidine and methionine loci (Au binding sites) are shown as cyan and red vertical lines, respectively.

5.4 Discussion and conclusions

In this chapter, the use of X-ray diffraction with isomorphous replacement techniques using synchrotron radiation to produce an axial one-dimensional electron density map of the MTT type I collagen *D*-period has been described.

With regard to the small angle/wide angle X-ray diffraction data collection setup, similar to that which was noted previously by Wess et al. (1995), the relatively long sample to detector distance (470 mm) used in the data collection during these experiments allowed the reflections that were detected to occupy an area covering a relatively large number of pixels on the detector. The noise level was therefore reduced, while still allowing measurement of the first order peak height and allowing data for orders up to the 30th to be collected with just the one beam line set-up.

The isomorphous replacement procedures successfully yielded gold and iodine MTT collagen derivatives. However, the iodine-treated collagen data patterns were visually indistinguishable from the native patterns, contrary to what was expected (Bradshaw et al., 1989; Orgel, 2000; Antipova, 2008), with a strong correlation between them. Further experiments using increasing concentrations of KI solution in PBS of 2, 4 and 10 mg/ml were conducted in an attempt to improve the extent of derivative formation. However, the data failed to save to the host computer at APS and were lost, so the success of this approach could not be appraised. An additional approach is the use of an oxidant such as chloramine T to potentiate the reaction (Bradshaw et al., 1989; Tsomides and Eisen, 1993; Orgel, 2000), though this approach was not used during these experiments. From Figure 5.17, it is clear from the iodine difference Patterson plots presented during the amplitude scaling exercise prior to phase determination and the resemblance of these scaled plots to the predicted Patterson plots, that the process had produced a useable iodine derivative. It was too weak to be detectable purely from inspection of the initial data patterns but the scaling algorithm employed proved successful. Similarly, the gold derivative resulted in an experimental difference Patterson similar to those predicted post-scaling (Figure 5.18).

The discrepancies between the predicted heavy atom difference Patterson plots and the experimental ones are thought to have two main causes. Firstly, the binding of iodine and gold to tyrosine and methionine respectively are mechanistically different and preferential to their binding to histidine (Tsomides and Eisen, 1993; Glišić et al., 2012). Iodination of both tyrosine and histidine is covalent but with formation of

iodotyrosine being favoured between pH 6 and 7, with increasing proportions of iodohistidine being formed at pHs above this range (Tsomides and Eisen, 1993). Binding of gold to methionine is thought to involve a redox reaction, in which the thioether sulphur atom is oxidised to a sulphoxide resulting in formation of metallic gold and localised gold-containing anions. Conversely, binding of gold to histidine is likely to be mediated by the nitrogen atoms in the amino acid's imidazole group, resulting in stabilisation and localisation of the gold ion (Glišić et al., 2012). The occurrence and rates of these reactions are likely to vary depending on the local environment of each amino acid. Some of the relative peak heights on the Patterson plot will differ from those predicted therefore, and this can be seen for both heavy atoms. Secondly, the simulated fibril assumed a constant axial distance per amino acid, which is known to vary. The telopeptides are known to have a lower axial translational distance than the triple helix. Hulmes et al. (1980) stated that the N-telopeptide chains are contracted 85% and the C-telopeptide 70% relative to the helical distance. With a helical amino acid translational distance of 0.286 nm, that would suggest that the N-telopeptide translational distance is 0.243 nm and the C-telopeptide distance is 0.200 nm. Folding of the telopeptides is likely to be an important factor here, whatever conformation(s) is/are adopted. The model used here, based on amino acid positions in cells on a spreadsheet, did not take these differing distances into account. The paths of collagen molecules through the three-dimensional unit cell or microfibrils are not parallel to the fibrillar axis and the inclination away from this axis varies according to whether it is in the overlap or gap region (Orgel et al., 2001). In addition, different regions along the triple helix are known to be less rigid and more flexible and susceptible to stretching under tension (Silver et al., 2001) and different amino acids tend to show predilections for slightly different translations distances (Cameron, Cairns and Wess, 2007). Therefore even along the triple helix, the 0.286 nm figure can only be considered an average with the spacing varying by 5% or more (Chapman et al., 1981). Discrepancies in the peak positions between the heavy atom Patterson plots and the predicted plots are likely as a result of these variations in the locations of the axial heavy atom binding sites.

The use of isomorphous derivatives has been stated as being a means of calculating phases without reliance on a model (Bradshaw et al., 1989). During the phase determination step, it became evident that some of the Harker phase probabilities produced were not clear cut. Several had two possible phases determined, though usually with one being considered more likely than the other. A greater number of phase probabilities being determined with more confidence may have been achieved

by using three or more isomorphic derivatives. Early work to elucidate the three-dimensional structure of haemoglobin used six isomorphous derivatives to determine phase angles (Perutz et al., 1960). In deducing their model-independent electron density map of RTT Bradshaw et al. (1989) employed isomorphous replacement with platinum binding to methionine residues, in addition to iodine and gold. Orgel et al. (2000) used iodine and gold, but used UV irradiation in addition with the iodine replacement to modify the tyrosine binding and therefore produce a variant iodine derivative.

Most (23 out of 30) of the first choice phases were used in the eventual electron density map but where there was a choice, correlation with the Orgel RTT model was regarded as a viable approach alongside an assessment as to how successively added phases contributed to the sharpening and defining of already-existent features in the nascent plot. Therefore the approach here could not be stated as model-independent. The hypothesis that "mouse and RTT collagen would have identical or close to identical electron density maps" could be argued as a self-fulfilling prophecy inasmuch as RTT collagen was a part of the solution for removal of ambiguity. However, it was only a part of the approach based on a reasonable assumption that mouse and rat tail should be close, structurally. Type I collagen α1-chains are 1056 amino acid residues long in rat and mouse with the α2-chains being 1023 residues long in both. Side-by-side comparison of both shows 16 differences in amino acid sequence between rat and mouse α1-chains with the figure being 41 differences for the α2-chain (UniProt, 2012).

The MTT electron density map presented shows obvious similarities with the RTT model but notable differences as compared with the predicted MTT model. Clearly, the two most striking differences in the model are the smaller overlap region length as a proportion of the entire *D*-period and the notable spikes in electron density map coinciding with the positions of the telopeptide regions. The model was based on the quarter stagger off-set and 234 amino acid *D*-period (Meek et al., 1979; Hofmann et al., 1980) (see Chapter 2, section 2.5 and Appendix 1) and the assumption that the N- and C-telopeptides are folded in order to bring each telopeptide lysine/hydroxylysine into close proximity with its triple helix cross-linking counterpart, i.e. $\alpha1(I)^N$ residues 9 and $\alpha2(I)^N$ residue 5 in register with α1(I) and α2(I) triple helix positions 930 and 933 respectively, and the $\alpha1(I)^C$ residues 16 in register with the α1(I) and α2(I) triple helix residues 87 (George et al., 1999; Brady and Robins, 2001; Malone et al., 2004; Malone and Veis, 2004). Previous research suggested that the telopeptides at both ends of the collagen molecules existed in a contracted conformation as compared with the triple

helix (Hulmes et al., 1977; Wess et al., 1998b) and there are differing theories as to the correct conformations. Orgel et al. (2000) suggested the N-telopeptides, though contracted, extended straight outwards from the triple helix, whereas Malone et al. (2004), through computer modelling using favoured conformation and energy reduction algorithms, showed evidence that they should be folded and exhibited specific binding with a region of a neighbouring triple helix, bringing the cross-linking Lys-Hyl pair into close proximity. This latter model was supported by other, previous work (Vitagliano et al., 1995; George et al., 1999). The resolution of the map presented here is 8 amino acids, so not high enough to determine conformations based on individual amino acid positions. However, the data presented here do tend to favour a model with all three N-telopeptide α-chains unfolded and extended and with the C-telopeptides in different conformations, one $\alpha 1$-chain folded in a manner as suggested by Orgel et al. (2000) and the other unfolded and extended. Malone et al. (2004) suggested a more heterogeneous arrangement and less hairpin-like bend in the C-telopeptide. Figure 5.25 shows the original simulated MTT fibril C-telopeptide model produced for this study (**A**) and the revised model (**B**), which better correlates with the MTT electron density map. Similarly, Figure 5.26 shows the folded N-telopeptide model originally included in the simulated fibril for this study (**A**) compared with the updated better-fitting extended N-telopeptide model favoured by the experimental work described (**B**).

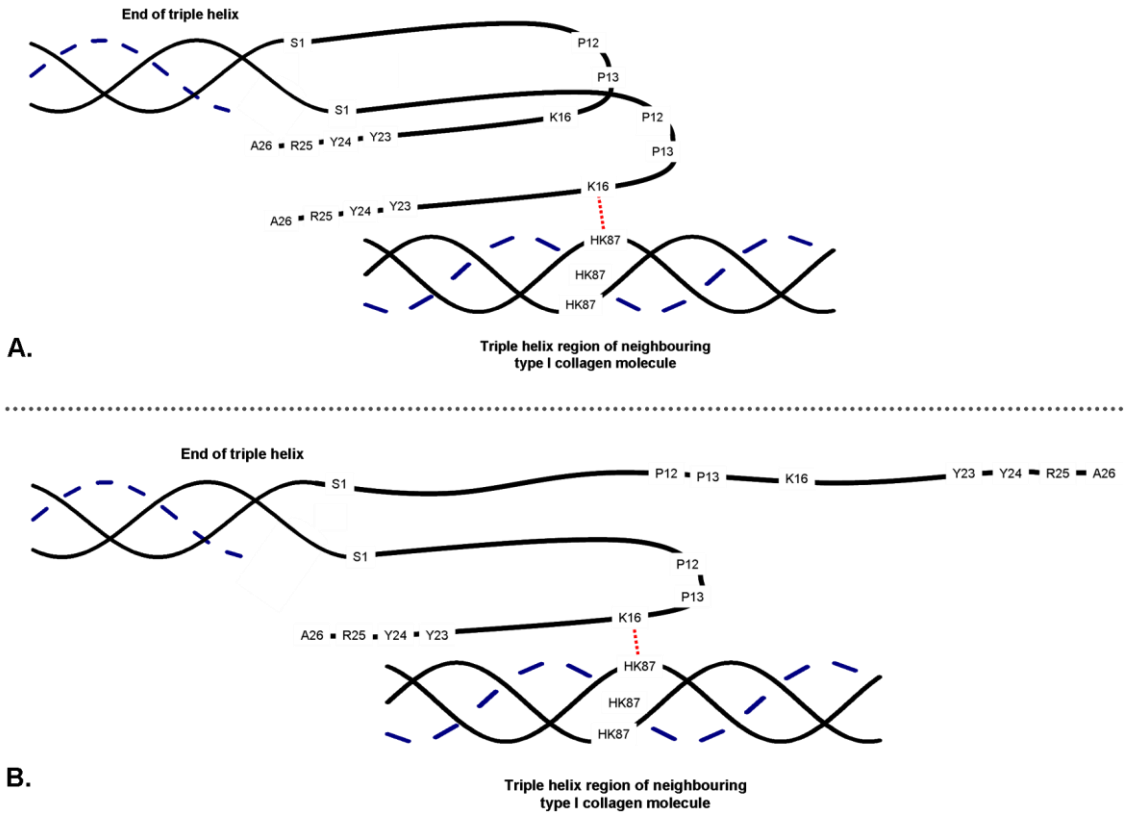


Figure 5.25 **A.** Model of C-telopeptide conformation, similar to that shown in RTT by Orgel et al. (2000). Both α1-telopeptide chains are folded into a hairpin, bringing the Lys16 (or Hyl16) into closest proximity with its cross-link counterparts Lys/Hyl87, present along all three α-chains in the triple helix. **B.** Adjusted model, giving an improved fit with the MTT experimental data. One of the α1-telopeptide chains is present in the same folded conformation as in **A**, with the other being extended outwards. Black lines represent α1-chains and blue dashed lines represent α2-chains. HK represents triple helical Lys/Hyl sites. Red dotted lines represent possible Lys/Hyl cross-links.

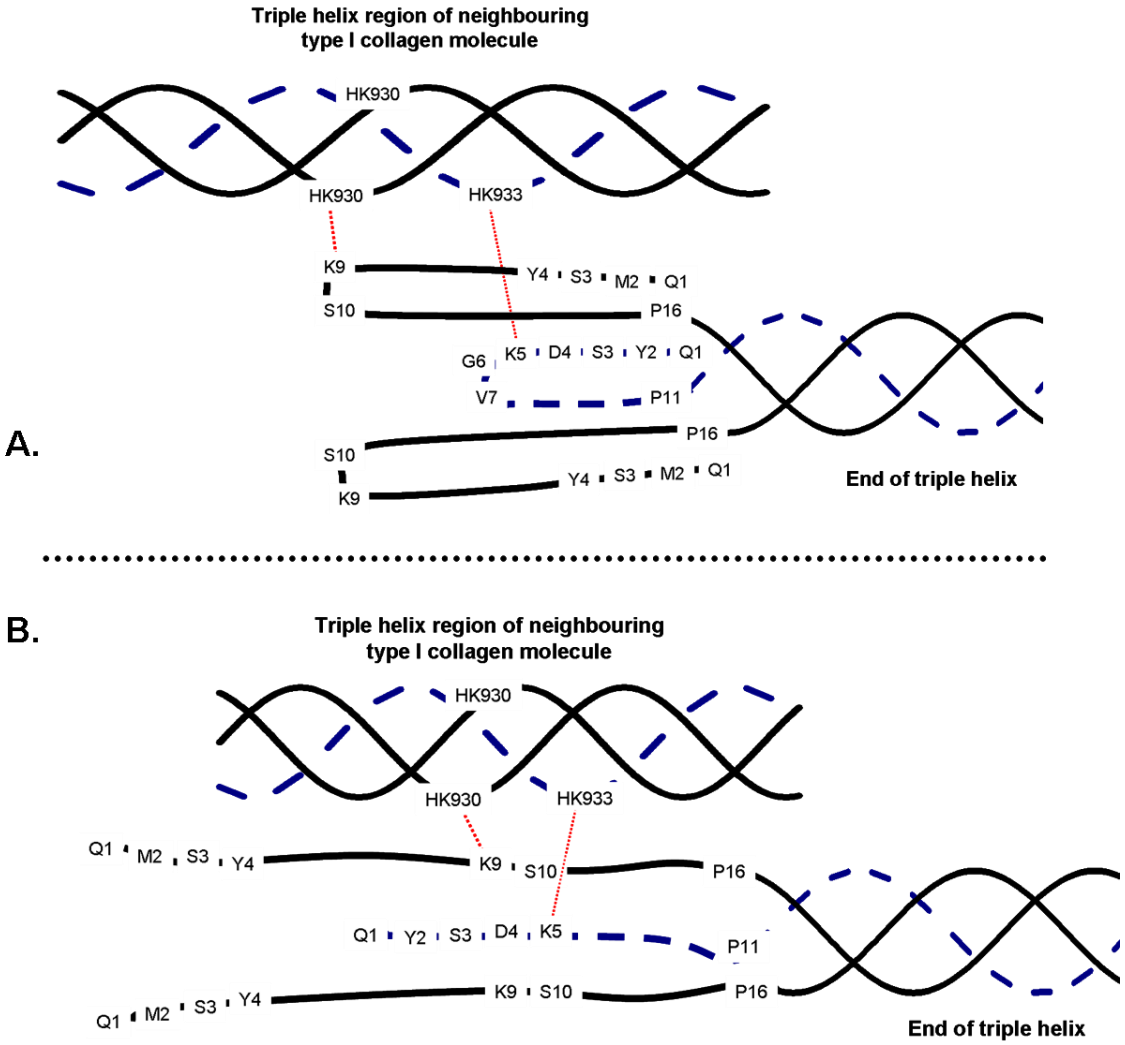


Figure 5.26 **A.** Model of N-telopeptide conformation, with both α1-telopeptide chains and the shorter α-2 telopeptide chain are folded into a hairpin, bringing the Lys9 (or Hyl9) of the α1-telopeptides into closest proximity with their cross-link counterparts Lys/Hyl930, along the α1-chains in the triple helix and Lys5 of the α2-telopeptide into closest proximity with its cross-link counterparts Lys/Hyl933, along the α2-chain in the triple helix . **B.** Adjusted model, giving an improved fit with the MTT experimental data and similar to that shown in RTT by Orgel et al. (2000). Both α1-telopeptide chains and the α2-telopeptide are extended outwards into the gap region. Key as in Figure **5.25**.

The iodine and gold difference maps presented in Figure 5.23 and Figure 5.24 were considered to show further evidence favouring the all extended N-telopeptide/one extended – one folded α1 C-telopeptide regions. These maps showed predicted versus experimental disparities in terms of peak height and positions, which would be explained in the same ways as for the heavy atom difference Patterson plots considered above. Some non-specific peaks are present in both heavy atom Fourier difference plots. Most notably on the iodine plot, there is a prominent sharp peak at the

89% position and on the gold plot, a corresponding double peak at 86% and 90%. This region corresponds with the *d* positive staining band as seen by TEM. In the iodine plot, there is also a peak at 52% corresponding to the *a1* positive staining band and a further sharp band at 59% in the *e2* positive staining band location. This is significant as these are all locations along the *D*-period, which are involved in binding PG molecules (Di Lullo et al., 2002; Sweeney et al., 2008) present in tendon, mainly decorin (Pins et al., 1997; Fessel and Snedeker, 2009) (see also Yoon and Halper (2005) for a review of proteoglycans present in tendon). It is important to remember that the X-ray diffraction data were collected from intact, freshly-dissected MTT and not isolated collagen. Therefore any PGs bound at specific sites along the fibrils will affect the pattern and bind heavy atoms, assuming they contain the amino acids to which they bind. Therefore an additional means for differences being introduced into the experimental heavy atom difference Patterson plots and difference Fourier plots for MTT tissue as compared with the predicted plots for MTT collagen, has been identified. This would also affect the electron density map itself as other components of the tendon matrix are included.

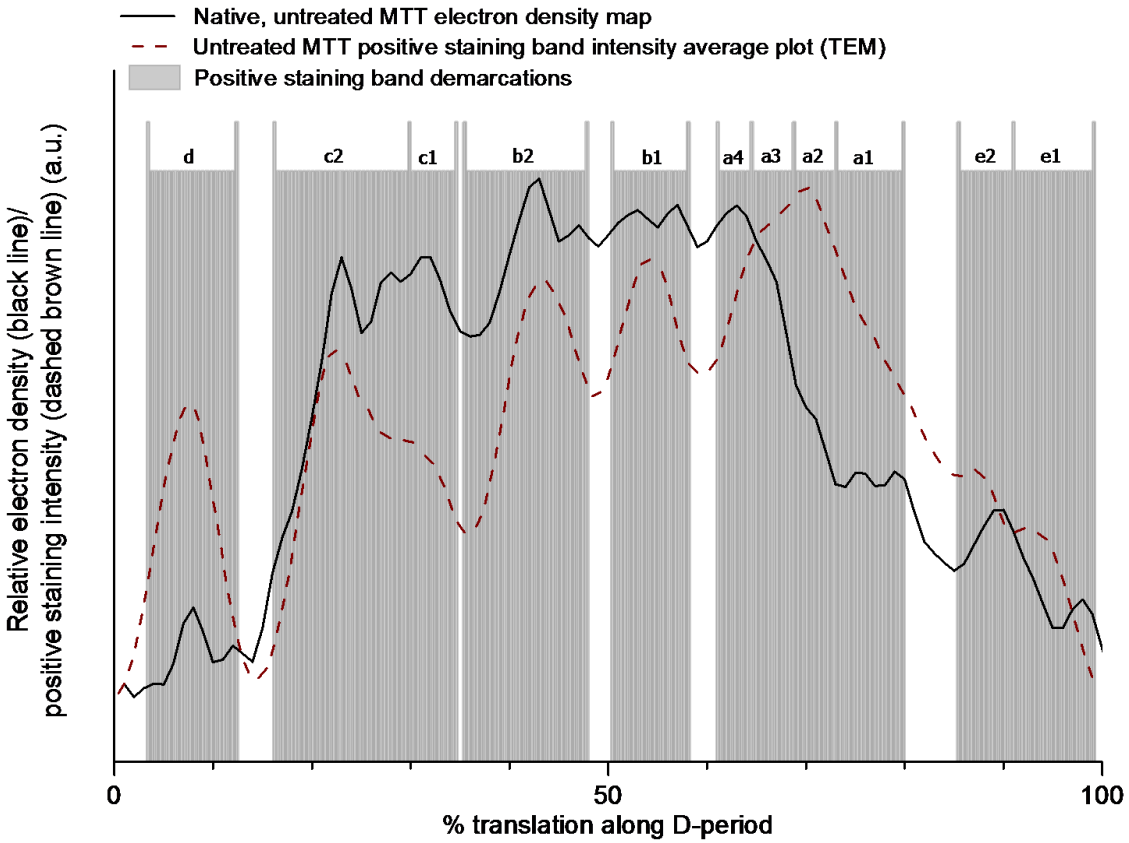


Figure 5.27 Comparison of the native MTT electron density map with the average (n=20) TEM positive staining band intensity plot presented in Chapter 4. The locations of the positive staining bands are overlaid.

For comparison, Figure 5.27 shows the MTT map overlaid with the native MTT positive staining band intensity average determined by TEM and presented in Chapter 4, section 4.3.3 (see Figure 4.4), and identifies how the positive staining bands coincide with the features within the electron density map, similar to the presentation in previous work that studied RTT (Hulmes et al., 1977; Bradshaw et al., 1989). Whilst there are discrepancies, the positions of the electron density bands presented here conform to the positions of the positive staining bands as has been reported previously (Chandross and Bear, 1973) as well as in the Hulmes et al. (1977) and Bradshaw et al. (1989) work. There appear to be few specific differences in peak positions, except for the slight shifting of the TEM positive staining *e* bands to the left relative to those in the electron density map. There may be a contribution of the differences between wet and dry fibrils, inasmuch as the dry fibrils observed in the positive staining-TEM procedure could some degree of axial contraction of fibrils within the gap region.

Although the two plots were produced using entirely different scales along the ordinate, they are presented here visually to the same linear maximum and minimum. It is obvious that the positive staining TEM band intensity trace lies below the electron density trace between the *c2* and *a4* bands – i.e. approximately in the overlap region – and above it in the gap region. This may be due to increased accessibility to the stain of basic residues within the gap region as compared with the overlap, or equatorial contraction artificially increasing the concentration of stain along each band in the gap region. Hence the apparent increase of the *a1-a2-a3* bands, the *d* band and to a lesser extent, the *e1-e2* bands relative to their electron densities.

It is appropriate to comment here that although this work has concentrated on obtaining an axial one-dimensional electron density map of MTT, equatorial reflections proved to be either undetectable or too faint for quantitation throughout these experiments. For example, the pattern shown in Figure 5.7 contains one of the best equatorial patterns recorded. It would be reasonable to assume that this is because each data collection used a single fibre of small diameter, roughly around 300 μm. The relatively precious nature of the samples and relatively low yield of fibres per tail dictated that bunching of fibres together or use of thicker sections of entire tendon in order to improve the resolution of this aspect of the pattern was not justifiable. It is anticipated that this approach in future work in order to elucidate packing arrangement of collagen molecules with fibrils would produce better results.

In this chapter has been presented work that has allowed the synthesis based on experimental data, of a one-dimensional axial electron density map of untreated MTT

using X-ray diffraction and isomorphous replacement techniques. It has a gap:overlap length ratio of 0.52:0.48 and showed a close similarity to Orgel et al.'s (2000) RTT map when adjusted to the same resolution. It also correlated best with a simulated fibril conformation created containing N-telopeptides, all three of which were not folded and were extended outwards along the fibrillar axis, and C-telopeptide regions, one α1-chain of which was folded bringing the lysine within it into closest proximity of its cross-linking counterpart Lys/Hyl and the second α1-chain unfolded and extended outwards. Additional weight of evidence was given to favouring this model by the peaks in the iodine and gold Fourier difference maps. This latter evidence was affected by noise and the apparent presence of non-collagenous components within the tissue, which bound the heavy atoms.

The extent of the differences between the MTT electron density map presented here and the RTT map presented by Orgel et al. (2000) (see Figure 5.21A) are unexpected. Although there are some differences between the primary sequences of the rat and mouse type I collagen α-chains which might explain minor differences, the most profound discrepancies relate to the C-telopeptide conformations of the α1-chains. It could be that these telopeptides are organised differently between the two species, though unlikely due to rat's and mouse's taxonomic closeness and the fact that the telopeptide primary sequences are identical. It is more likely that differences in experimental approach are factors. For example here, tissue was dissected into, and temporarily stored in, PBS buffer whereas not during Orgel et al.'s (2000) work. In addition, the degree of tensioning employed to remove the crimp in the tendon during data collection may have differed between the two studies resulting in differences in the telopeptide conformations as a result of differing amounts of applied stress (see Mosler et al. (1985)). Acknowledging differences between the scope and focus of Orgel et al.'s (2000) study and that presented here, these discrepancies are nonetheless interesting.

This MTT electron density map can be considered reasonable and a relevant baseline and starting point, against which to assess any changes during glycation incubations considered in the next chapter.

Chapter 6: Construction and comparison of electron density maps for glycated and non-glycated mouse tail tendon collagen using X-ray diffraction and isomorphous replacement techniques

6.1 Introduction

The numerous effects of glycation on the properties of type I collagen have been described earlier in this work (see Chapter 1, section 1.3) and elsewhere as referenced. However, the specific changes in the physical molecular structure of collagen that bring about and contribute to these effects have not been extensively reported in the literature.

Although some studies have produced valuable evidence and information regarding the location and nature of glycation in collagen, much of this work has involved fragmentation and chemical modification of samples prior to analysis (Le Pape et al., 1984; Tanaka et al., 1988b; Reiser et al., 1992b). The knowledge yielded has tended to relate to primary sequence locations and specificities of glycation, also often relying on assumptions that relatively newly-discovered substances and adducts are not themselves affected by the processes involved in the experiments. This led for example, to a long-standing importance being attached to pentosidine as a dominant cross-link *in vivo* (Sell and Monnier, 1989; Monnier, 1990; Reddy et al., 2002) with other more physiologically and biochemically-relevant AGE cross-links such as glucosepane being overlooked due to their susceptibility to degradation by some of the hydrolytic methodologies in use (Biemel et al., 2002; Sell et al., 2005). Acknowledging this, Wess et al. (1993) used neutron diffraction to study glycation of RTT *in vivo* using diabetic and control rats. Although treatment of the tissue by deuteration reduction was in itself a chemical modification it was not thought to have affected the locations of the glycation products, so the group were able to study the entire collagen tissue and sites of glycation along the *D*-period, which appeared to favour Hyl-containing sites. The differential reduction treatments also allowed some information about the chemical nature of the cross-links to be deduced. However more advanced non-reducible AGE

cross-links could not be detected using this method. They suggested that X-ray diffraction would be able to detect these AGEs.

X-ray diffraction work has already been carried out to elucidate the effects of glycation on type I collagen-containing tissue structure. Research carried out by Tanaka et al. (1988a) used X-ray diffraction to investigate structural changes due to glycation of RTT *in vitro*. Their major finding was the expansion of the packing arrangement of collagen molecules within the fibril during glycation with a loss of packing order over time. This was seen on the X-ray diffraction pattern as a progressive loss of the characteristic equatorial and near-equatorial reflections over the increasing incubation times used. They found this expansion to occur along a vector perpendicular to the fibrillar axis and hypothesised it was caused by glycation-related cross-links forming between molecules and pushing them apart (Tanaka et al., 1988a). The meridional reflections remained sharp showing that the axial order remains well defined. The only significant changes in the meridional reflections identified were in the relative intensities of orders 14 to 19, these mainly having been identified as being related to the non-helical portions of the collagen molecules, i.e. the telopeptides (Brodsky et al., 1982). Further X-ray diffraction evidence supporting the work of Tanaka et al. (1988a) showed that an expansion of the fibrillar collagen molecular packing spacing occurs in extensor tendons from toes of human diabetic patients as compared with control subjects alongside similar changes in the relative intensities of the same meridional orders (James et al., 1991). Work using TEM shows that the increase in molecular packing spacing is accompanied by an increase in fibrillar diameter and an increase in the percentage area across the tendon occupied by fibrils. However, based on positive staining patterns along fibrils, axial fibrillar order was maintained (Bai et al., 1992) .

In Chapter 4, using a methodology similar to that employed in previous work (Hadley et al., 1998; Hadley et al., 2001), TEM study of the PTA positive staining of MTT was used to produce band intensity electron density data, the changes in which could be used to show some of the sites and distribution of glycation along the *D*-period. However the fibrils observed were removed from their tendon environment, chemically modified by the PTA stain and desiccated during the preparation procedure for the microscopy. No previous work has been done to compare electron density maps of glycated and non-glycated native collagen in intact collagen fibres from tendon or other tissues. The work presented in this chapter aims to produce electron density maps for glycated and control MTT incubated in buffer for the same period. An incubation time period of 2 weeks was chosen and the sugars and polyol selected for the glycation

incubations were galactose, ribose and sorbitol. This time period was considered both pragmatic and adequate. The two sugars and sorbitol had produced significant changes in TEM positive staining patterns over 2 weeks in Chapter 4, section 4.3.3, and both galactose and ribose had also produced notable changes in fluorescent AGE profiles over that time (Chapter 3, section 3.3, Figures 3.3B and 3.5B) whereas sorbitol had produced none (Chapter 3, section 3.3, Figure 3.1). It was anticipated that sorbitol would not produce significant changes in electron density within the fibril anywhere along the *D*-period and would therefore act as a viable control, as demonstrated previously by Tanaka et al. (1988a). At the time of the design of these experiments, it was decided that glycation with glucose and fructose would not be studied here due to the inherent limitations in the amount of synchrotron beam-time available. Glucose was excluded due to its relative lack of significant AGE formation (Chapter 3, section 3.3, Figure 3.2) and the absence of positive staining changes it produced (Chapter 4, section 4.3.3, Figure 4.9). Fructose was excluded in deference to faster-acting ribose, which unlike fructose, was seen to have produced considerable changes in the TEM positive staining pattern over the 2 week treatment period chosen (Chapter 4, section 4.3.3, Figures 4.11A and 4.12B).

The main aims of the work presented within this chapter are to test the following hypotheses. Firstly, that products of *in vitro* glycation in type I collagen of MTT tissue can be detected and located along the *D*-period by X-ray diffraction using heavy atom isomorphous replacement for phase determination, as a result of changes (increases) in electron density produced by AGEs. It is anticipated that these will show similarities to the corresponding positive staining changes reported in Chapter 4. Secondly, in the light of the significant changes in the positive staining pattern of fibrils following incubation in the buffered-saline incubation medium alone (i.e. the “buffer control”) reported in Chapter 4, it is anticipated that the electron density map of the buffer control produced here will show differences from that produced in the previous chapter.

6.2 Materials and methods

All glycation incubations were carried out in Dr Orgel’s laboratory within the Technology Park at IIT, Chicago, IL. As in the previous chapter, data collection experiments in this chapter were performed on the BioCAT (18-ID) beamline at the APS synchrotron at Argonne National Laboratory near Chicago, IL.

6.2.1 Materials

The solid incubation buffer ingredients, PIC solution and sugars were as described in Chapter 2, sections 2.1 and 2.2. Consumables used for the data collection experiments were as described in section 5.2.1 and the software used for data extraction were as described in Chapter 2, section 2.1 and Chapter 5, section 5.2.1.

6.2.2 Glycation incubation of MTT

The obtaining and dissection of mouse tails was performed as outlined in Chapter 2, section 2.3. The mouse tails were transported to the Orgel group's laboratory at IIT in the same manner as described for transport to APS, described in Chapter 5, section 5.2.2. The tails used in this chapter were also from 3 month and 2 day old male CD-1 mice. Strips of tendon of at least 4cm in length were removed, taking care to maintain the axial structural integrity of each section as much as was practical. Each length was placed in a Petri dish containing 0.9% saline solution with 1% v/v PIC added.

The incubations were carried out broadly as described in Chapter 2, section 2.4 using 10 ml each of the solutions described for the MOPS-saline buffer control and the 0.2M galactose, ribose and sorbitol preparations (see Chapter 2, section 2.2.1). Each was added to separate 50 ml conical-bottomed plastic screw-topped tubes. At least three separate MTT sections were added to each tube and the tubes sealed and mixed thoroughly ensuring the whole of every section was fully immersed in its solution. All the tubes were placed in a 37^{o}C water bath for 2 weeks.

At the end of the incubation period, each solution was carefully removed from its respective tube with a pipette and the MTT tissue remaining washed three times with 0.9% saline solution. A further 10 ml of the saline was added to each tube and the tubes kept refrigerated at 4^{o}C until being transported to APS, where they were again stored refrigerated until use.

6.2.3 Post-incubation preparation of MTT fibres for X-ray diffraction analysis

The isomorphous replacement procedures and installation of MTT fibres into the sample holder were performed broadly as described in Chapter 5, section 5.2.2. As a result of the weak iodination apparent in the previous chapter, the isomorphous

replacement protocol was varied inasmuch as 4 mg/ml and 10 mg/ml KI solutions in PBS were used separately here, both for 1 hour, for iodine isomorphous replacement. The Au derivative preparation procedure remained unchanged.

As all MTT samples had been stored post-glycation and transported in physiological (0.9%) saline, all non-isomorphically-replaced MTT samples were washed in PBS for at least 10 minutes prior to insertion into the sample holder. The isomorphically-replaced samples were briefly washed in PBS, as described in Chapter 5, section 5.2.2.

6.2.4 X-ray diffraction experiments and data collection

The beam line setup, placement of the sample holder in the beam line, alignment and data collection procedures were broadly as outlined in sections 5.2.3 and 5.2.4. Data were collected over time periods varying between 2 and 5 seconds and the aluminium filters were not utilised. The effective resolution of the detector used was 2048 x 2048 pixels, i.e. 2 x 2 binning, giving an effective pixel size of 80 μm..

6.2.5 Data analysis

The data were extracted from the X-ray diffraction patterns and the amplitudes measured and phases determined as described in section 5.2.5. Patterson and electron density plots were produced for the buffer control and sugar/sorbitol-treated MTT. These were studied to ascertain the extent to which changes in electron density resulting from glycation of MTT collagen could be detected using X-ray diffraction.

The electron density plot data from each of the three glycated groups were individually linearly scaled against the buffer control data and plotted on the same axes for comparison. Difference electron density traces were produced by subtracting the buffer control MTT electron density trace data from the glycated data. The locations and relative magnitudes of changes in electron density resulting from glycation could then be studied.

6.3 Results

6.3.1 Two week buffer control MTT incubation

Figure 6.1 shows meridional X-ray diffraction patterns recorded from non-isomorphically replaced MTT and iodine and gold treated samples. It can be seen that as with the non-incubated native MTT in Chapter 5 the relative intensities of each meridional reflection in the non-isomorphically replaced and gold-treated MTT patterns differed noticeably. Conversely, differences between the non-isomorphically replaced and both iodine treated MTT patterns are less obvious.

In Figure 6.2 can be seen meridional data sector-extracted from the non-isomorphically replaced 2 week buffer control MTT samples. In comparison with the data presented in the previous chapter (section 5.3.2, Figure 5.9) it is noticeable that the patterns are less intense following the buffer incubation giving smaller peak heights. Therefore fewer reflection peaks could be measured. As few peaks could be located reliably and measured accurately beyond the 22nd order reflection for this group it was considered appropriate to extract peak heights from the reflections for meridional orders 1 to 22.

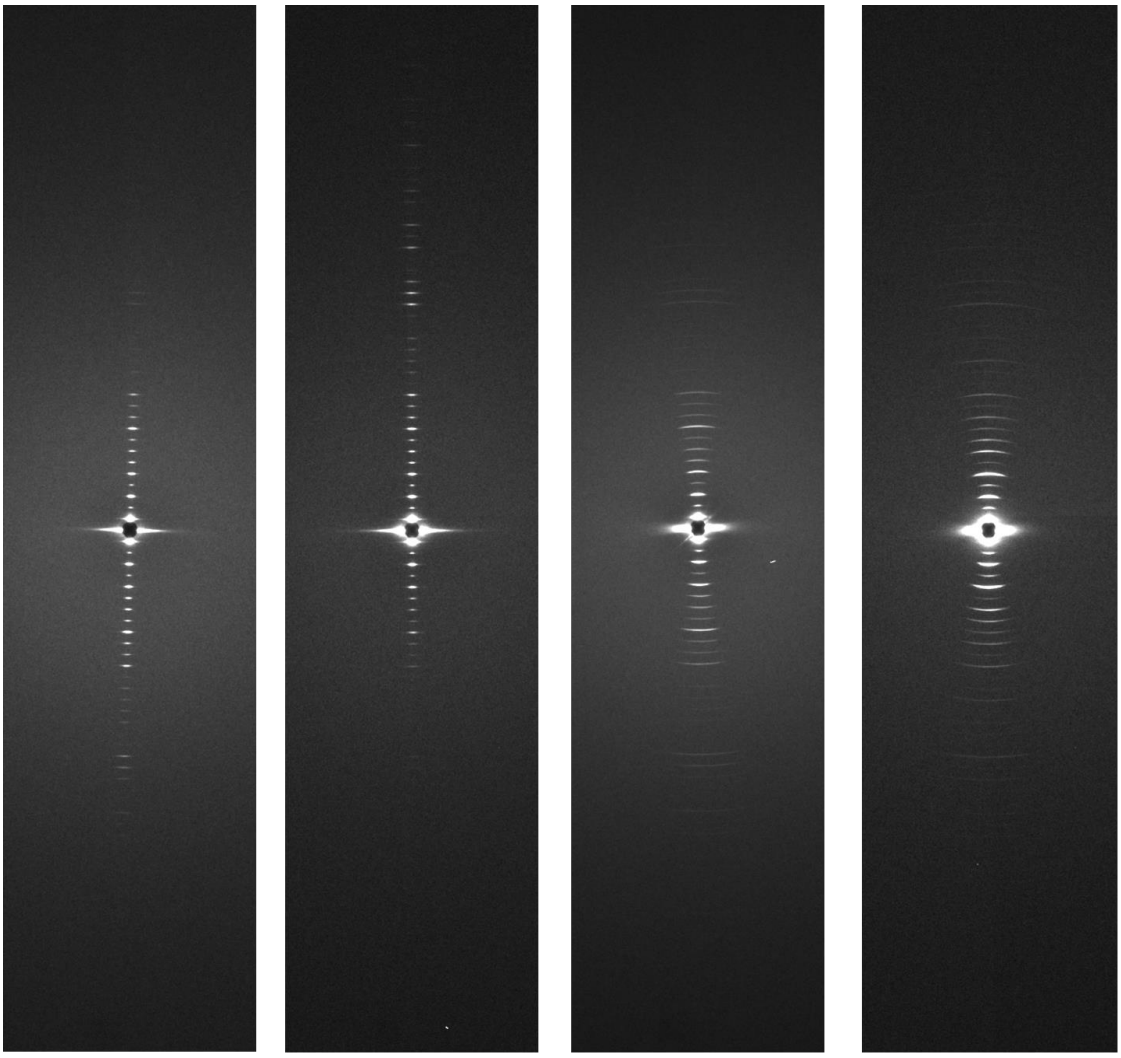


Figure 6.1 Meridional X-ray diffraction patterns recorded from MTT incubated for 2 weeks in MOPS-saline buffer (the buffer control). All patterns are rotated 90^{o} to the right to be consistent with convention. For comparison, the non-isomorphically replaced MTT pattern is shown between that obtained from iodine isomorphically-replaced tissue prepared using 10 mg/ml KI in PBS and the gold isomorphically-replaced tendon prepared as described previously. In addition, a pattern recorded recorded from a further MTT sample iodinated using 4 mg/ml is shown on the left. The light specks down and to the right of the beam stop (dark region in the centre of the pattern) in the non-isomorphically replaced pattern and at the bottom of the iodine isomorphically replaced pattern using 10 mg/ml KI, are spurious noise and could be masked prior to data extraction using FIT2D.

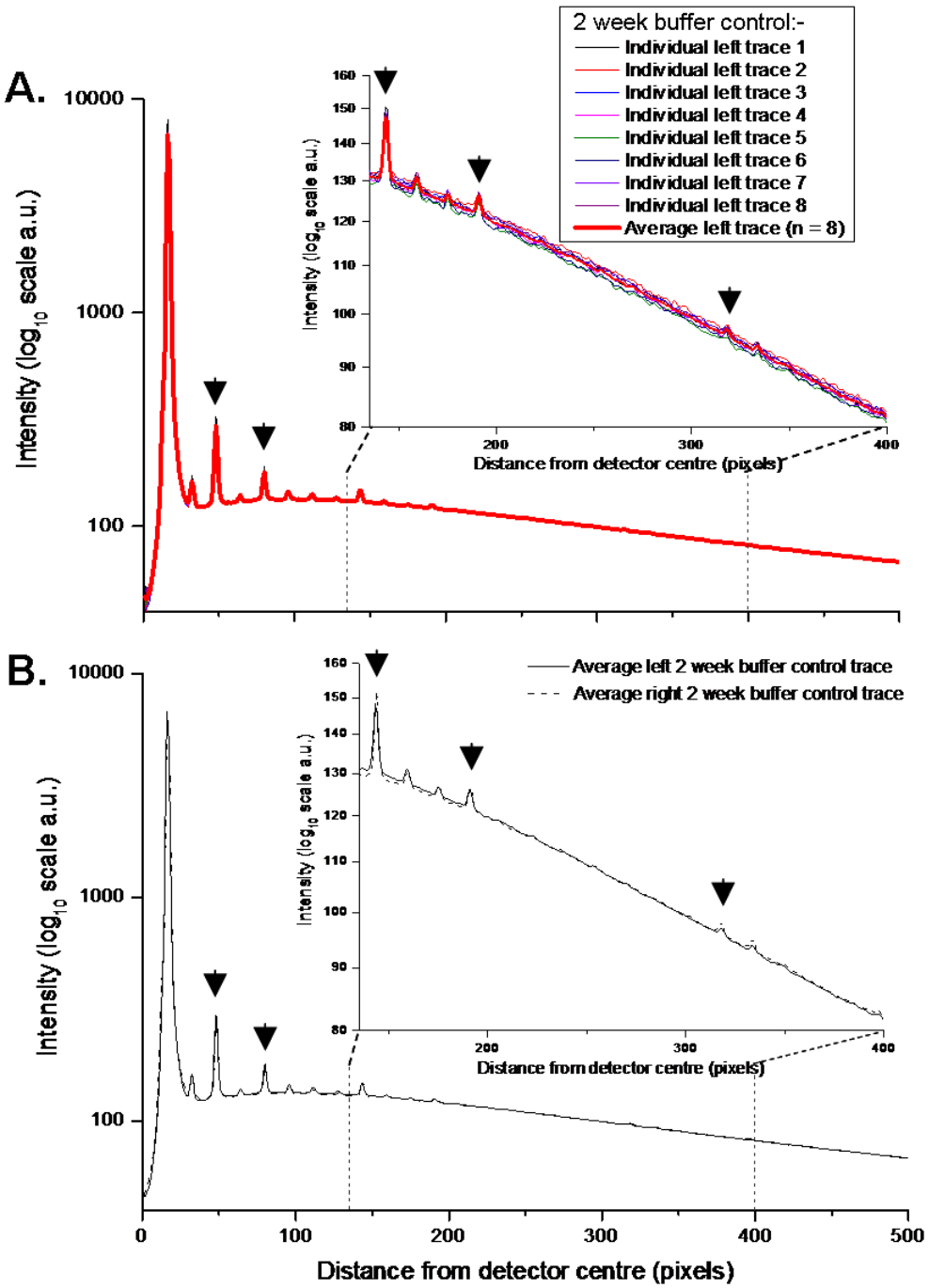


Figure 6.2 From the non-isomorphically replaced 2 week buffer control MTT samples, **A.** Individual X-ray diffraction meridional pattern traces taken from one side (left, or upper as seen in Figure **6.1**) of the detector on a semi-log plot. The overlaid thick red line is the average (n = 8). **B.** The average meridional traces taken from both halves of the X-ray diffraction pattern. The left trace (solid line) is from the same data as in A. For the right trace, n = 7 individual traces. For both A and B, the insets show magnifications of the visible reflection peaks beyond the 9th order reflection and the black arrows indicate from left to right, orders 3, 5 (main plots), 9, 12 and 20 (inset magnified plots).

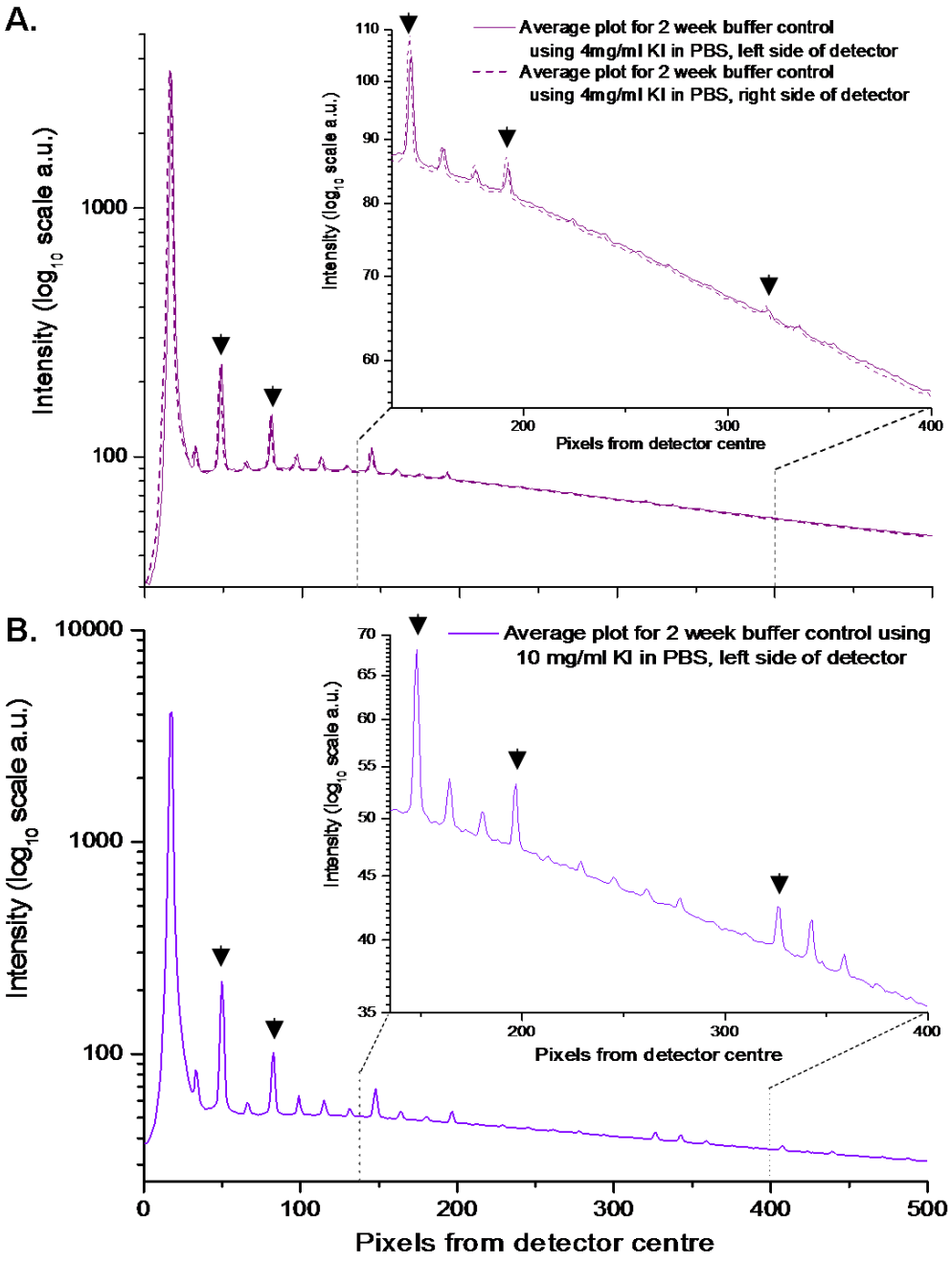


Figure 6.3 Semi-log plots of X-ray diffraction average meridional data for 2 week buffer control MTT isomorphically replaced with iodine **A.** using 4 mg/ml KI solution in PBS (n = 6 for left data, n = 7 for right data) and **B.** using 10 mg/ml KI solution in PBS (n = 8). For both A and B, the insets show magnifications of the visible reflection peaks beyond the 9th order reflection and the black arrows indicate from left to right, orders 3, 5 (main plots), 9, 12 and 20 (inset magnified plots).

As stated previously (section 5.4), data from earlier experiments examining use of different concentrations of KI in PBS in order to optimise and improve the iodine-isomporphous replacement procedure were lost. Here again iodine isomorphous

replacement was tried using higher concentrations of KI, i.e. 4 and 10 mg/ml, in PBS and the integrated intensity traces of the meridional data from these are shown in Figure 6.3. There was an asymmetry about the 10 mg/ml KI derivative patterns inasmuch as the upper (left) half of it was more intense than the lower (right) half for all the data collected. Therefore only the upper data were used. The 10 mg/ml KI-treated tendon upper patterns showed a clearer resolution of the higher orders as compared with the 4 mg/ml KI-treated buffer control MTT collagen, making peak height measurement more accurate and less susceptible to noise. Therefore it was decided that the 10 mg/ml KI be used for iodine isomorphous replacement.

6.3.1.1 Amplitude determination

Figure 6.4 shows plots of the amplitudes for each isomorphous treatment group following $\log_{10}$ scaling of the duplicate data sets relative to one another, within the non-heavy atom and gold treatment groups and then Lorentz correction in all three groups. Even with the omission of duplication for the 19^{th} order reflection for the non-isomorphically replaced MTT group and therefore no precision data (so an effective %CV value of 0%), the Au treatment showed a marginally higher precision (sum of %CVs = 114.2) than the untreated group (sum of %CVs = 120.2). Presented in Figure 6.5 are the scaled and Lorentz-corrected amplitude data for the iodine and gold isomorphous replacement treatment groups compared in turn with the untreated 2 week buffer control MTT data. Generally the positions of the gold amplitudes relative to the “no heavy atom treatment MTT” are similar to those seen with the freshly dissected MTT in the previous chapter. With the iodine treatment, in this case the amplitudes can be seen to show greater variation as compared with the non-isomorphically replaced MTT. It is reasonable to consider that the isomorphous replacement with iodine has occurred more strongly in these experiments and using the greater concentration of KI solution than was seen previously, especially for orders 13, 15, 17 and above. This includes orders 14 to 18 that are most associated with the non-helical telopeptide regions (Brodsky et al., 1982). By virtue of these regions containing all but one of the tyrosine residues in the collagen molecule, these would bind much of the iodine that isomorphically attaches to collagen.

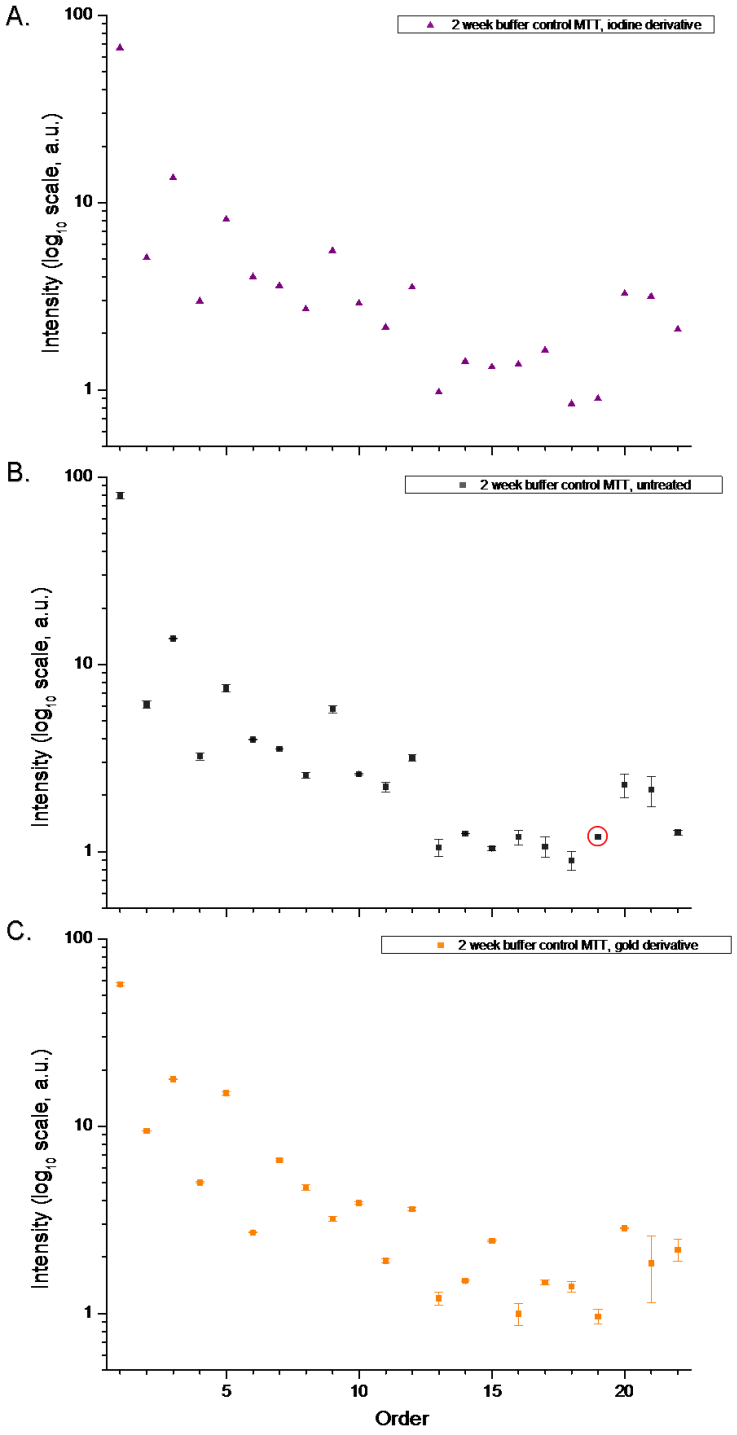


Figure 6.4 Semi-log plots of scaled, Lorentz-corrected amplitude data for 2 week buffer control MTT samples for orders 1 to 22, for **A.** iodinated using 10 mg/ml KI, **B.** non-isomorphically replaced and **C.** Au-treated samples. Error bars represent 1 SD either side of the mean of duplicate (left and right average) values, except for the iodinated data (**A**) where only the pattern on the left side of the detector was used, and the 19th order of the non-isomorphically replaced MTT (**B**), circled in red, where only an intensity value from the right side of the detector could be measured with any confidence. Numbers of individual peak traces upon which the average traces were based: **A.** left n = 8, no right data used; **B.** left n = 8, right n = 7; **C.** left n = 14, right n = 12.

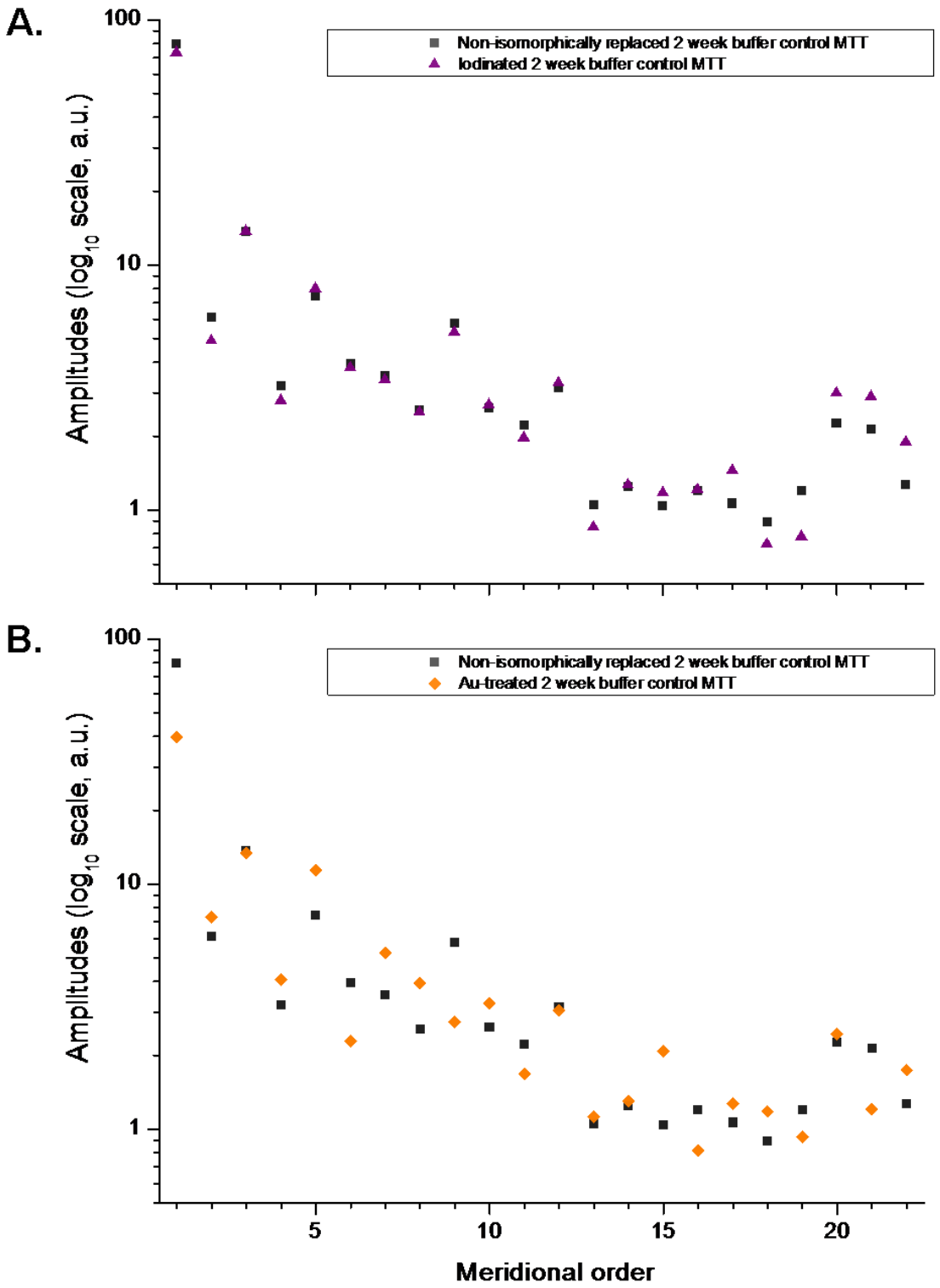


Figure 6.5 Semi-log plot comparisons of the mean amplitudes for the 2 week buffer control non-isomorphically replaced and **A.** iodine and **B.** gold heavy atom treatment groups for orders 1 to 22. The heavy atom amplitudes were $\log_{10}$ scaled against the non-heavy atom treated amplitudes and then each dataset was Lorentz-corrected.

6.3.1.2 Patterson plots

Patterson plots for the 2 week buffer control isomorphous treatment groups are shown in Figure 6.6. It can be seen from these and most especially when the first order is excluded, that both of the heavy atom derivatives are distinguishable from the non-treated samples. Therefore the iodination using 10 mg/ml KI solution in PBS appears to have brought about stronger binding to the MTT collagen than the 1 mg/ml solution employed for the native, freshly prepared tissue in the previous chapter.

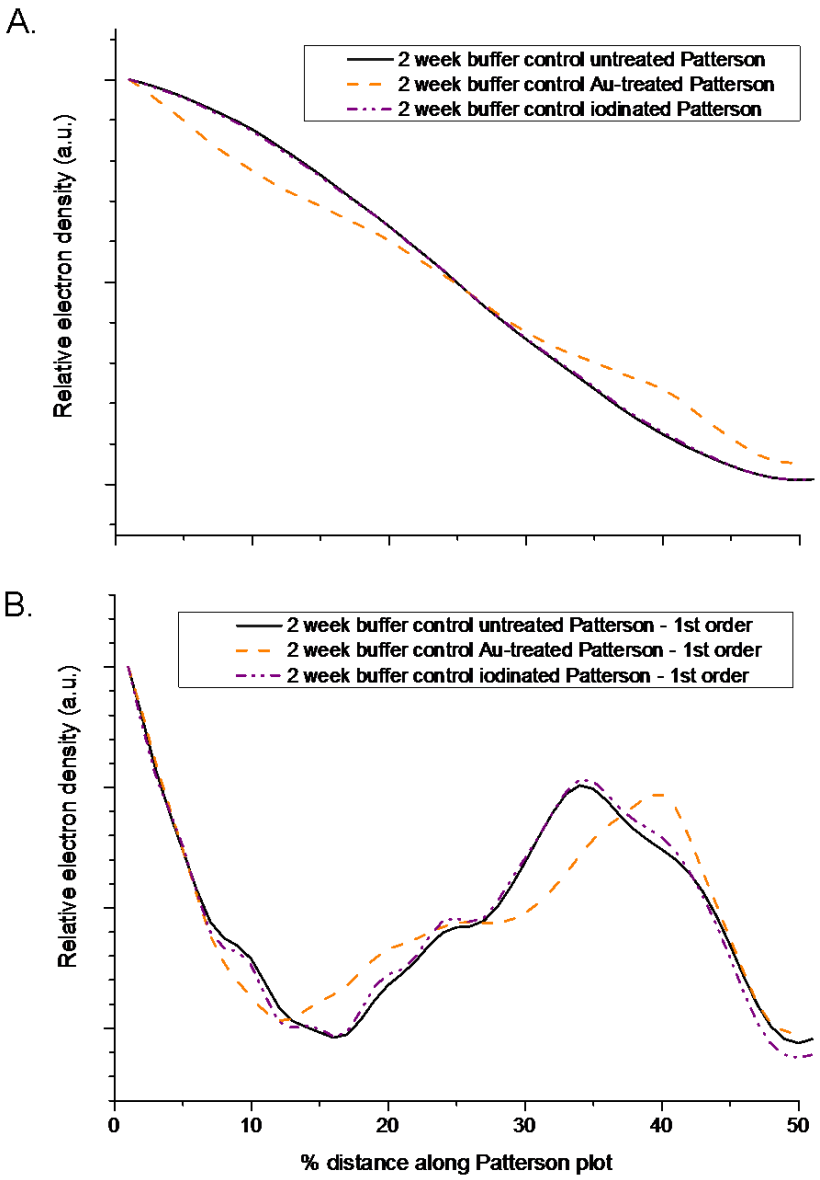


Figure 6.6 Comparison of the untreated, Au-treated and iodinated Patterson plots for the 2 week buffer control MTT Fourier synthesised from **A.** orders 1 to 22 and **B.** orders 2 to 22, i.e. excluding 1st order. As entire Pattersons are always centrosymmetric only half of the plots are shown, for clarity.

Heavy atom difference Patterson plots for iodine and gold binding to 2 week buffer control MTT collagen are shown in Figure 6.7 and Figure 6.8 respectively, in the same manner in which they were presented for freshly-dissected tissue in Chapter 5 (section 5.3.3, Figure 5.17 and Figure 5.18). Although it can be seen that there are strong similarities between the experimental and predicted Patterson plots for both heavy atoms, the experimental plots do differ from those presented for fresh, non-incubated MTT collagen.

In the case of iodine, whilst the experimental peaks at 5%, 41% and 46% are in better alignment with their predicted peaks, all peaks between 10% and 29% have been shifted further left with respect to their corresponding predicted peaks as compared with the fresh, non-incubated MTT collagen. It is also notable that in terms of peak heights, there is a lesser predominance of the experimental peaks here associated with the spacings between the non-helical collagen termini/telopeptide regions at 35%, 41% and 46% when compared with the other peaks further left.

With respect to the gold difference Pattersons the predicted plots were shifted 1.8% left relative to the experimental here (compare with 2.8% for the non-incubated, fresh MTT collagen) to achieve the closest fit. There are two striking changes brought about by a 2 week incubation in the buffer solution. Firstly the large experimental Patterson peak at 40% has shifted 5% to the right and away from its predicted position. In the case of the histogram-predicted Patterson, this brought it into alignment with a smaller predicted peak absent from the Fourier-predicted Patterson plot. As in the case of the iodine difference Patterson plots, this may suggest alterations in the conformations or the spacings between the telopeptide regions after 2 weeks incubation of the MTT in the buffer solution. The relatively unchanged peak height relative to the other peaks makes changed affinity for the methionine and histidine residues for the gold atoms appear less likely. Secondly, there have been significant changes in the original four peaks sitting between 10% and 30% on the original non-incubated MTT Patterson. The peak at 15% has disappeared with the peak at 20% having become smaller and flatter. Conversely, the experimental Patterson peaks originally sitting at positions 11% and 25% have shifted approximately 2% to the right and become more prominent in height relative to the rest of the plot. They remain identifiable as being consistent with peaks present on the predicted Patterson plots produced using both methods.

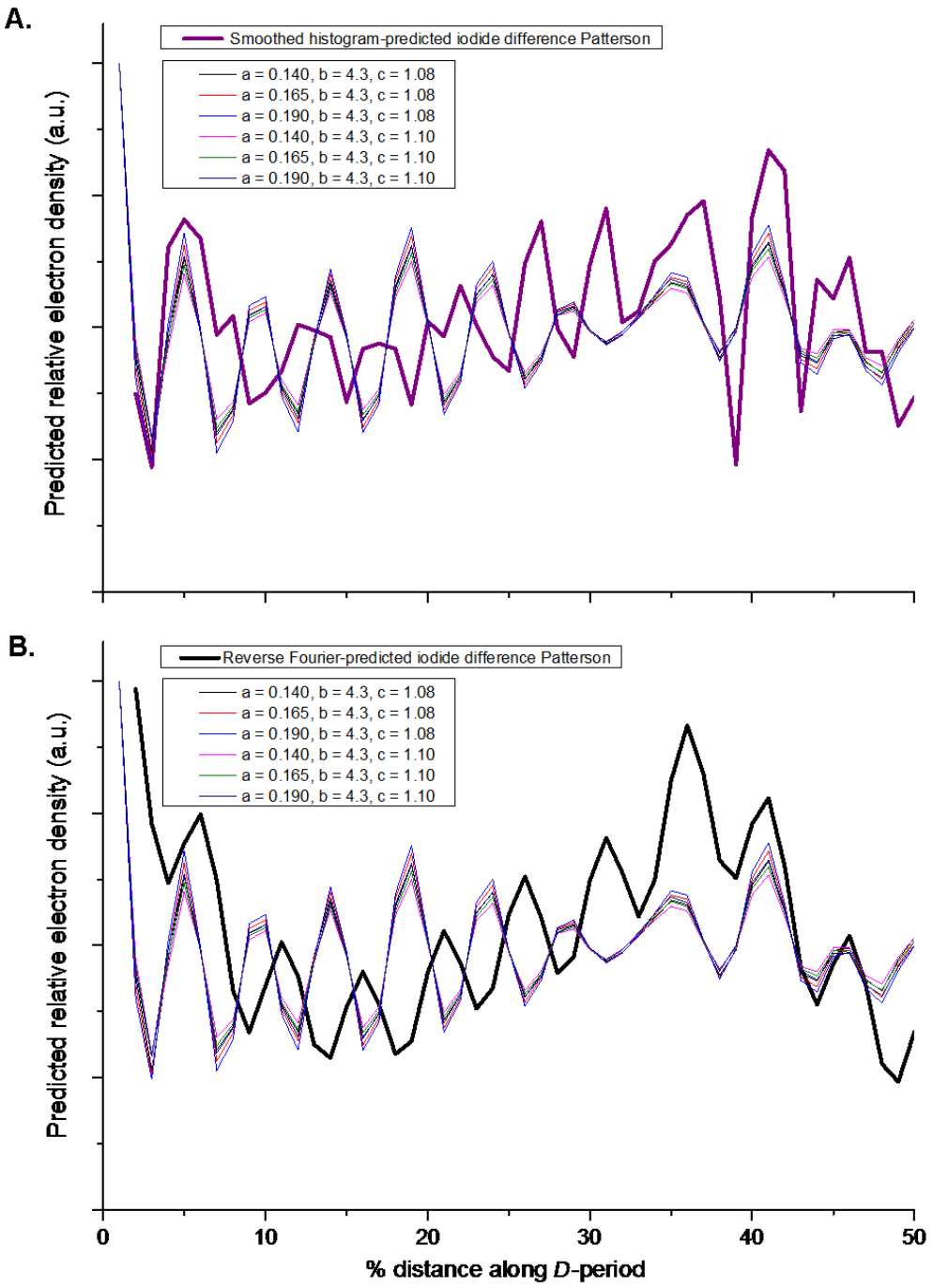


Figure 6.7 Two week incubation MTT buffer control experimental iodine difference Patterson plots, based on subtractions of non-isomorphically replaced MTT amplitude data from iodine isomorphically replaced native MTT amplitudes after scaling of both using the scaling factors (a,b,c) given. By comparison with the predicted Patterson plots, the best six sets of scaling factors (insets) for the experimental data were selected. **A.** Scaled experimental traces overlaid with smoothed histogram-predicted Patterson plot, and **B.** the same experimental traces compared with the Fourier-predicted Patterson.

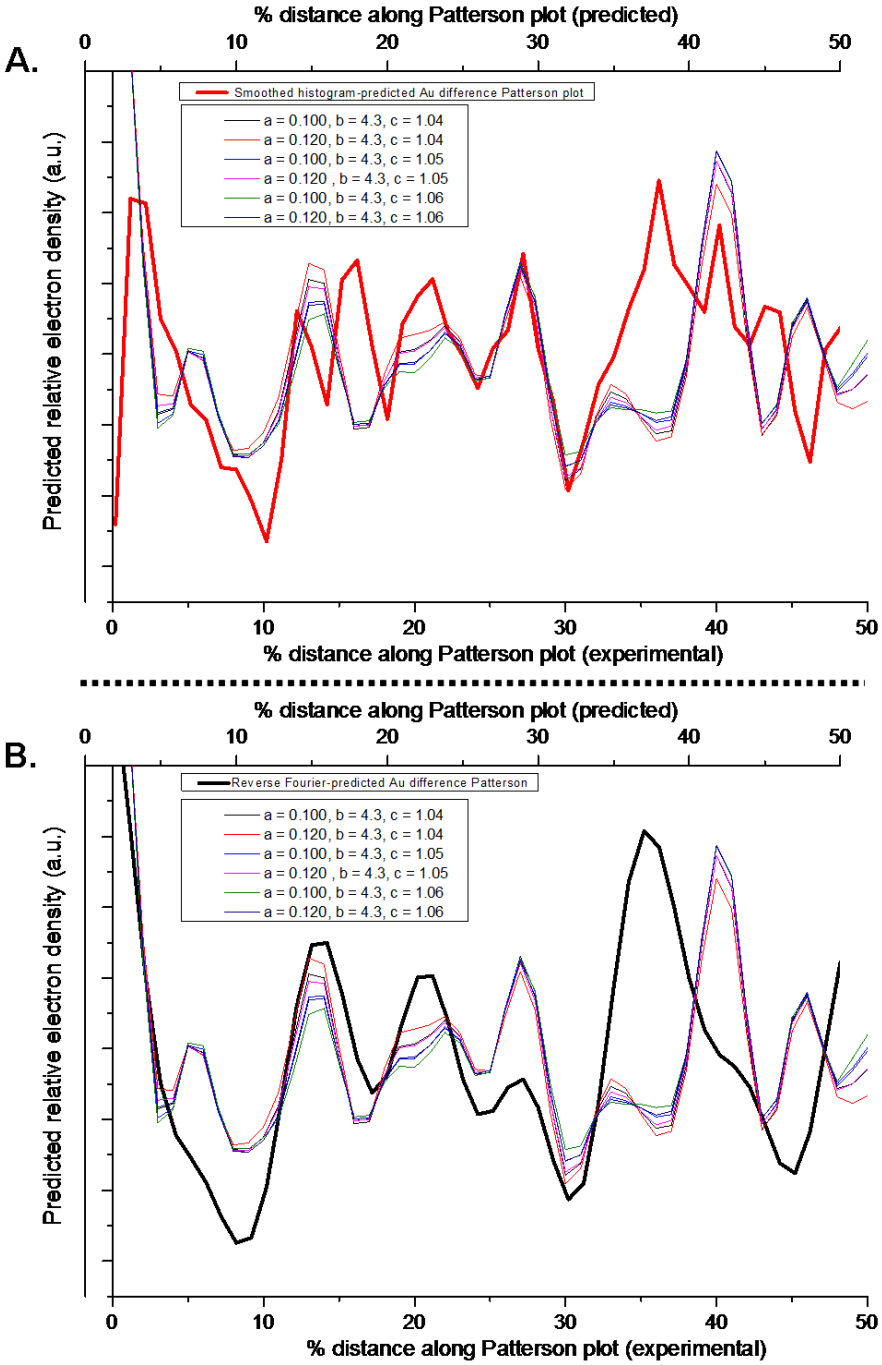


Figure 6.8 Two week buffer control MTT experimental gold difference Patterson plots, analogous to the iodine difference data shown in Figure **6.7**. The experimental traces shown were produced using the best sets of scaling factors (a,b,c) as before. **A.** Experimental data compared with smoothed histogram-predicted Patterson plot, and **B.** compared with Fourier-predicted Patterson. As there seemed to be an overall left shift of the experimental data as compared with the predicted away from visual best fit in both cases, the predicted Patterson plots have been shifted left by 1.8%.

6.3.1.3 Phase determination

The phases determined for the 22 reflections used, using the scaled heavy atom derivative and untreated MTT are shown in Table 6.1 (see also, Appendix 3 for the *individual* values). The rationale and criteria used for phase selection were as described in Chapter 5, sections 5.2.5.5 and 5.3.4.

Order	First preference phase (mean, in radians)	Number of individual phase values	% error	Second preference phase (mean, in radians)	Number of individual phase values	% error
1	2.9792	4	2.1	4.9016 †	2	1.1
2	0.3502 †	7	2.6	4.2701	2	0.5
3	4.7965 †	3	0.8	N	-	-
4	3.3689 †	5	3.9	N	-	-
5	4.3452 †	3	0.9	N	-	-
6	1.5294 †	2	1.0	0.5094	2	1.0
7	3.3962 †	5	2.6	N	-	-
8	1.8514 †	2	1.6	5.4735	2	1.4
9	0.8802 †	4	1.1	N	-	-
10	1.8060 †	6	2.4	N	-	-
11	4.3373 †	4	1.2	N	-	-
12	5.1080 †	3	0.4	N	-	-
13	0.5656 †	3	0.4	4.3739	3	0.1
14	2.1234 †	2	0.1	2.6531	2	0.8
15	3.6157 †	3	1.3	N	-	-
16	4.9404 †	5	3.2	N	-	-
17	2.0123 †	5	4.2	N	-	-
18	5.0334 †	5	2.3	N	-	-
19	6.2125 †	4	2.5	N	-	-
20	3.8841 †	4	1.3	N	-	-
21	3.4242 †	2	1.8	5.1376	3	3.9
22	5.2646 †	3	1.4	1.2088	3	2.2

Table 6.1 Relative phase probabilities determined using Harker plots for the 2 week buffer control MTT and based on the difference Patterson-scaled amplitude values of the two heavy atom isomorphous MTT derivatives and the non-isomorphically-replaced MTT tissue. As previously, bar plots were used to visualise phase probabilities. The mean phase probabilities given in the left half of the table were considered the most likely phases with possible alternatives given in the right half. *n* is the number of individual phases used to calculate each mean phase. The "% error" for each phase is a measure of precision and is the sample-based SD divided by 2π. † The phases that were used in the eventual electron density map.

From this table, it can be seen that only one of the phases used in the electron density maps for the 2 week buffer control MTT collagen was not a first preference phase. This was that for the 1st order reflection. The level of confidence in producing this set of phases was higher than for the native, untreated MTT collagen presented in Chapter 5.

6.3.1.4 Electron density map for the *D*-period of the two week MTT buffer control collagen

The electron density trace of the two week buffer control MTT collagen *D*-period is shown in Figure 6.9 and was constructed using a Fourier synthesis using the structure factors shown in Table 6.2. The gap and overlap regions are marked on the trace. Based on this the ratio of the lengths of the gap and overlap regions is 0.53:0.47, slightly discrepant with the values of 0.52:0.48 reported in the previous chapter for untreated, native MTT and consistent with the similar values stated previously by other workers (Brodsky et al., 1982; Fraser et al., 1987; Bradshaw et al., 1989; Orgel et al., 2000).

In Figure 6.10 is shown the buffer control electron density map overlaid with the electron density map of the freshly dissected, untreated MTT in two different ways, for comparison. It can be seen from these that whilst the same axial gap:overlap length ratio is maintained, the positions of the peaks representing higher electron density have shifted relative to the basic *D*-period structure. Therefore it is possible to align the two maps either primarily on the basis of the gap-overlap unit cell structure (Figure 6.10A) to produce the best correlative alignment (Pearson's r value = 0.954, R^2 = 0.910) or visually on the basis of the electron density bands (Figure 6.10B), ignoring the resultant discrepancy in the gap-overlap boundary positions between them. This latter arrangement shows two major differences between the two plots. Firstly, an apparent retraction of the N-terminal telopeptide end of the molecules/overlap region and secondly, an apparent increase in the electron density of and broadening of the C-terminal telopeptide end.

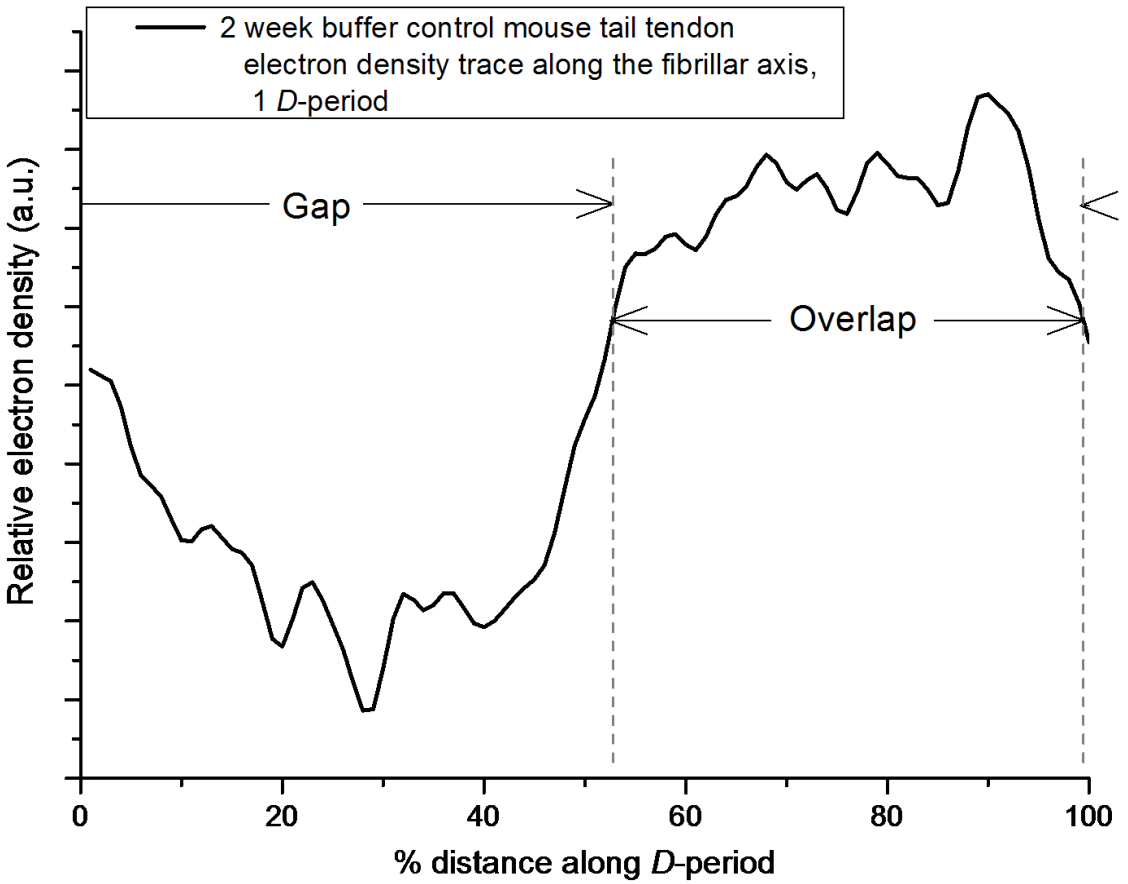


Figure 6.9 Electron density map of MTT type I collagen after 2 weeks incubation in buffer solution. This was constructed using structure factors from the meridional reflections, orders 1 to 22 resulting in a theoretical resolution of 3.05 nm equivalent to a translation distance of 10.6 amino acids. Orientation of trace: N → C, left to right.

Order	Structure factor *amplitudes*	Structure factor *phases*
1	79.4	4.9016
2	6.09	0.3502
3	13.7	4.7965
4	3.22	3.3689
5	7.45	4.3452
6	3.96	1.5294
7	3.53	3.3962
8	2.55	1.8514
9	5.76	0.8802
10	2.59	1.8060
11	2.21	4.3373
12	3.16	5.1080
13	1.05	0.5656
14	1.25	2.1234
15	1.04	3.6157
16	1.19	4.9404
17	1.06	2.0123
18	0.89	5.0334
19	1.19	6.2125
20	2.26	3.8841
21	2.13	3.4242
22	1.26	5.2646

Table 6.2 Structure factors used in the Fourier synthesis of the electron density map for one *D*-period of MTT collagen following a two week incubation in glycation buffer. The amplitude values are Lorentz-corrected arbitrary units and the phases are in radians.

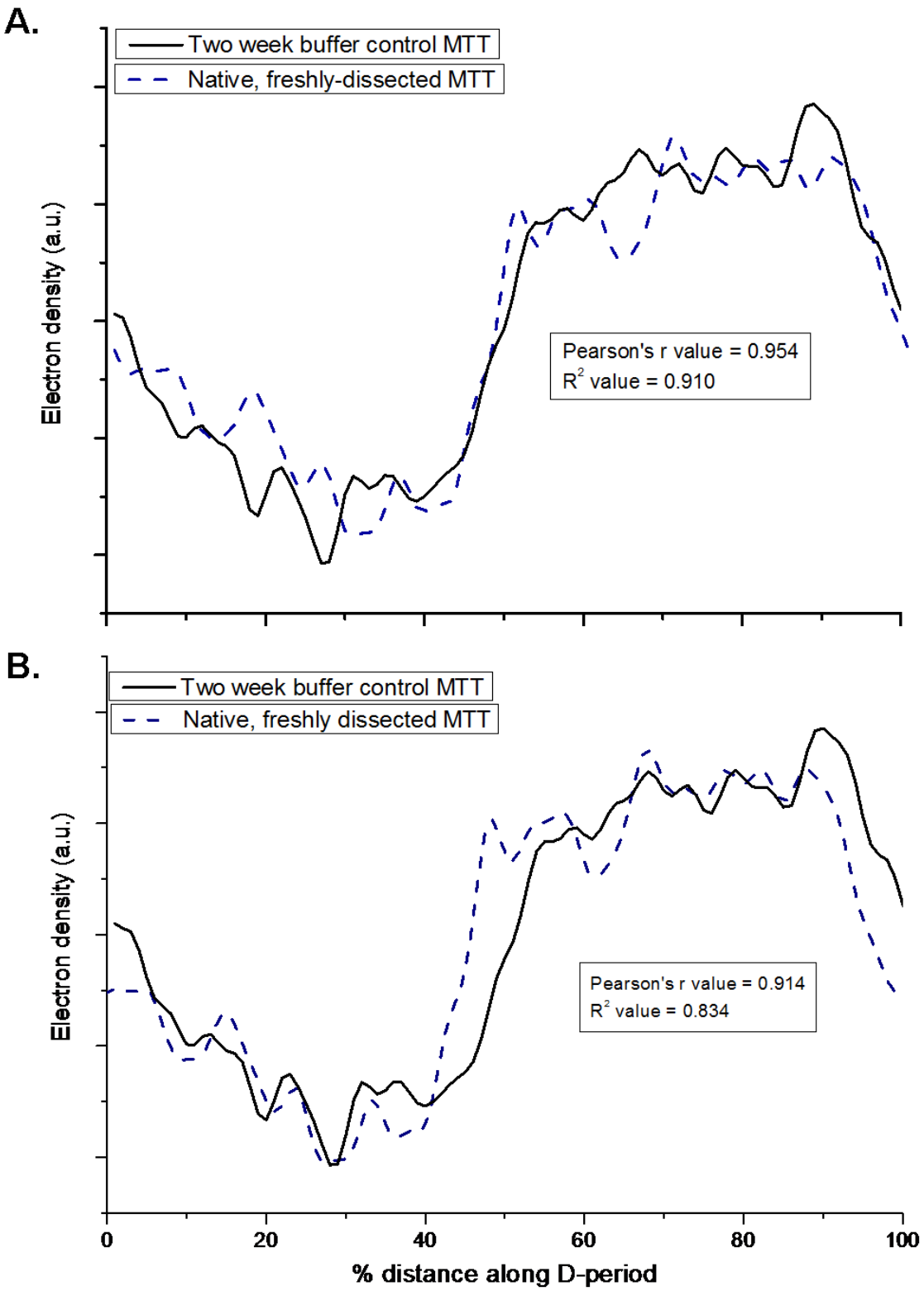


Figure 6.10 Axial electron density plot of one *D*-period of MTT collagen post-two week incubation in glycation buffer overlaid with the untreated MTT electron density trace constructed in Chapter 5 **A.** to best alignment according to cross-correlation r-value and **B.** to best visual alignment of detail, disregarding rigid adherence to the locations of the gap-overlap boundaries.

The two week buffer control plot is presented in Figure 6.11 aligned with the positive staining band identification plot. It can be seen that the higher electron density peaks broadly coincide with the positions of these bands. The major inconsistency when comparing the buffer control and untreated MTT plots as aligned to the band identities on this plot is the relative lack of a *c2* band electron density component in buffer-incubated collagen. Closer inspection of this suggests the possibility that the buffer incubation has caused the *c1-c2* bands to migrate towards the *b2* band, the *c1* band forming a shoulder at the N-side of the principal *b2* band's peak at a 38% translation distance along the plot. Therefore the three *c1-c2* band peaks have shifted 5-6% N→C during the buffer incubation.

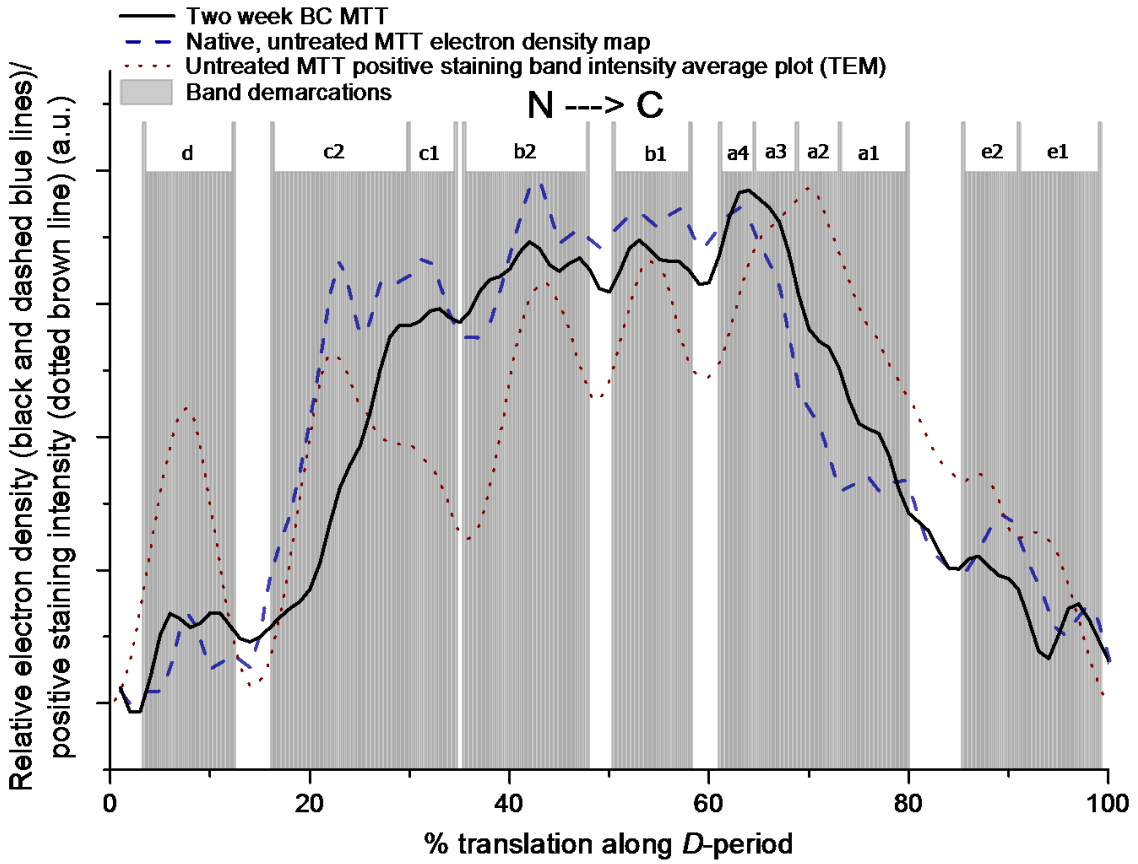


Figure 6.11 Two week buffer control MTT collagen electron density plot overlaid with positive staining band locations. Traces for untreated, fresh MTT collagen – visually aligned for detail as in Figure 6.10B - and the experimental positive staining band intensities for native, non-incubated MTT from the TEM work presented in Chapter 4 are also shown for comparison.

Another striking change is that there appears to have been a general increase in the relative electron density of the *a* bands as compared with the rest of the overlap region,

with incubation. Other notable changes are an apparent flattening and splitting into two peaks, of the *e2* band and a similar splitting of the *d* band.

As in the case of the native, untreated MTT, predicted models were optimally aligned and fitted to the 2 week buffer incubation MTT electron density map and these are shown in Figure 6.12. The model being compared in Figure 6.12A is the one found to best fit the untreated MTT previously, with unfolded, extended α1-chains in the N-telopeptide region and one folded and one extended α1-chain at the C-telopeptide termini of the molecules. It can be seen that adjustment of the conformation of the N-telopeptide ends of the molecules in the spreadsheet model would require increasing the level of folding in this region increasing the electron density at the N-termini of the molecules. This clearly would not improve the fit. Conversely as seen in Figure 6.12B, adjusting the model such that both α1-chains in the C-telopeptide regions of the molecules are folded as suggested by Orgel et al.'s (2000) work, addressed the increased magnitude of the electron density seen for the two week buffer control MTT and produced a visually better fit. However inconsistencies between model and experimental plot still exist, most obviously the apparent experimental contraction of the *c* bands inwards at the N-terminus of the overlap region and the apparent increased breadth of the *a* bands as compared with the all-folded C-telopeptides model.

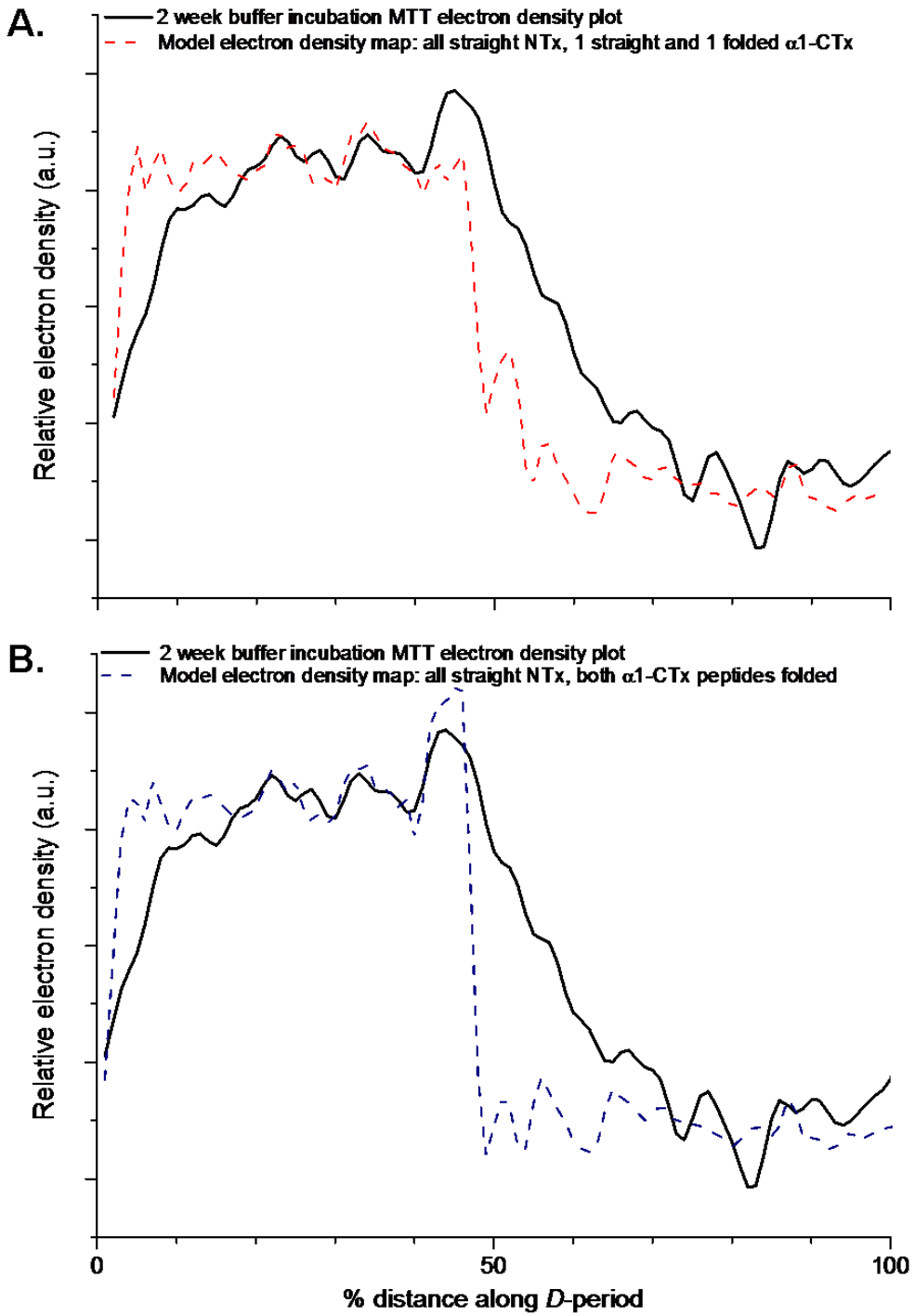


Figure 6.12 Two week buffer incubation MTT collagen electron density map compared with: **A.** predicted electron density model map with all three N-telopeptide terminal regions straightened (unfolded) in each molecule and one α1-chain straightened and one folded in the C-telopeptide termini of each molecule, as found to fit best with the fresh, untreated MTT collagen experimental data in Chapter 5; and **B.** predicted electron density with all three N-telopeptides extended and both α1-peptide C-telopeptide chains folded as per modified model from Chapter 5 (see section 5.3.5 Table **5.5**) and as originally suggested by Orgel et al. (2000). For both predicted models presented here, gap: overlap ratios = 0.55:0.45. The 2 week buffer control plot is shifted along the abscissa for optimal alignment with the models. N→C is left to right.

In an attempt to gain more information about changes in the relative distances between various loci along the *D*-period, heavy atom electron density plots were produced based on the amplitudes used in the scaling difference Patterson plots and the non-isomorphically-replaced buffer control MTT phases. These are shown aligned with the expected amino acid binding sites for both heavy atoms in Figure 6.13 according to the extended N-telopeptides, folded α1-C-telopeptides model mentioned above. The gold plot in Figure 6.13A seems to indicate three of the five most significant experimental peaks showing a notable discrepancy with their predicted positions. The associations of these experimental and predicted peaks (as they are thought to be) are indicated with arrows. When compared with the comparable plot for native, non-incubated MTT (Figure 5.24) the peaks here appear less nicely aligned with the predicted binding positions. The peak at 7% and the broad split peak at 36% (centre position) show a spacing of 29%, which is a modest reduction in comparison to the 31% seen in the corresponding non-incubated MTT plot so consistent with some degree of contraction within the overlap region. However, the peak at 61% does appear to have shifted 6% to the right as compared with the untreated MTT plot, contrary to what might be predicted on study of the equivalent *e* band positions in Figure 6.11. For the iodine electron density plot (Figure 6.13), the fit appears to be rather better, though a greater proportion of the amino acid binding sites are located in the peptide terminal regions with only one tyrosine not being located in a telopeptide.

Neither heavy atom difference plot adds to the weight of evidence as to the relative changes in peak positions seen in the electron density map for the 2 week buffer-incubated MTT as opposed to the untreated, though it is not electron density *per se* that is being represented here. In addition, as cross-phasing has been used (Orgel, 2000), i.e. the native phases, although the phase differences between non-derivatized and isomorphically replaced MTT tended to be relatively small, a degree of inaccuracy and approximation has been introduced as a result. Both the gold and iodine plots do show a significant broadening of the peaks associated with the C-telopeptide regions (36% median position on the gold plot, 39% median position on the iodine plot) post-buffer incubation, consistent with the increased breadth of the electron density peak associated with the *a* bands seen in Figure 6.11.

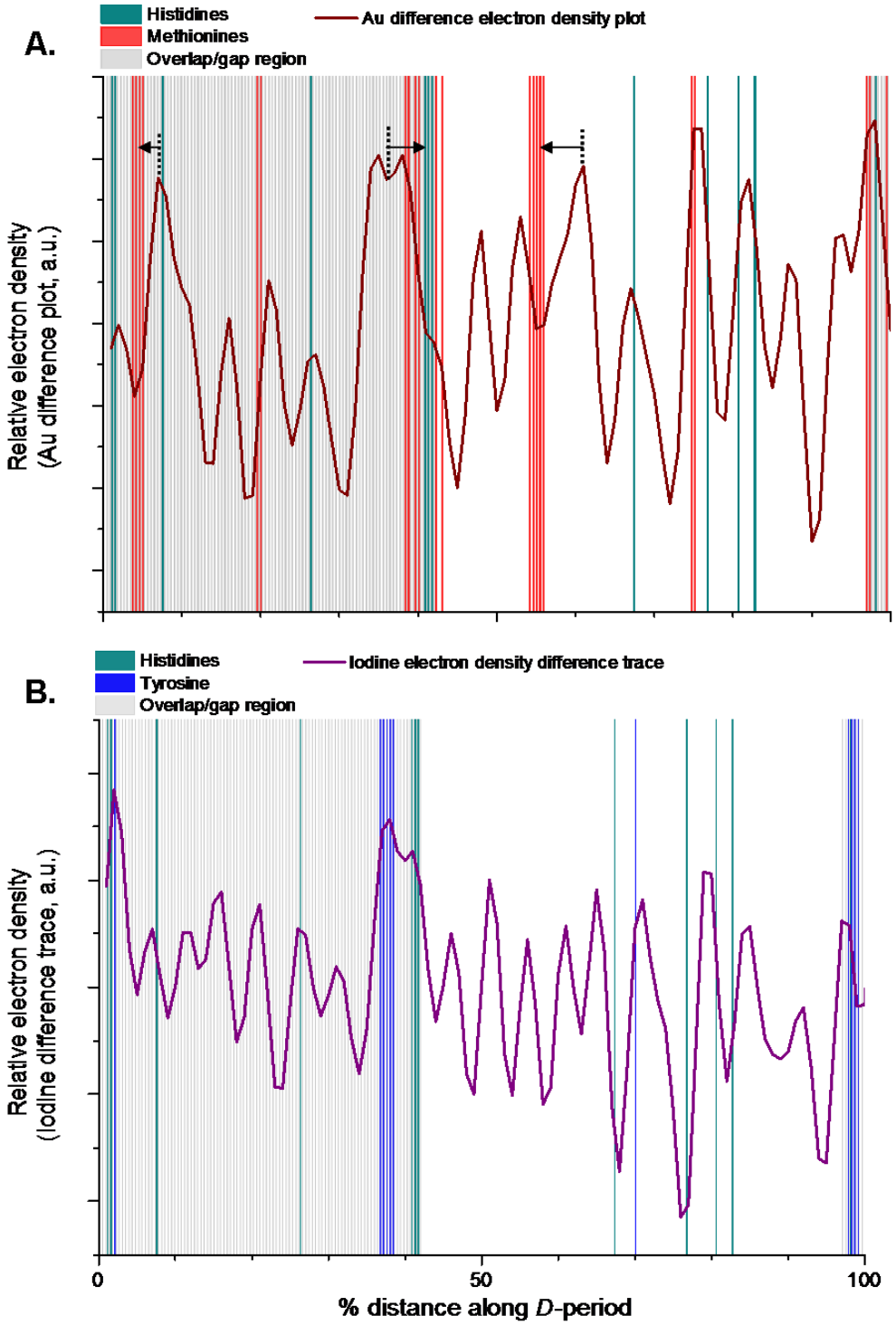


Figure 6.13 Heavy atom difference plots of electron density for 2 week buffer control MTT, isomorphically-replaced with **A.** gold and **B.** iodine. These are overlaid with the expected amino acid binding positions according to the model with extended (non-folded) N-telopeptides and both α1-C-telopeptide peptide regions folded as per the original model presented (see Orgel et al. (2000)), with constant amino acid translational distances along the *D*-period. In **A.** the arrows indicate the modelled peaks that the Au-experimental peaks are thought to be associated with. In **B.** the iodine atom trace appears to fit the atom's major binding regions more precisely. N→C direction is left to right.

6.3.2 Effect of glycation on the X-ray diffraction data

Following 2 week incubations of MTT in the buffer-saline solution and separately in the three 0.2M sugar/polyol solutions in the same buffer, there was no discernable evidence of equatorial reflections in the X-ray diffraction patterns detected. Meridional X-ray diffraction patterns were recorded as shown in Figure 6.14. Broadly, few differences were visually detectable between these except for some degree of faint reflection arc-lengthening, most obvious with some of the lower and more intense order reflections in the diffraction pattern. This occurred despite the MTT sections being held under tension in the sample holder during data collection. Although it was noted previously that the number of continuous order intensities it was possible to measure accurately with any confidence had been reduced from 30 to 22 by incubation of the MTT tissue in the buffer solution, inspection of these patterns does not indicate that addition of a 0.2 molar concentration of any of the sugars resulted in a further reduction of this number.

As for the buffer control and its heavy atom derivatives presented above, reflection intensities and therefore amplitudes were measured for all order reflections between 1 and 22. Figure 6.15 shows comparisons of meridional amplitudes of the MTT incubated for 2 weeks in the respective sugar solutions. In the case of the galactose incubation, there appear to be decreases in amplitude relative to the buffer control, for reflection orders 2, 13 and 19 with a marginal change for order 11. Conversely, there are increases for orders 15, 17 and 22 with another marginal change for the 10^{th} order. For the ribose MTT, orders 2, 3, 4, 13, 17 and 19 show a decrease with the first order also showing a small decrease. Orders 5, 6, 7, 10, 14, 16, 18 and 22 are increased for the ribose incubation, in addition with order 8 showing a less significant increase. Sorbitol showed the fewest discrepancies with as compared with the buffer control MTT, showing amplitude decreases for orders 1 and 2, and amplitude increases for orders 15 and 22, with further smaller increases for orders 7 and 10.

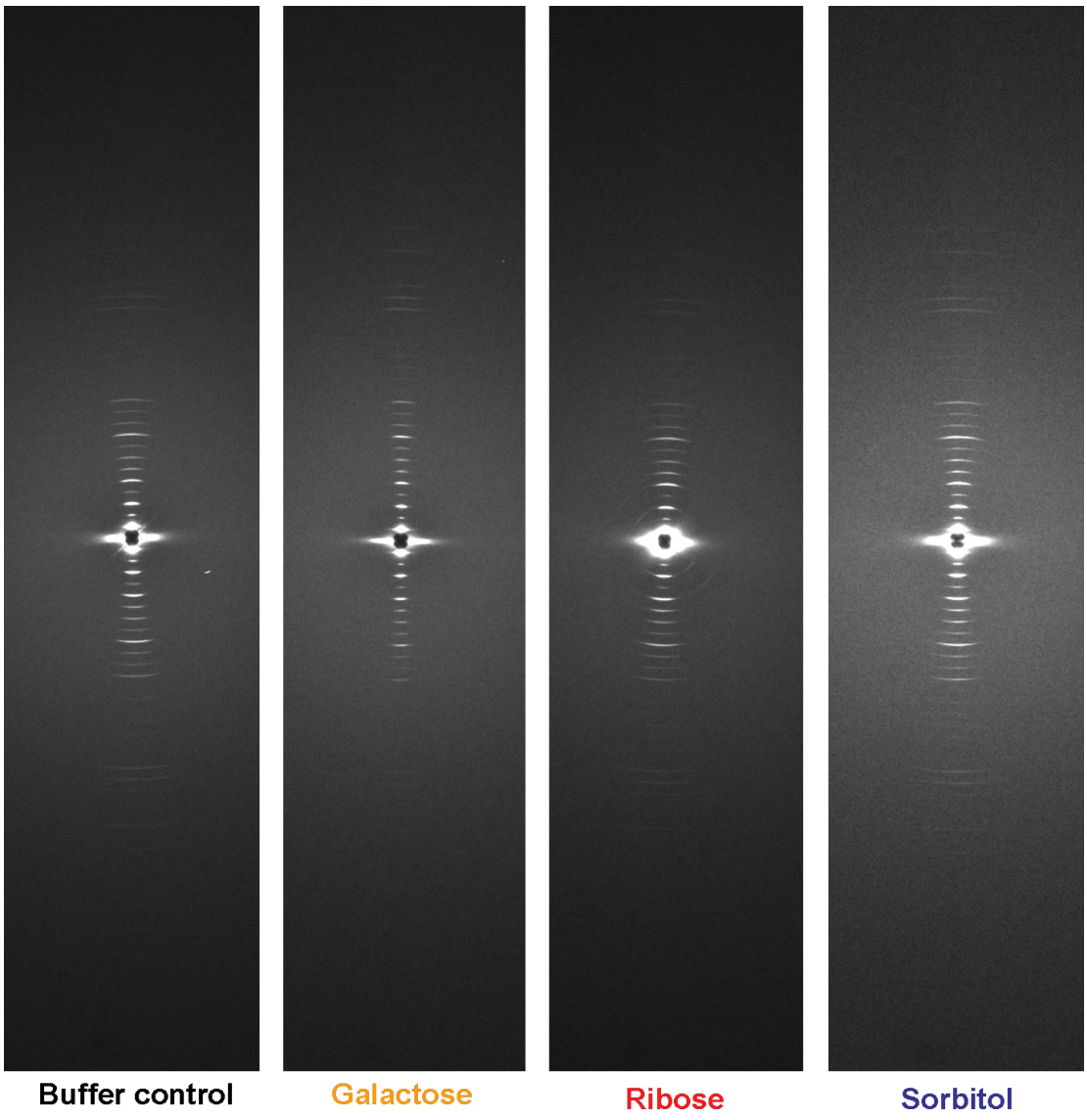


Figure 6.14 Meridional X-ray diffraction patterns obtained, left to right, from MTT incubated for 2 weeks in MOPS-saline buffer, and buffer containing 0.2M galactose, ribose and sorbitol, as labelled. As previously, the images have been rotated clockwise by 90^{o} to show as per convention. The top halves of the meridional patterns represent data collected from the left half of the detector, the bottom representing that from the right.

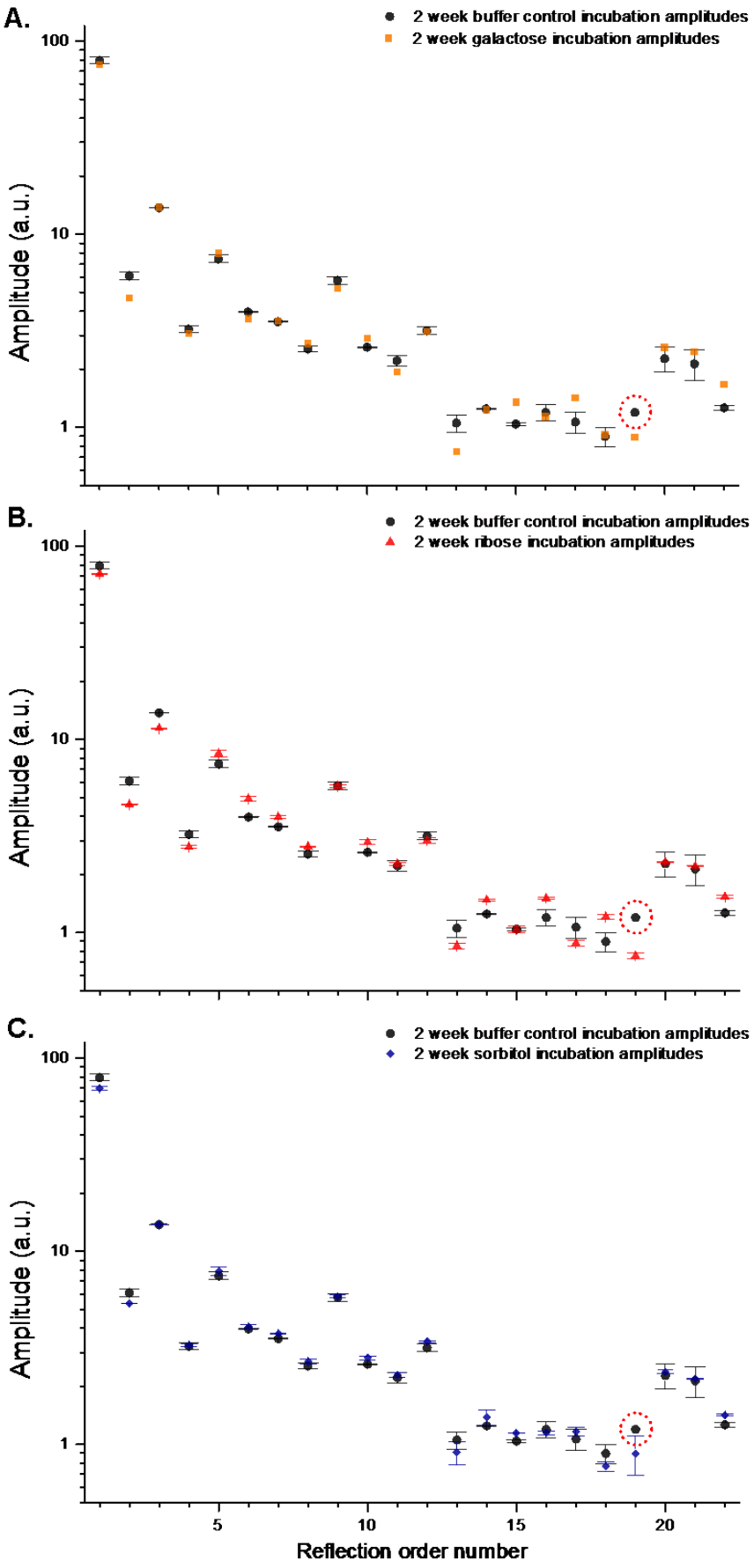

A.
2 week buffer control incubation amplitudes
2 week galactose incubation amplitudes
Amplitude (a.u.)
100
10
1
B.
2 week buffer control incubation amplitudes
2 week ribose incubation amplitudes
Amplitude (a.u.)
100
10
1
C.
2 week buffer control incubation amplitudes
2 week sorbitol incubation amplitudes
Amplitude (a.u.)
100
10
1
5
10
15
20
Reflection order number

Figure 6.15 (Previous page) Semi-$\log_{10}$ plot showing comparisons of the mean X-ray diffraction pattern meridional reflection amplitudes from 2 week buffer control MTT with those measured from the data collected from MTT after **A.** 2 week galactose incubation, **B.** 2 week ribose incubation and **C.** 2 week sorbitol incubation. All reflection intensities were measured using averaged traces from each half of the detector data. The error bars about each point show the sample SDs of the duplicates from both detector halves, where both halves produced usable data. The numbers of individual patterns upon which the averaged data traces were based were: **A.** for galactose, n = 9 from the left side of the detector with no data from the right being extracted due to it being significantly less intense. Therefore no error bars are shown for galactose; **B.** for ribose, n = 10 for both halves of the detector; **C.** for sorbitol, n = 11 for both halves of the detector. For the buffer control data, n = 8 for the left (upper as shown) and n = 7 for the right (bottom as shown) halves of the detector. The red dotted circles around the 19th order reflection amplitude for the buffer control are placed to draw attention to the fact that this amplitude could only be measured with confidence from the data extracted from the right (lower as shown) half of the detector, therefore there is no error bar. All of the glycated data were $\log_{10}$ scaled against the buffer control data so it was comparable and all data were then Lorentz-corrected prior to plotting.

The ribose-incubated MTT showed the most differences at 14 plus 2 marginal changes. In the case of galactose it was 6 differences plus 2 further small ones. Sorbitol incubation produced the fewest reflection amplitude changes relative to the control, at 4 plus 2 more marginal differences. All the reflection order amplitudes that changed relative to the buffer control in the case of more than one sugar treatment, did so in the same direction – i.e. increase or decrease - with the exception of order 17, which was raised in the case of the tendon treated with galactose but had fallen following ribose treatment.

Consistent with previous work (Tanaka et al., 1988a; James et al., 1991), in the cases of both ribose and galactose, amplitudes of a cluster of meridional reflections that correspond to the non-helical telopeptide regions were noticeably changed (Hulmes et al., 1977; Brodsky et al., 1982). These were orders 13 to 19. However in the case of ribose especially, changes in the amplitudes of several other order reflections were evident, as described above.

6.3.3 Effects of glycation on axial electron density plots for MTT

6.3.3.1 Patterson plots

The amplitude data for reflection orders 1 to 22 for four experimental MTT glycation incubation groups are shown in Table 6.3. From these data, Patterson plots were prepared for each of the MTT samples incubated for 2 weeks in buffered saline (buffer

control), and 0.2M ribose, galactose and sorbitol solutions in buffer. These are shown in Figure 6.16. Due to the comparatively large amplitude of the 1st order reflection, inclusion of this order in the Fourier synthesis of the Patterson plots largely obscures other detail revealed by the other smaller orders and results in almost perfectly overlaid traces (Figure 6.16A). Removal of this order allows for clearer differentiation between the treatment groups. Comparison of the sugar incubation Patterson plots in Figure 6.16B with the buffer control shows that the sorbitol incubation plot correlates with it most closely (Pearson's r = 0.999, R^2 = 0.998) followed by the galactose plot (Pearson's r = 0.996, R^2 = 0.992). Visual inspection shows the ribose incubation Patterson to deviate most significantly from the buffer control trace (Pearson's r = 0.977, R^2 = 0.955).

Order	Buffer control incubation amplitudes	Ribose incubation amplitudes	Galactose incubation amplitudes	Sorbitol incubation amplitudes
1	79.4	99.4	78.1	78.0
2	6.09	5.57	4.72	5.97
3	13.7	14.4	14.1	15.3
4	3.22	3.27	3.06	3.61
5	7.45	10.5	8.10	8.77
6	3.96	5.96	3.67	4.50
7	3.53	4.76	3.58	4.15
8	2.55	3.29	2.72	2.98
9	5.76	7.01	5.31	6.47
10	2.59	3.47	2.91	3.10
11	2.21	2.64	1.93	2.52
12	3.16	3.55	3.14	3.77
13	1.05	0.95	0.74	1.00
14	1.25	1.69	1.22	1.52
15	1.04	1.16	1.34	1.26
16	1.19	1.72	1.13	1.27
17	1.06	0.98	1.41	1.28
18	0.89	1.36	0.91	0.85
19	1.19	0.84	0.88	1.77
20	2.26	2.71	2.60	2.62
21	2.13	2.57	2.47	2.41
22	1.26	1.75	1.67	1.57

Table 6.3 Lorentz-corrected amplitudes for each meridional reflection used in the Fourier syntheses of the Patterson and electron density plots for the 2 week buffer control, ribose, galactose and sorbitol incubation MTT samples.

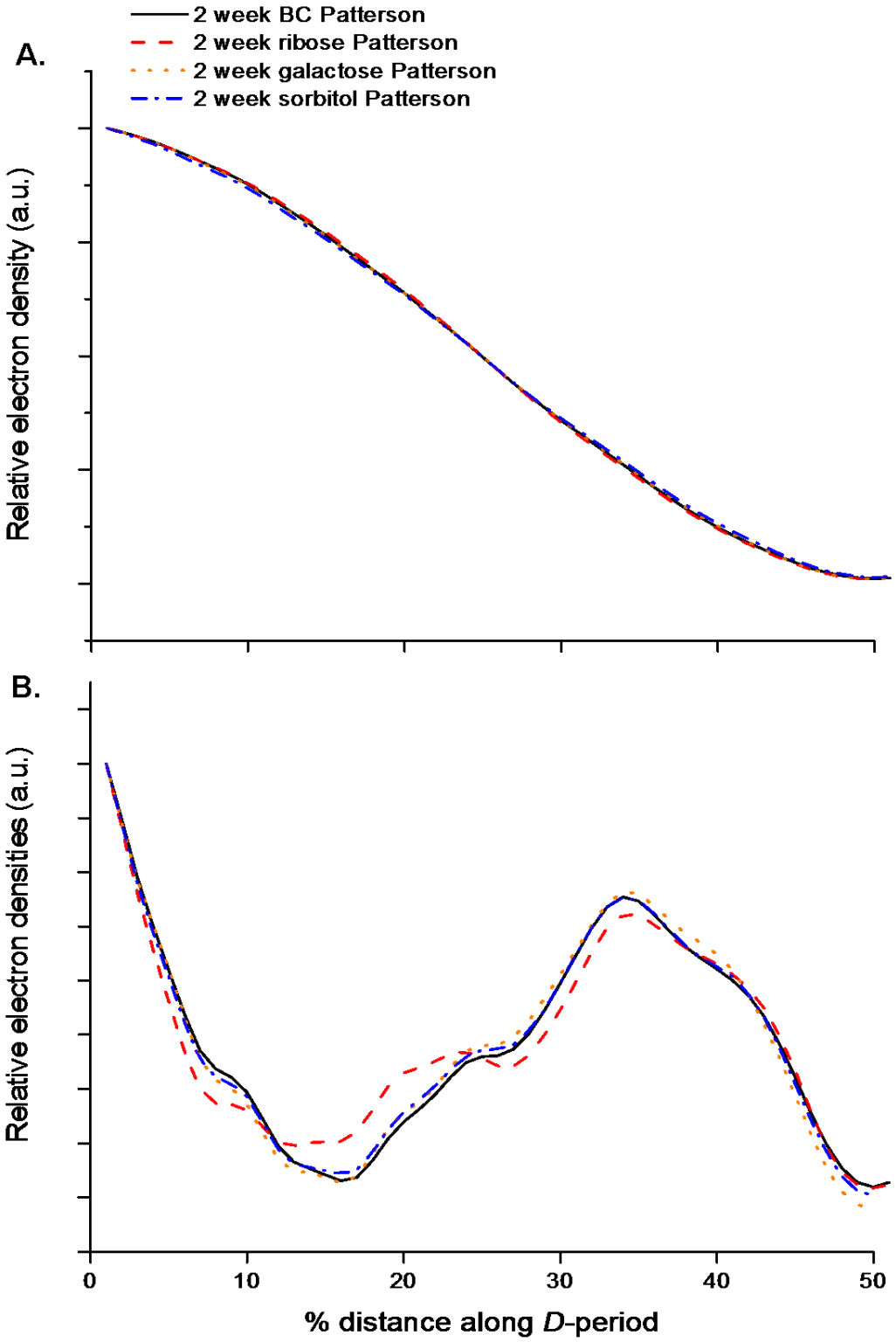


Figure 6.16 Patterson plots for the four 2 week incubation MTT groups, i.e. buffer control and the ribose, galactose and sorbitol incubations. **A.** All Patterson plots are synthesised using the amplitudes for orders 1 to 22; **B.** Patterson plots synthesised from reflection amplitudes for orders 2 to 22, i.e. the 1st order reflection amplitude is excluded. Due to the Patterson plots' centrosymmetry, only the first 50% is shown for each.

6.3.3.2 Electron density plots for glycated MTT samples

Using the amplitudes from Table 6.3 and the 2 week buffer control incubation phases shown in Table 6.2, a set of structure factors was produced for the order 1 to 22 reflections for each incubation MTT treatment group. From these, axial electron density plots over one *D*-period were produced for each. In turn the ribose, galactose and sorbitol were linearly-scaled against the buffer control plot for comparison. These are shown in Figure 6.17, Figure 6.18 and Figure 6.19 respectively. As with the Patterson plots, due to the comparatively high magnitude of the 1st order compared with the others electron density plots were also prepared after exclusion of the 1st order structure factor and these are also shown. All overlaid electron density plots are shown in closest scaled alignment rather than electron density relative to one another *per se*.

Overall and unsurprisingly, of the three glycation treatment groups the ribose-treated MTT appears to show the biggest differences in its electron density plot from the buffer control MTT (Figure 6.17A: Pearson's r = 0.999, R^2 = 0.998; Figure 6.17B: Pearson's r = 0.989, R^2 = 0.978). Galactose shows minor differences as compared with the control in the entire trace, which become more obvious when comparing the plots constructed having excluded the 1st order structure factor (Figure 6.18A: Pearson's r > 0.999, R^2 = 0.999; Figure 6.18B: Pearson's r = 0.995, R^2 = 0.991). Sorbitol – originally intended as the "control sugar" – appeared to show slight visually apparent entire electron density plot differences as compared with the buffer control. On removal of the 1st order from the synthesis process however, these differences did not become any more visually apparent (Figure 6.19A: Pearson's r = 0.999, R^2 = 0.998; Figure 6.19B: Pearson's r = 0.999, R^2 = 0.997), in contrast with the galactose and ribose plots.

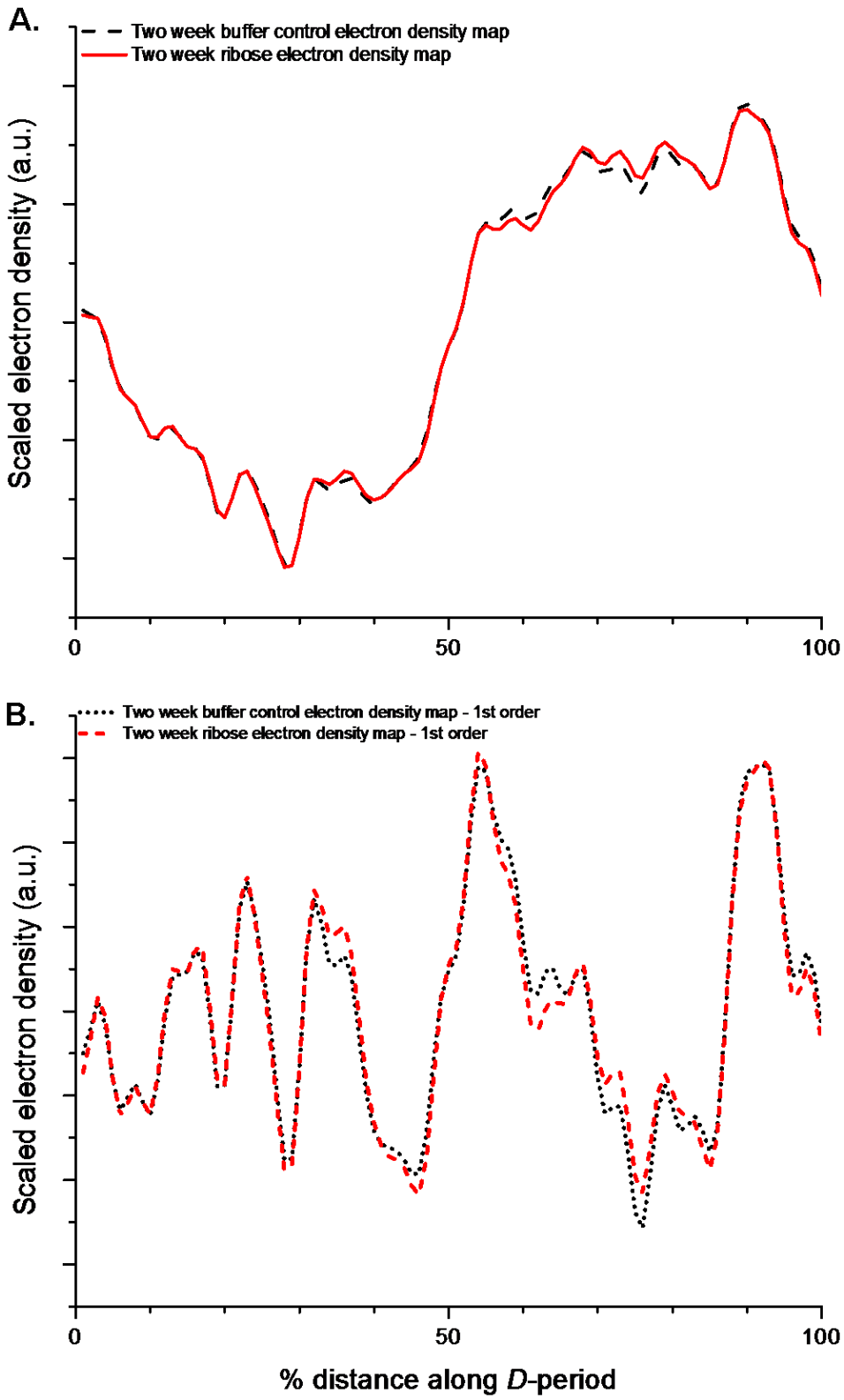


Figure 6.17 Electron density plots of ribose-incubated MTT linearly scaled against and overlaid with the buffer control plots **A.** constructed using structure factors for orders 1 to 22, and **B.** constructed using structure factors for orders 2 to 22, i.e. excluding the 1st order.

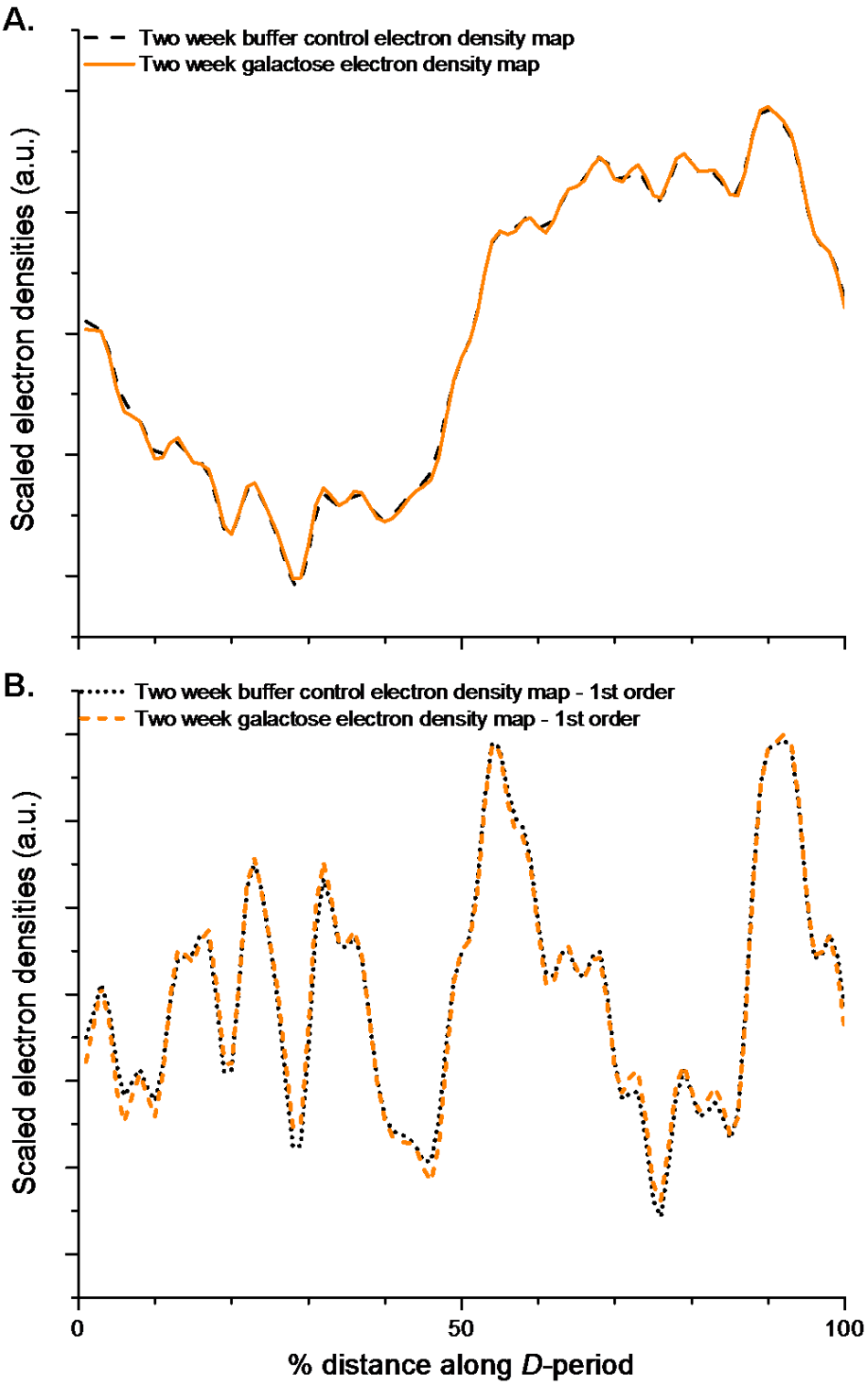


Figure 6.18 Electron density plots of galactose-incubated MTT linearly scaled against and overlaid with the buffer control plots **A.** constructed using structure factors for orders 1 to 22, and **B.** constructed using structure factors for orders 2 to 22, i.e. excluding the 1st order.

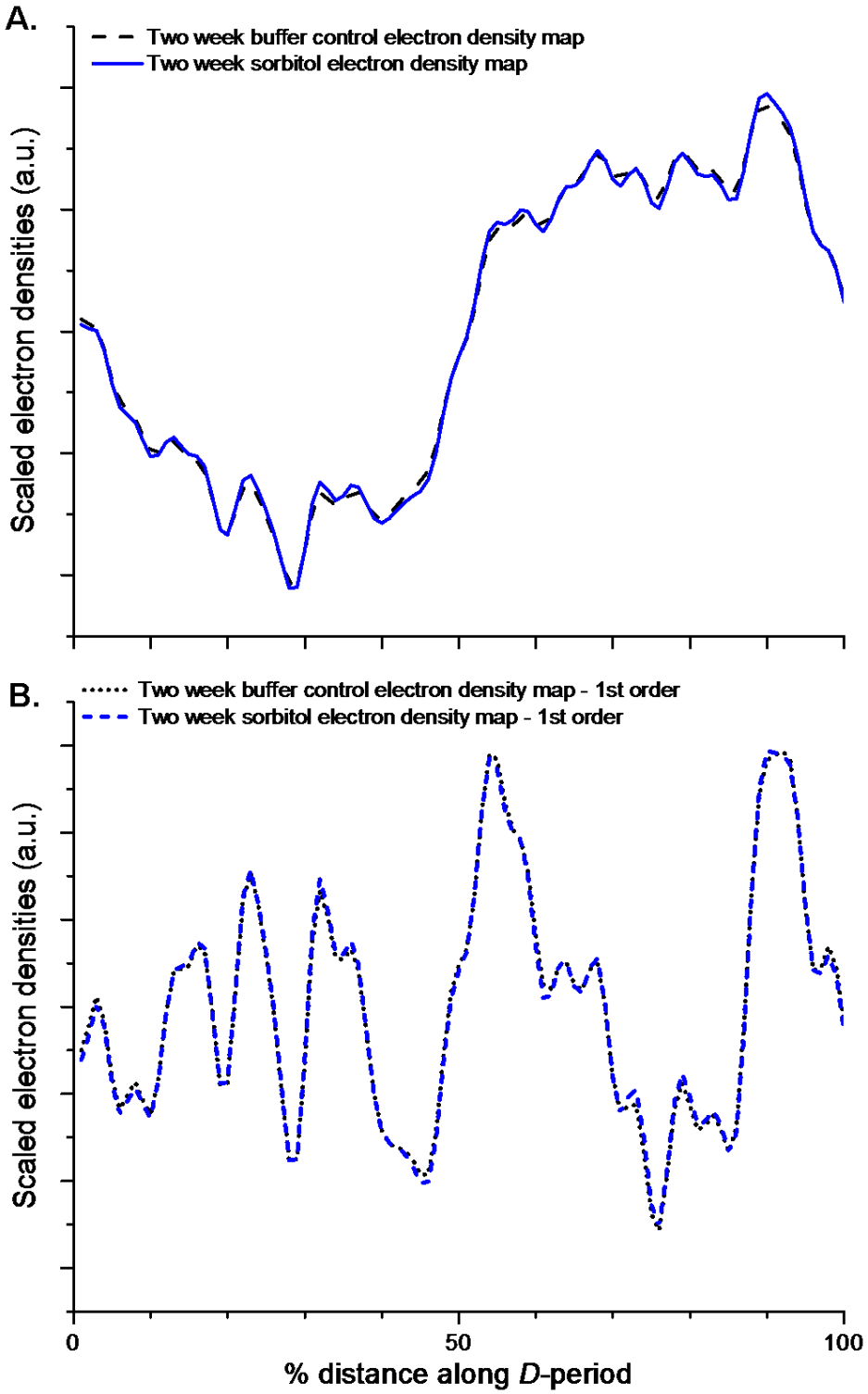
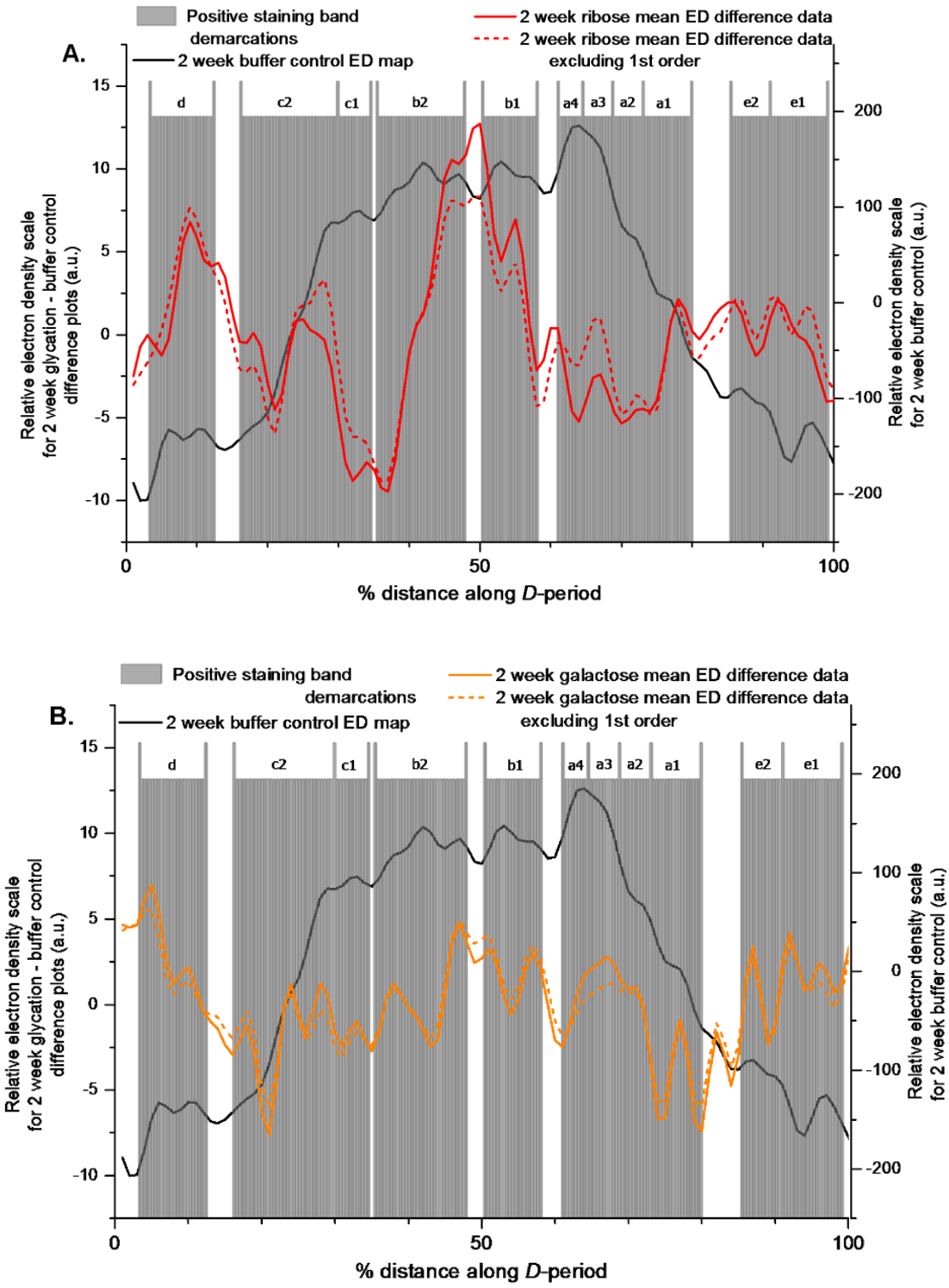


Figure 6.19 Electron density plots of sorbitol-incubated MTT linearly scaled against and overlaid with the buffer control plots **A.** constructed using structure factors for orders 1 to 22, and **B.** constructed using structure factors for orders 2 to 22, i.e. excluding the 1st order.

6.3.3.3 Locating electron density changes resulting from glycation using difference plots

In Figure 6.20 are shown the difference plots of electron density for the MTT treated for 2 weeks with ribose, galactose and sorbitol, calculated by subtracting the mean buffer control electron density plot from that of each sugar in turn. This was carried out for the electron density plots constructed from the structure factors for orders 1 to 22 plus also from the structure factors for orders 2 to 22, i.e. excluding the 1st order. The alternatives for each sugar are plotted on the same scale for comparison. In the case of ribose and galactose, it can be seen that in each case both difference plots (i.e. with and without 1st order structure factor inclusion) produced show striking similarities both in terms of peak positions and magnitude of electron density difference. The 1st order in these cases has therefore not significantly distorted the electron density plots for these sugars relative to the buffer control.

The most significant changes in electron density are seen in the difference plot for 2 week ribose treatment MTT (Figure 6.20A). The biggest increase in electron density with this treatment is associated with the *b1* and *b2* (lower half) positive staining bands with another large peak associated with the *d* band. More moderate increases are evident in the *c2*, *a1* and *e1-e2* bands. The total ranges of electron density difference values from minimum to maximum according to the plots including and excluding the 1st order (solid and dashed red lines, respectively), are 22.2 a.u. and 17.2 a.u. respectively. The minimum to maximum range of the 2 week buffer control MTT electron density trace is 392 a.u., so the maximum detected percentage changes in electron density as a proportion of this at any point along the *D*-period with the ribose treatment are 5.7% with and 4.4% without 1st order inclusion.


A.
Positive staining band demarcations
2 week buffer control ED map
2 week ribose mean ED difference data
2 week ribose mean ED difference data excluding 1st order
d
c2
c1
b2
b1
a4
a3
a2
a1
e2
e1
Relative electron density scale for 2 week glycation - buffer control difference plots (a.u.)
Relative electron density scale for 2 week buffer control (a.u.)
% distance along D-period
B.
Positive staining band demarcations
2 week buffer control ED map
2 week galactose mean ED difference data
2 week galactose mean ED difference data excluding 1st order
Relative electron density scale for 2 week glycation - buffer control difference plots (a.u.)
Relative electron density scale for 2 week buffer control (a.u.)
% distance along D-period

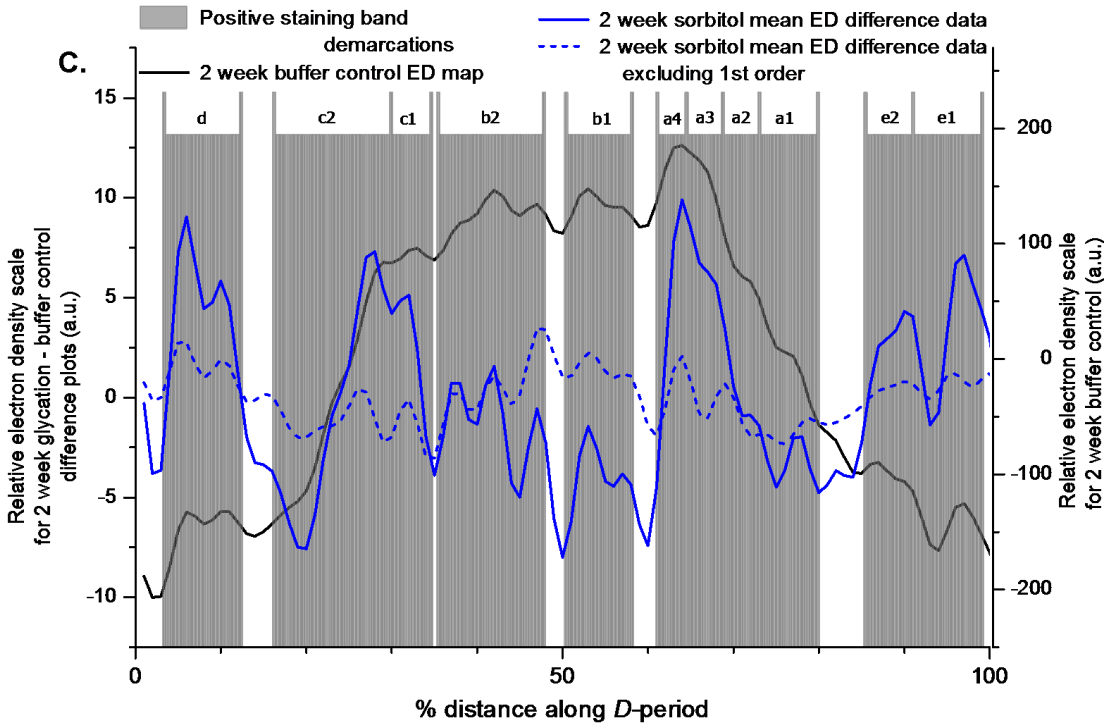


Figure 6.20 (Previous page and above) Difference electron density plots (coloured lines) for **A.** ribose, **B.** galactose and **C.** sorbitol, aligned and overlaid with 2 week buffer control MTT electron density plot (dark gray line) and positive staining TEM banding locations. The solid coloured lines represent the difference plots produced by subtracting the complete buffer control treatment electron density trace from that obtained for the sugar post-linear scaling of the two. The dashed coloured lines represent the same, but where the 1st order reflection structure factor had been excluded from the Fourier synthesis of both electron density plots. All sugar difference plots use ordinate electron density difference data plotted against the y-axis on the left and this scale is the same for **A**, **B**, and **C**, for ease of comparison. The *y*-axis scale on the right of each plot is applicable to the 2 week buffer control MTT plot.

In the case of the 2 week galactose treatment (Figure 6.20B), the changes are smaller. Most significantly, there is an increase in electron density in the *d* band. Smaller increases are also apparent in the *b1* band and lower half of the *b2* band, the *e1-e2* bands and broadly, across the *a2-a3-a4* bands. The range of electron density difference values with galactose treatment is 14.6 a.u. with 1st order structure factor inclusion in the calculation (solid orange line) and 11.6 a.u. without (dashed orange line), giving maximum proportional measured changes in electron density of 3.7% and 3.0% respectively.

From the sorbitol MTT results, it can be seen that there is a significant discrepancy between the electron density difference plots derived from the 2 week sorbitol and buffer control plots constructed with and without the inclusion of the first order (Figure 6.20C, solid and dashed blue lines, respectively). Although all of the peaks appear to

be in almost exactly the same places, their relative heights are altered. The trace derived from the entire electron density plots suggests significant measurable electron density increases as a result of the sorbitol incubation in the *a3-a4*, *c1-c2*, *d* and *e1-e2* positive staining band regions. Conversely the difference trace derived from the sorbitol and buffer control electron density plots constructed in the absence of the 1st order would tend to suggest that any changes in electron density which have occurred after the sorbitol incubation have been small and/or non-specific. It is likely that this difference plot discrepancy is as a result of an inaccuracy in the measurement of the 1st order peak height. Expressed as a proportion of the sum of the other amplitudes, i.e. orders 2 to 22, the mean 2 week sorbitol MTT 1st order amplitude gave a value of 1.017. Comparison of this with the values 1.175, 1.239 and 1.156 for the buffer control, ribose and galactose treatment groups respectively would seem to show a significantly greater discrepancy between the sorbitol value relative to the buffer control, than either of these sugars (-0.158 versus +0.064 and -0.019). An artefactual augmentation of the relative magnitudes of other electron density characteristics as compared with the control seems to have resulted, thereby producing two notably differing difference traces as a result.

The data for changes in electron density resulting from the sorbitol treatment are equivocal though on balance, it is more likely that the difference plot produced from the electron density plots produced with the exclusion of the 1st order is most representative of the true electron density changes. Therefore it would be difficult to conclude with confidence that significant changes have occurred after the 2 week sorbitol incubation.

6.4 Discussion and conclusions

The first part of the work described in this chapter was the production of an axial electron density map for one MTT *D*-period following incubation in the MOPS-saline glycation buffer for 2 weeks. It is evident from Figure 6.10 that direct comparison of the axial electron density plot for the 2 week buffer control MTT with that for freshly dissected native tissue showed significant discrepancies. Whilst the peaks in the buffer control MTT plot are preserved when compared with the untreated MTT or can be accounted for in terms of peak shifts - the most notable of which appears to be the migration of the *c1-c2* positive staining bands towards the *b1* band - the differences between the two plots is still clear.

One possibility as to the cause of this is that inaccuracies in the structure factors have combined to produce a cumulative effect. Of the two structure factor components, the amplitudes are the more straightforward to measure. Study of Figure 6.4 indicates the duplicate precision between the sides of the detector to have been satisfactory. Substitution of these two sets of amplitudes individually into the Fourier synthesis of both electron density maps produces plots which retain the same shape and characteristics as that presented (Figure 6.9). The phases are considerably more involved and difficult to measure. Comparison of the Patterson plots for the untreated and buffer control MTTs – constructed from just the amplitudes – as shown in Figure 6.21 shows them to be noticeably different. Though the differences are apparently small and non-specific, they appear to be significant and would support the expectation of a difference between the electron density plots observed.

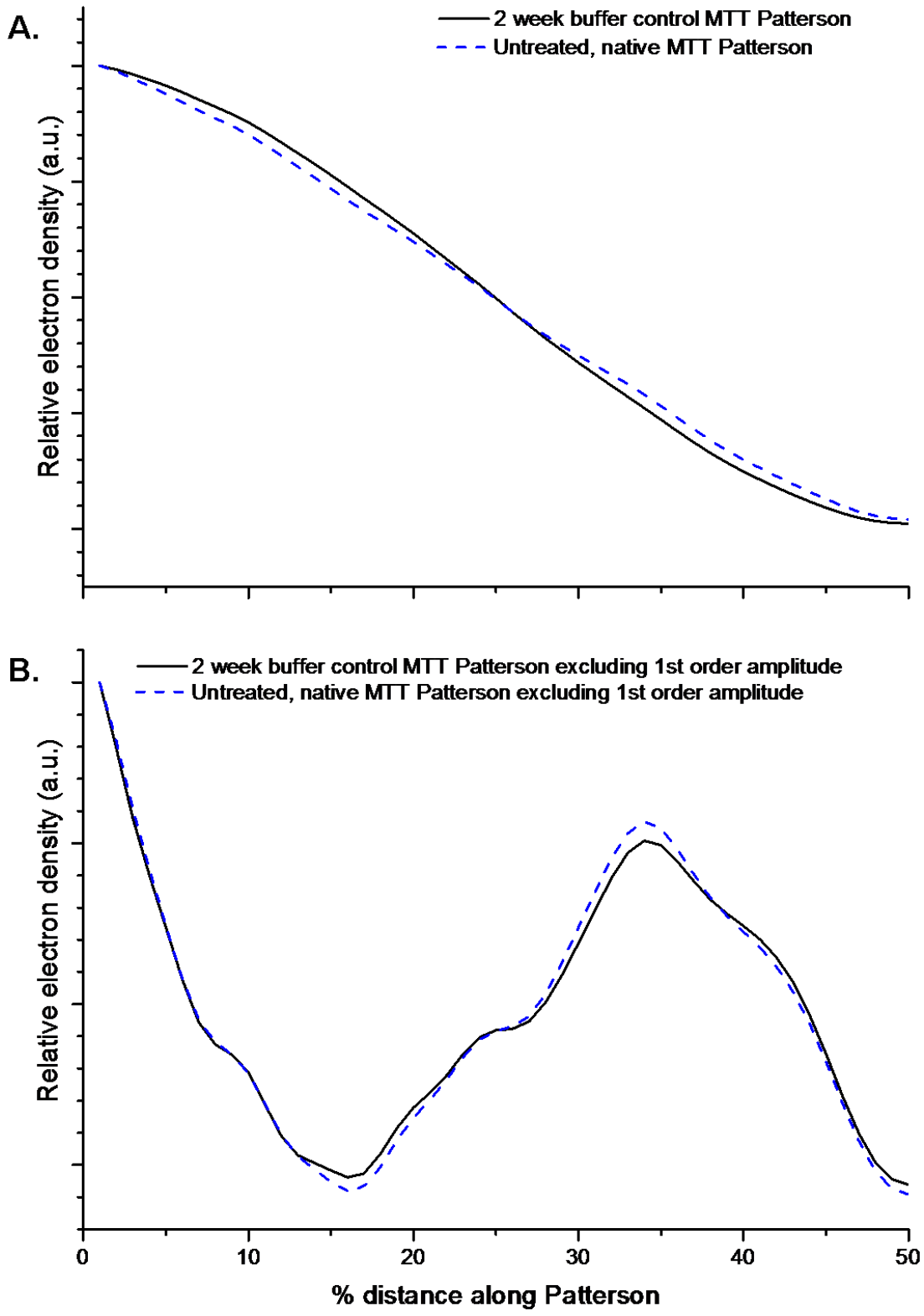


Figure 6.21 Overlay comparisons of **A.** complete Patterson plots and **B.** Patterson plots constructed with the omission of the 1st order amplitude, for one *D*-period of 2 week buffer control MTT and native, untreated MTT. Only 0 – 50% of the whole plot is presented due to the Patterson plots' centrosymmetry.

Changes in the Patterson plots of proteins are difficult to pinpoint due to the huge number of possible vector spacings between centres of electron density in the macromolecule and therefore the smooth and comparatively featureless curves obtained. As the number of centres of electron density represented are smaller and specific, comparisons of the heavy atom difference Patterson plots for the untreated and 2 week buffer control MTT groups, i.e. the means of the experimental plots originally presented in Figure 5.17 and Figure 5.18, and Figure 6.7 and Figure 6.8, can yield more information about the nature of the shifts that are occurring. These are presented in Figure 6.22.

For the iodine difference Patterson (Figure 6.22A), although there are differences between the two plots – most notably, the peak at 34% for the untreated MTT that is not evident for the 2 week buffer control MTT and the apparent peak height scaling schism for the two plots, for the range 0 – 30% as compared with 30 – 50% - they appear broadly the same. By comparison, although the shapes are generally preserved the gold Patterson plots show profound differences between untreated and 2 week buffer control MTT. Acknowledging the necessary shift of the predicted and experimental gold difference plots relative to one another to achieve the best comparative overlay, inspection of Figure 5.18 showed a strong comparison in the case of the untreated MTT whereas for the buffer control MTT (Figure 6.8), there is a greater discrepancy between predicted and experimental Patterson plots. This is seen more clearly in Figure 6.22B where the experimental data are presented on the same percentage distance scale (abscissa). The changes seem mainly to be associated with a shift of several peaks to the right, thus a generalised increase in the most significant vector spacings. The most profound identifiable change is that the largest experimental peak at 35% has shifted to the 40% position. Several other changes are apparent, e.g. general changes in the peak heights and positions between 9% and 30%, including an apparent loss or shift of the untreated MTT peak at 15% following buffer incubation. However except for the large peak shift, it is difficult to be specific as to which peak (or portion of a peak) is moving and in which direction. For example, the increased electron density buffer control MTT peak at 27% could be taken to be equivalent to the untreated MTT peak at 25%, shifted right and with increased electron density. However, it could also contain a component of electron density shifted right from the apparently-flattened peak at 20% or even a contribution, left-shifted from the large peak at 35% or its shoulder at 29%. Therefore interpretation of much of the gold difference Patterson would be overly-reliant on guess work and would amount to little more than speculation.

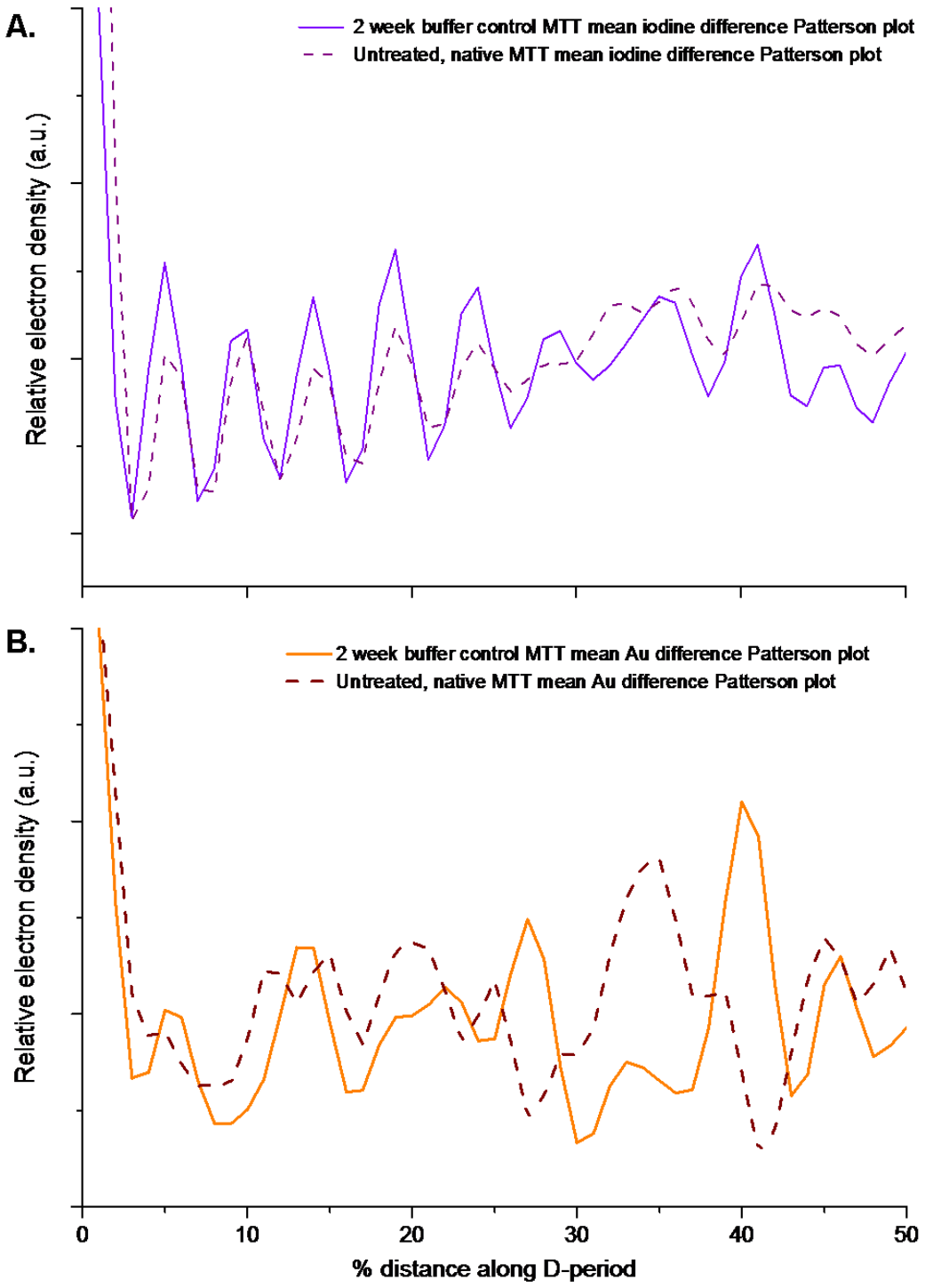


Figure 6.22 Comparisons of the experimental mean heavy atom difference Patterson plots for one *D*-period of 2 week buffer control and untreated, native MTT for **A.** iodine and **B.** gold binding. As previously, only the lower 50% of the Patterson is shown. Each trace is the mean of the six individual traces shown in Figure **5.17** and Figure **6.7** (**A**), and Figure **5.18** and Figure **6.8** (**B**).

The comparative lack of changes in the iodine difference Patterson between untreated and 2 week buffer control MTT would seem to suggest that as the iodine binds to tyrosine and histidine residues and the histidine residues provide common binding loci for both iodine and gold atoms, the changes in structure brought about by incubating collagen in the glycation buffer show a relatively minor effect on tyrosine position and associated electron density. As 12 of the 13 tyrosine residues found in the *D*-period are located in the telopeptide regions, this might suggest relatively small involvement of these in the observed effects of the buffer incubation. Conversely, as gold binds to methionine as well as histidine, if one assumes that the minor scale of change associated with histidine electron density and position is as reflected by the iodine difference data, the profound changes seen between the experimental untreated and 2 week buffer control gold difference Patterson plots would relate to methionine residues, which are spread widely along the *D*-period outside the telopeptide regions (Figure 6.13A). The large peak shift from 35% to 40% is principally associated with a 5% increase in the spacing between the resolvable clusters of “H” and “M” residues located at positions along the *D*-period either within or alongside the N- or C-telopeptide regions. Of all these amino acids, only the two methionines in position 2 of the α1-chains are in the non-helical regions at the N-terminal, with four further methionines close to the N-termini of, but within the triple helical region (i.e. M19 (α1-chains); M14 and M17 (α2-chain)) (See Figure 5.15A). The remaining histidines and methionines identified as being close to the telopeptide regions within a *D*-period are at various loci along the triple helices without being especially close to either terminus within their own molecules. It is reasonable to presume that these residues hold importance, nonetheless, in providing an environment conducive to correct functional telopeptide conformation being maintained. This is especially important for the C-telopeptides, as none of the histidine or methionine residues found adjacent in the fibril to this molecular region are close to either of their molecules’ termini, and as considered in Chapter 5, section 5.4, there appears to be ambiguity in current knowledge as to the correct configuration of these non-helical termini.

The increased Patterson spacing observed here would tend to suggest that some degree of molecular slippage and stretching might have occurred within the microfibrillar structure, perhaps partially as a result of the tension employed during the X-ray data collection component of the experiments, as has been described previously (Mosler et al., 1985) and that the buffer treatment increased the susceptibility of the tissue to this. This would be consistent with the entire electron density map plots as

presented in Figure 6.10 in section 6.3.1.4, wherein it was noted that the major positive staining band-associated electron density peaks could not be optimally aligned with those from the untreated MTT whilst still maintaining close agreement with the positions of the gap-overlap regions.

Lysyl oxidase activity in MTT (and RTT) tissue as compared with tendon from other sources, is low and so enzymic cross-link formation and maturation continues well beyond the mouse reaching adulthood at 2 – 3 months (Avery and Bailey, 2005; Svensson et al., 2013). This might suggest a greater potential for this slippage and stretching of molecules to occur without cross-link-mediated constraint in the 3 month old model used here.

With regard to the C-telopeptide α-chains, according to the comparison between the experimental and model electron density plots for the 2 week buffer-incubated MTT and the iodine atom binding difference electron density plots (section 6.3.1.4, Figure 6.13) the conformation is closer to that previously suggested (Orgel et al., 2000), i.e. both folded into a hairpin formation, than the one folded α1-chain – one straight α1-chain conformation suggested in section 5.4 for the untreated MTT collagen. It could be therefore that the slippage and altered localised molecular stretching characteristics brought about by the buffer incubation have made a more contracted C-telopeptide folding structure more thermodynamically favourable than was observed in the untreated MTT tissue. This may be responsible for the fact that the iodine atom difference Patterson for the buffer-treated MTT collagen shows improved peak resolution above 27% as compared with the untreated (Figure 6.22A). The uniformity of the more compact α1 C-telopeptide conformations augments and sharpens the peaks, as opposed the heterogeneity suggested for the untreated MTT.

The nature of the region-specific changes occurring in the buffer control MTT tissue were considered in Chapter 4 as significant changes seem to have occurred in the relative intensities of the TEM positive bands over certain time periods of incubation, i.e. between 1 and 2 weeks, and 3 and 4 weeks (section 4.3.3, Figure 4.15) and over the full 4 weeks. They are assumed to be electrostatic in nature, analogous to the likely binding of negatively charged phosphate buffer anions to positively charged amino acid side-chains (Watkins et al., 1987). It is relative electron density which is being measured along the *D*-period here as opposed to relative numbers of unbound or unaltered positively charged amino acid-rich sites. However as the changes studied using TEM are expected to be precursor stages to any detectable electron density changes, it is reasonable to hypothesise that the buffer control-related changes might

show some resemblance when comparing both sets of data from the same incubation time periods.

Between the start of incubation and 1 week, there were no significant TEM positive staining changes. However, between 1 and 2 weeks, marked relative reductions in band intensities were noted concentrated mainly in the *a2-a3-a4* and *c1-c2* bands. In Figure 6.23 are shown mean positive staining traces for untreated and 2 week buffer control MTT *D*-periods, overlaid for comparison with the *a* and *c* band regions highlighted for clarity. The positions of the peaks appear to be unaltered following the buffer incubation. Even so, salient changes can be ascertained from the comparison. They are **1** a significant drop in the relative band intensities of the *c1* and *c2* bands and **2** an apparent broadening of the *a* bands' region, no evidence of which can be seen after just 1 week's incubation. This latter may be as a result of the previously noted preferential loss of *a2-a3-a4* relative band intensity or perhaps some peak shifting or moving of positively charged residues has actually occurred as a result of the incubation.

It should be borne in mind that as considered in Chapter 4, section 4.4.1, the difference between the two traces highlighted in Figure 6.23 may be at least in part as a result of the data from the untreated and the 2 week buffer control MTT groups being mean traces for tissue taken from two separate, individual mice. However, some consistency appears to exist between TEM positive staining data and the electron density changes observed after buffer incubation (see Figure 6.11). The entities being studied differ in their nature as do the protocols involved, e.g. dehydrated specimen versus hydrated, tensioned specimen versus non-tensioned, therefore direct comparison of phenomena observed – such as movement of peaks – should be employed with caution. However the principal electron density changes associated with the buffer control incubation, apparently relating mainly to the molecular termini and their adjacent molecules triple helical regions at this time period, seem to have removal or redistribution of positively charged amino acids at the same or proximal *D*-period loci as a precursory or concomitant occurrence.

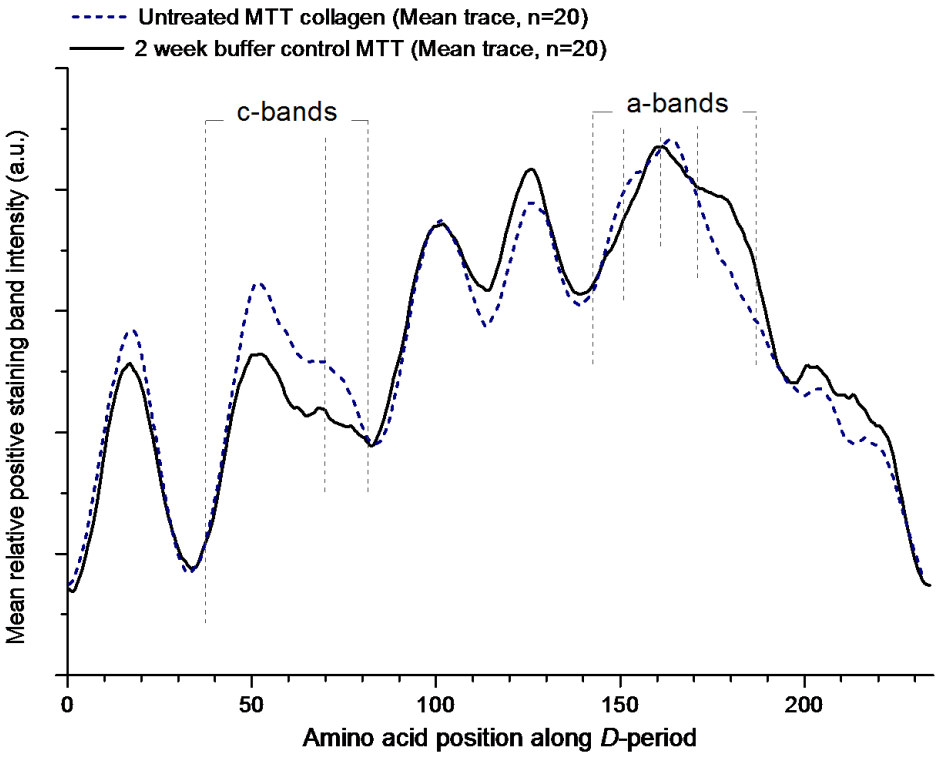


Figure 6.23 Mean positive staining TEM band intensity traces for the untreated MTT and 2 week buffer control MTT, using data from the work presented in Chapter 4.The overlaid traces shown are individually scaled by the OriginPro graphing software along the ordinate for optimal presentation and not against one another. The positions of the *a* bands (*a4, a3, a2, a1*, left to right) and *c* bands (*c2, c1,* left to right) are indicated by the dashed lines.

The most likely explanation as to the causes of the change in the fibrils' susceptibility to molecular stretching and slippage is some removal or negation of charge by buffer system, resulting in changed properties of the more flexible, positively-staining spring-like regions of the molecule (Silver et al., 2001). This might prevent these regions from resistance or recovery from a tensile force as well as fundamentally changing interactions between individual staggered collagen molecules within the fibril.

6.4.1 Detection of glycation-related changes in electron density

The differences between the amplitudes for meridional reflection orders 1 to 22 in the ribose- and galactose-treated MTT as compared with the buffer control are presented in section 6.3.2 and Figure 6.15. The previous studies using X-ray diffraction to study changes in type I collagen-containing tissue associated with glycation both identified meridional orders 14 to 19 as exhibiting the most notable post-glycation changes (Tanaka et al., 1988a; James et al., 1991), suggesting that the telopeptide regions

demonstrated the most significant structural changes (Hulmes et al., 1977; Brodsky et al., 1982). Whilst in this study, both ribose and galactose treatment showed changes in these orders, more widespread differences were evident, so any electron density effects detected were expected to be located in both the telopeptide and helical regions.

On the basis of the differences seen, comparison of post-glycation MTT Patterson plots with the buffer control (see section 6.3.3.1 and Figure 6.16) suggests that the biggest alterations in electron density are produced by glycation with ribose, with any changes produced by galactose being smaller as would be predicted by the proportion of each in the acyclic form (Hayward and Angyal, 1977; Bunn and Higgins, 1981).

Combination of the meridional reflection amplitudes from the sugar treatment groups with the phases previously calculated for the corresponding (2 week) buffer control produced the structure factors that were used for the construction of the sugar-treated MTT electron density maps. The rationale of this approach here, previously described as “cross-phasing” (Orgel, 2000), is based upon assumptions regarding the effect of glycation specifically on the collagen protein. The axial changes are presumed to be entirely as a result of AGE and cross-link formation, the latter of which would form between specific and the most predisposed amino acid residues, so stabilising the structure and minimising movement of loci of differing electron density relative to one another along the microfibril. In addition, the level of imprecision inherent in the phase determination procedure in comparison with the relatively small anticipated differences between the control and glycated MTT samples would be likely to outweigh the benefits of further phase determination in each glycation incubation group. As the “phases” tend to locate or place the centres of electron density, the magnitudes of which are indicated by the corresponding structure factor “amplitudes”, any such errors might lead to an artefactual movement of peaks and therefore to a phenomenon of *peak chasing*, i.e. applying a certain level of guess work to relate shifted (or perhaps, lost or new) peaks to those present within the control trace, in the lack of any other available evidence.

Another consideration was that histidine residues, important in the binding of both the iodine and gold heavy atoms used in the phase determinations in this study, might be involved in AGE cross-link formation as has been determined in serum albumin (Slatter, Avery and Bailey, 2004). This might alter the heavy atom distribution along the *D*-period and affect the phase determination process. However, this is thought not to

occur due to steric and positional constraints within collagen's structure (Slatter et al., 2004) and should not be a factor.

The electron density plots presented in section 6.3.3.2 and shown in Figure 6.17 and Figure 6.18 show that the changes following the 2 week ribose and galactose incubations respectively are visibly small as compared with the buffer control plot. Alongside the electron density difference traces (see section 6.3.3.3 and Figure 6.20) they again confirm the reactivity of the two sugars to be as predicted (Hayward and Angyal, 1977; Bunn and Higgins, 1981), i.e. ribose > galactose, as suggested by the relative electron density changes expressed as a proportion of the entire range of the electron density plot (5.7% and 3.7% respectively, based on the entire electron density plots produced).

Comparison of the glycation electron density difference plots with the corresponding relative positive staining band intensity TEM difference data presented in Chapter 4 are shown in Figure 6.24, aligned along the abscissa. With respect to the 2 week ribose comparison (Figure 6.24A), there are similarities. The sizeable significant reduction in relative *b1* and *b2* band TEM positive staining intensity (shown by positive peaks, as previously described in Chapter 4) resulting from the 2 week ribose incubation, appears to be matched by an increase in relative electron density across these bands. Similarly, the increase in the *d* band positive staining intensity is also reflected in the electron density plots. However, for the *d* band, there seems to have been a shift of the location of the electron density change to the right. This may in part be due to the buffer-related *d* band splitting seen in Figure 6.11. The small change in the *c2* band's relative intensity also appears to be matched, but by an electron density change band shifted to the right (N→C) by 5%. This is precisely the magnitude of *c1-c2* buffer-related shift suggested to occur in section 6.3.1.4. As the positive staining bands were not seen to shift (Figure 6.23), an apparent positional discrepancy is seen here.

The major discrepancy in the comparison between the two ribose plots relates to the biggest TEM positive staining band intensity change reported for the 2 week ribose incubation, i.e. the peak covering the *a1-a2-a3* bands. There is no convincing match in electron density change. This is consistent with the fact that the positive staining intensity changes reported in this region in Chapter 4 appeared to be due to a non-glycatory binding effect (see section 4.3.3 and Figure 4.14), as was discussed in section 4.4. The electron density changes seen in the *e1-e2* bands appear less significant in comparison to the positive staining intensity change peak heights from Chapter 4, if one compares the relative peak heights with those for the *d* and *b1-b2*

bands. As the non-glycatory changes referred to above also appeared to be spread across the *e* bands, this might be expected.

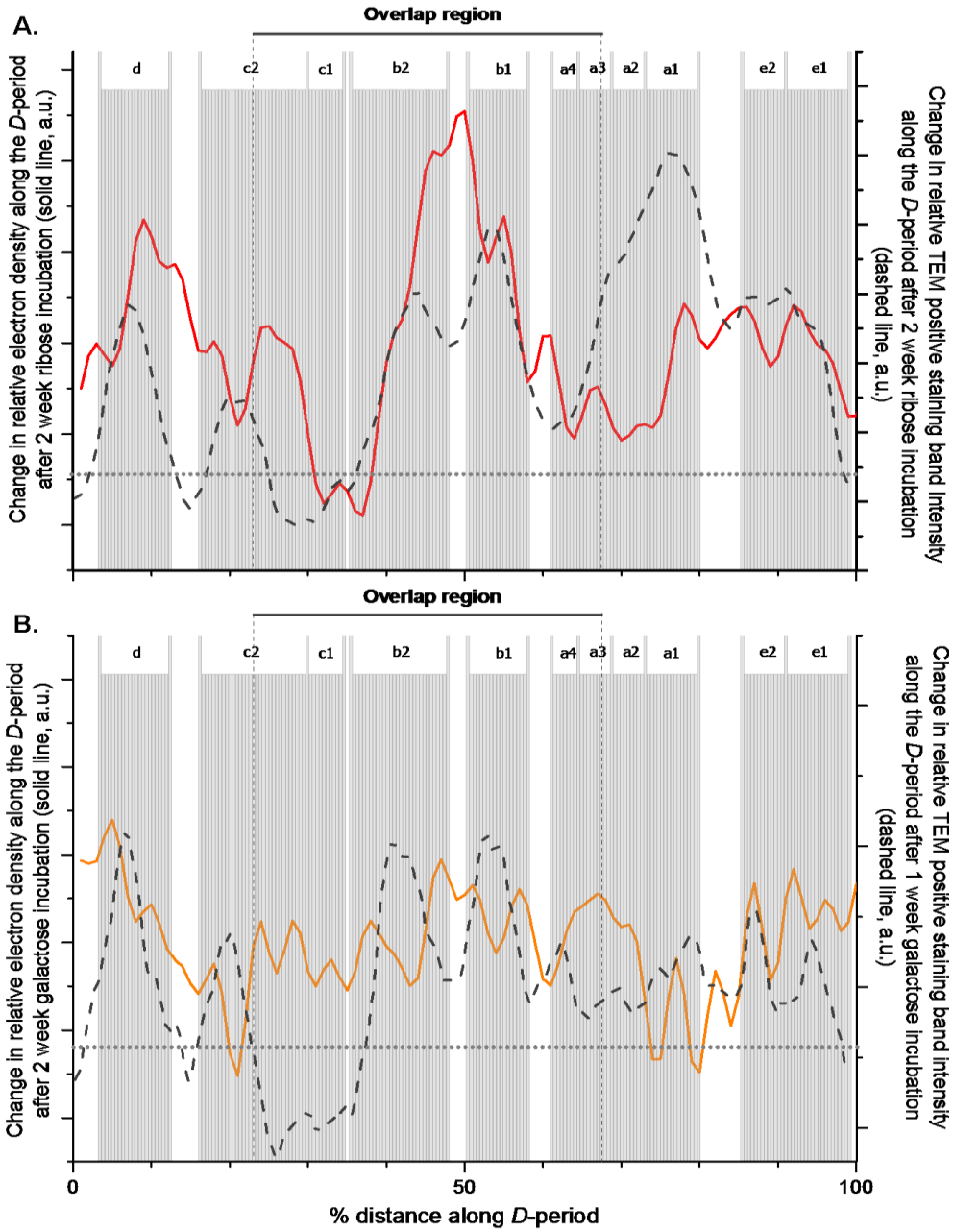


Figure 6.24 Entire 2 week glycation electron density difference plots (solid coloured lines) overlaid with TEM positive staining relative band intensity difference plots (dashed gray lines) from Chapter 4, section 4.3.3, for **A.** ribose and **B.** galactose, for comparison. Note that for ribose, the 2 week TEM difference data are presented (see Chapter 4, Figure **4.10**A) whereas for galactose the 1 week TEM data is overlaid (see Chapter 4, Figure **4.12**B) as no significant site/band specific changes were measured for the 0 → 2 week time period. The horizontal dotted lines are the lines of significance presented for the TEM data in Chapter 4, for reference. The ± 1sd lines have been omitted for clarity.

Any relationships between the two data plots presented for galactose, are less clear (Figure 6.24B). In this case the electron density difference plot for 2 weeks of galactose incubation is overlaid with the TEM positive band intensity changes associated with 1 week of incubation in galactose, as no significant location-specific positive band intensity changes were identified as being associated with the 2 week incubation in this sugar, in Chapter 4. Therefore any association between peak positions could be taken as being indicative of a lag phase between the initial galactose reaction with amino acid amines and the progression of further reactions to form more electron dense AGEs and associated cross-links. The significant specific TEM positive banding changes for galactose after 1 week were relative reductions in the *b1-b2* bands and the *d* band, with a more marginal reduction in the *a4* band. The *d* band does appear to have some increase in electron density associated with it. Any corresponding increase in electron density in the *b1-b2* bands is less clear, especially against the noisy background seen. There is also a peak showing an increase in electron density in the *e2* band which appears to correspond precisely with a decrease in positive staining after 1 week of galactose treatment. Although this peak was not considered significant from the TEM work (see section 4.3.3, Figure 4.10A), the closeness of these two peaks in terms of position and relative height in Figure 6.24B adds to the weight of evidence that significant glycation and AGE formation has occurred at this location.

Except for the *d* band, the 2 week galactose electron density difference data do not show much similarity to the *significant* 1 week galactose TEM changes presented in Chapter 4. It appears from the results in that chapter that the magnitude of the changes attributable to glycation are smaller for galactose than ribose (see section 4.3.3 and Figure 4.14), as expected (Bunn and Higgins, 1981). Therefore any increases in electron density as a result would be expected to be smaller for galactose. Also, as any positive staining TEM changes after 2 weeks' galactose incubation were non-specific (Figure 4.10), electron density changes over this period might be expected to be similarly non-specific, as appears to be the case in Figure 6.24B. The *d* band electron density increase for galactose after 2 weeks, which apparently corresponds to a positive staining TEM change after 1 week, could be evidence of a lag-phase between sugar-amine adduct formation and AGE appearance at this location. Similarly, this could be true for the corresponding peaks in the *e2* region.

As in Chapter 4 the interaction of sorbitol with MTT collagen is considered. This polyol was initially selected as a control non-glycating sugar, apparently a reasonable choice following its previous successful inclusion as a control in Tanaka et al.'s (1988a) X-ray

diffraction study of RTT collagen glycation, and in the light of a lack of any biochemical "glycation" products reported in Chapter 3 (section 3.3.1, Figure 3.1). Nonetheless significant changes in the relative TEM positive staining band intensities were reported after 2 and 4 weeks, most probably due to binding of the sorbitol along the *D*-period as presented and discussed in Chapter 4 (see section 4.3.3 and Figure 4.14, and section 4.4.2). The data presented in Figure 6.20C show a discrepancy between the electron density difference plot calculated using electron density maps for sorbitol and the buffer control constructed from structure factors of orders 1 to 22 ("entire" maps) and that calculated with the exclusion of the 1st order data. The latter shows changes which can probably be disregarded as being too small to be of significance. The difference plot which included the 1st order in its calculation appears to contradict this, inasmuch as showing some large, significant increases in the *a3-a4*, *c1-c2*, *d* and *e1-e2* positive staining band regions. In view of the close relationship between the electron density difference plots calculated both ways for ribose and galactose – including the magnitude of the scaling – the basis of the sorbitol discrepancy is puzzling. It has been reported that the polyols bind to serine residues in the *X* position of the Gly-*X*-*Y* triplet via hydrogen bonding along the collagen molecule, without adversely affecting its stability or electrostatic charge distribution (Usha et al., 2006). Also, with regard to its effects on the detection of removal of positively charged amino acids over 2 weeks, as studied in Chapter 4, associations with hydrophilic regions appear to be relatively important (see section 4.4.2, Figure 4.16 and section 4.4.3.1, Table 4.2). It could be that the sorbitol is therefore binding generally along the *D*-period and affecting the interaction of the protein with, for example, water molecules and so affecting the overall electron density along the unit cell. It is difficult to envisage specifically how this would affect the 1st order amplitude without affecting the amplitudes of other major structure factors, especially as the distribution of the 63 *X*-serines and hydrophilicity is not uniform along the unit cell. Nonetheless, there is a "bowing" effect on the sorbitol electron density difference trace constructed with the inclusion of the first order, where the baseline in the middle of the *D*-period centred around the *b1-b2* bands appears notably lower than that at the edges (*d* and *e1* bands). Along with the fact that inclusion of the 1st order in the ribose and galactose difference plot calculation was unproblematic, this does lead to the assumption that the sorbitol discrepancy is artefactual. Correction of the 1st order amplitude for sorbitol relative to the sum of the other structure factor amplitudes, from a multiple of 1.017 times the sum of the amplitudes of orders 2 to 22, to a multiple of 1.190 times that sum (1.190 being the average multiple from the 2 week buffer control, ribose and galactose incubations),

resulted in a sorbitol difference plot (not shown) that lacked any notable and specific increases in electron density. Therefore it was in agreement with the plot constructed with the exclusion of the 1st orders.

In summary, work has been presented here that has demonstrated first and foremost, that the isomorphous replacement phase determination and X-ray diffraction methods employed appear to be sufficiently sensitive to detect some changes in electron density resulting from formation of AGEs and also able to locate them along the *D*-period. Both ribose and galactose show some change along the *D*-period associated with glycation and apparently in agreement with the TEM data. The changes seen are more pronounced in the case of ribose, and factors such as the lower rate of galactose glycation and the effect of the chosen time period, as discussed above, are likely to be important reasons for this.

Significant changes caused to the axial electron density of native MTT type I collagen by buffer treatment have been shown, both in terms of peak height and position. Therefore in establishing a baseline, the importance has been demonstrated of adequately accounting for the effects of buffer components themselves during experimental structural studies *in vitro*.

Chapter 7: Discussion and conclusions

7.1 This study

The techniques TEM and X-ray diffraction are commonly used in biology for the study of tissue and macromolecular structure. The main purpose of this thesis was to demonstrate the feasibility of using specific applications of these techniques to study glycation of type I collagen. In doing so it was hoped that results would be obtained that would elucidate information about the preferred sites and extent of glycation along the collagen fibril. It was anticipated that the data from these methodologies would be complementary and that any differences and discrepancies that might arise could nonetheless yield further information about the collagen glycation process, and limitations of these structural techniques in its study.

In Chapter 3, the patterns of formation of products of glycation were studied for MTT treatment with glucose, galactose, fructose and ribose over 1, 2, 3 and 4 weeks. The structures of these sugars and rationales for their use in this study were given in Chapter 1, section 1.4.2. It was not the intention here to identify these products – with the exception of pentosidine, which was measured – but rather to gain an insight into the differences between the products of glycation for each sugar and therefore the existence of different mechanisms for the chemistries of AGE formation. It was hoped that, allowing for their differing predicted rates of glycation (see Bunn and Higgins (1981)) and given their structural closeness, galactose and glucose would show almost identical “fingerprints” of detected glycation products. Galactose might therefore be identified as a faster-glycating alternative and so a more experimentally-pragmatic alternative to glucose. Whilst there were strong similarities between the products detected for the two sugars, unidentified extra products were seen with galactose therefore it would not be a precise substitute. Previous biochemical evidence would seem to support a similarity between products of glycation with glucose and galactose (Syrový, 1994). However, glycation by galactose has not been studied as widely as glucose, or fructose and ribose. Future work to ascertain the nature of the novel galactose products and to see if their formation – in the light of their absence from the glucose AGE profile – is related directly to altered Amadori fragmentation products or altered kinetics of AGE formation due to the increased rate of galactose glycation, would be interesting.

Fructose was found to produce more unidentified novel glycation products in MTT collagen whereas ribose appeared to be more constrained. Whilst it was the fastest glycator, the great majority of ribose AGE product detected here was pentosidine (Sell and Monnier, 1989; Grandhee and Monnier, 1991). On the basis of the lack of other AGEs in MTT incubated in the sugar, ribose was clearly more restricted in the mechanisms it could initiate towards AGE formation. This is probably because of its apparent propensity to form a stable AGE cross-link (pentosidine) rapidly (Sell and Monnier, 1989; Grandhee and Monnier, 1991). It was interesting that pentosidine formation began to plateau between 3 and 4 weeks even though its concentration was <1 cross-link per molecule of collagen. Although significant depletion of glycatable amino acids was unlikely from this, the expansion of and loss of order within the molecular packing structure of the collagen molecules (Tanaka et al., 1988a) would be likely to be concerted and cumulative and affect further pentosidine (and other AGE) formation where earlier in the process it might have formed. In addition not all AGEs that were formed during the MTT incubations in ribose would have been detected via this method. For example CML is an AGE known to be formed during ribose glycation (Mikulíková et al., 2007; Pageon et al., 2007) and likely to be formed in significantly greater quantities than pentosidine (Culbertson et al., 2003; Pageon and Asselineau, 2005). Other undetected AGEs could have a structural effect and of course, deplete lysine and arginine residues.

It was considered desirable to include both buffer controls and a non-glycating sugar “osmotic” control in the glycation experiments throughout this project and one ramification of the work presented in Chapter 3 was the establishment of both buffer and sorbitol incubations as acceptable controls. No appreciable changes in the HPLC chromatograms were seen over the entire incubation time period for either of these groups and the 4 week chromatograms for both were similar. These controls were of particular value during the positive staining TEM work presented in Chapter 4. Previous work (Hadley et al., 1998; Hadley et al., 2001) had demonstrated the potential for using measurement of positive charge distribution along fibrils to assess removal of basic amino acid residues during glycation and AGE formation. In Chapter 4 the scope of this was extended to using different sugars over a greater number of time periods to study the progressive effects and their variations. Unexpectedly, it was found that the glycation buffer itself seemed to affect the distribution of positive charge along the collagen *D*-period. It was considered that the most likely mediator of this effect was the MOPS itself, via electrostatic attachment of the ionised negatively-charged sulphonate group to the positively charged lysine and arginine residues, which was not alleviated

by the post-incubation washing step. Therefore PTA stain binding to these sites would be blocked. However, the mechanism for any site specificity of this is unclear. The extent to which other buffers might bring about a similar effect in experiments employing this methodology is unknown. Hadley et al. (1998; 2001) used untreated collagen as their control, and so any effect of phosphate buffer has yet to be ascertained. Nonetheless it is hypothesised that MOPS might exert a possible inhibitory effect on glycation (Watkins et al., 1987; Nemet et al., 2011) via this binding to glycatable basic amino acids, therefore preventing the sites from reacting with the sugar.

A further complicating factor was the finding that even taking the buffer effects into account, the sorbitol control itself produced significant, broad recorded changes in the positive staining pattern along the *D*-period, especially initially across the *a2-a1-e2-e1* positive staining bands. Some kind of localised metabolism of sorbitol to a reactive sugar such as fructose giving rise to glycation could be ruled out. So could the involvement of an osmotic effect, taking into account previous work suggesting that sorbitol is excluded from collagen fibrils but glucose and fructose are not (Kuznetsova et al., 1998) (and presumably, galactose and ribose would not be) and yet the 2 week sorbitol difference trace correlated strongly with those seen with the “strongly” glycating sugars over certain time periods (2 week ribose, 3 week fructose and 3 week galactose). A mechanism by which less *D*-period locus-specific glycation locations might be determined was implied, i.e. by binding of sugars to more hydrophilic regions therefore concentrating them along these parts of the fibril close to potential glycation sites. As sorbitol does not itself glycate, it is considered that the sorbitol incubation difference trace effects are most likely to be brought about by sorbitol binding (see Usha et al. (2006)), some remaining bound throughout the washing step and interfering with the binding of the PTA during the TEM sample grid staining procedure.

To what extent the correlating positive charge difference traces for the other glycating sugars indicate an identical mechanism occurring to that exhibited by sorbitol, as opposed to generalised increased glycation in the *a2-a1-e2-e1* positive staining banding region by that sugar could not be answered by these data. However, the X-ray diffraction electron density data was valuable in this. The electron density difference data for MTT incubated in ribose for 2 weeks is discussed in section 6.4.1 (see also Figure 6.24A). The 2 week ribose positive staining changes in the *b1-b2, d* and (to a lesser extent) *c2* bands appear to be matched by increases in electron density over the same time period, whereas the increases in electron density concomitant with the *a2-*

a1 and *e2-e1* bands are proportionately a good deal less. Therefore it appears that in this region a secondary non-glycatory process has occurred here with ribose, as seen with sorbitol.

The “sorbitol effect” has provided an insight into a process that is detected by the TEM positive staining technique but does not appear to be related to an increase in electron density and therefore, presumably, AGE cross-link formation. The selection of any amino acids that might be preferentially glycated from the *a* and *e* banding regions generally might be seen to be problematic and due consideration and caution should be applied in attempting to determine preferentially-glycated amino acids where such sorbitol changes are seen. Nonetheless on the basis of the data in Chapter 4, the locations of a number of lysine residues along the *D*-period were suggested to have been preferentially-glycated, they being lysines 416 and 884 on the α1- and α2-chains and lysine 651 on the α2-chain, in the *d* band; lysines 264 on the α1- and α2-chains, and 729 on the α1-chains in the *b2* region; and possibly, lysines 290 and 756 on the α1- and α2-chains in the *b1* region. Surprisingly, cross-referencing of the significant difference data peaks with the simulated fibril from Chapter 2, section 2.5 produced a reasonable rationale for identifying these peaks, especially when taking into account other factors favourable to glycation such as proximity of acidic amino acid side-chains and other basic amino acids. However, as discussed, the preferred locations of glycation suggested here do not agree with other studies (Le Pape et al., 1984; Wess et al., 1990; Reiser et al., 1992b; Wess et al., 1993; Hadley et al., 1998; Mikulíková et al., 2007). The differences may be as a result of different approaches employed. This study is unlike the other cited *in vitro* investigations, inasmuch as a MOPS buffering system was used. In addition although several sugars were studied, due to its moderate rate of glycation and notwithstanding that, its production of site specific changes after just 1 week of MTT incubation – which were visually apparent even from the original electron micrographs (see section 4.3.3 and Figure 4.6) – galactose-incubated MTT after 1 week was selected here as the primary determinant of preferential glycation sites. None of the other studies used this sugar. This raises the question of what constitutes a preferred glycation site? Reiser et al. (1992b) considered that glycation is an ongoing process and that preferential sites of glycation might change progressively in the light of glycation-mediated changes in collagen structure (Tanaka et al., 1988a; Bai et al., 1992) and presumably, on depletion of the “most preferred” sites. Table 4.3 in Chapter 4, section 4.4.3.3 suggests the changing relationship between MTT collagen glycation with galactose, and time. Expansion of

and loss of order in the molecular packing structure across the fibril as a result of AGE and cross-link formation (Tanaka et al., 1988a) as well as changed hydration state of fibrils as a result the cross-linking process (Miles et al., 2005) are likely to change accessibility of regions and loci to sugar molecules. Li et al. (2013) found no overall altered fibrillar water content due to cross-linking however, though this does not exclude the possibility of localised effects along the fibril. Use of collagen from different sources already glycated to differing extents, is likely to contribute to differing conclusions as to the location- and structural-specificity of the glycation process.

The X-ray diffraction work presented in Chapter 5 and Chapter 6 made use not only of meridional reflection amplitudes from recorded diffraction patterns, but also of isomorphous replacement using two heavy atoms (iodine and gold) to enable these reflections' phase determination. Therefore construction of a one-dimensional electron density map of the MTT type I collagen *D*-period in a manner similar to previous work (Bradshaw et al., 1989; Orgel et al., 2000; Antipova and Orgel, 2010) that used either types I and II collagen from other sources, was possible.

Chapter 5 shows the results and procedures involved in producing an electron density map for type I collagen from mouse tendon for the first time. The map produced correlated well with the rat tendon type I collagen map produced by Orgel et al. (2000) showing most of the salient features of the rat model - which would be expected given the similarity of the primary α-chain sequences between the two species and their taxonomic closeness - though noting that the rat model was used as a model to help decide on selection of some of the phases in the event of ambiguity. The mouse collagen model did not correlate or fit as well with the simulated electron density maps produced from the original "telopeptides folded" simulated fibril (Chapter 2, section 2.5). Various adjustments of the telopeptides in this led to a best fit with a model containing all-extended N-telopeptides and one folded, one extended α1-chain C-telopeptide (the α2-chain C-telopeptide is not considered, due to its short length). Of the models pertaining to the conformations of the N-telopeptides and C-telopeptides, the model presented here most closely agreed with that of Orgel et al. (2000), differing only inasmuch as that work favoured a both α1-chains-folded C-telopeptide model. The iodine and gold heavy atom difference Fouriers (section 5.3.5 and Figure 5.23 and Figure 5.24) tended to favour the conformations suggested by the modelling in this study. However, these difference Fouriers also appeared to demonstrate that the actual heavy atom binding sites along the entire *D*-period do not fit precisely with the linear model of constant amino acid translation used here, even along the triple helix,

due to the non-linear path of collagen molecules through the *D*-period (Fraser et al., 1987; Wess et al., 1998b; Orgel et al., 2001; Orgel et al., 2006) and differences in the translational differences along the collagen molecules, e.g. due to the presence of regions of greater and lesser flexibility (Silver et al., 2001).

Having constructed a baseline one-dimensional electron density map for MTT collagen in Chapter 5, using the same method, work was performed in Chapter 6 to study the effects of glycation on electron density along the MTT *D*-period over 2 week MTT incubations. As with the changes seen with the positive staining pattern during buffer incubations in Chapter 4, the effects of the buffer incubation on the electron density map were unexpected. The changes as compared with the untreated map (section 6.3.1.4 and Figure 6.10) indicated similar peak positions overall, but that peak shifts and changes in electron density occurred during buffer incubation, which were concentrated around the molecular termini/telopeptide regions. In the N-telopeptide region, this seemed to manifest as a contraction of the molecular termini in towards the overlap region, whereas in the C-terminal region, it produced a broadened region of higher electron density. This seemed to tie in with a change in the buffer control MTT positive charge distribution between 1 and 2 weeks, which seemed to be concentrated in the telopeptide regions (section 4.3.3, Figure 4.15A). The most likely explanation for these changes would seem to be that buffer (MOPS) is binding to positively charged amino acid residues, associated with the more flexible regions of collagen molecules (Silver et al., 2001), and differentially affecting the relative spacing along the fibril. It is hypothesised that the properties of the springier flexible regions of the collagen molecules are altered by negation of positive amino acid charges on MOPS binding. The mutually repulsive forces of the co-localised acidic (negatively charged) amino acids therefore render these regions more susceptible to the tension applied during the X-ray diffraction experiments, and this is what is causing the changed electron density map.

The glycation-related electron density change investigations in Chapter 6 showed the potential value of X-ray diffraction with isomorphous replacement in ongoing glycation studies. As discussed above, the 2 week ribose glycation electron density data helped interpret the positive staining-TEM changes from the same group. These suggested that the general sorbitol-like changes across the *a2-a1* and *e2-e1* positive staining banding regions were not mainly glycation-related, as they were not accompanied by a concomitant increase in electron density. Conversely, in the *b1-b2*, *d* and – to a lesser extent – *c2* bands, there was a corresponding large increase in electron density. In the

case of the *b1-b2* bands, this appeared to be spread between the bands, which might indicate localised spreading of reactive glycation products between the two sites, giving rise to a region of particularly high AGE formation. For the *d* and *c2* bands, increases in electron density were present, but apparently shifted to the right by the peak shifting phenomenon and therefore not precisely aligned to the respective positive staining TEM bands.

Study of the electron density map for the 2 week galactose-treated MTT showed less definitive and identifiable changes. Compared with the *1* week galactose treatment positive staining differences, these showed a consistent electron density increase in the *d* band region but little evidence of other notable, location-specific changes. Here, it is likely that either **1.** the 2 week galactose changes in MTT collagen positive staining band intensity – which were not found to show any site-specific significant changes – have largely "washed out" the site specific changes seen after the 1 week galactose incubation, or **2.** as the glycation-related changes from galactose treatment of MTT generally would be expected to be significantly less than with ribose, the galactose changes may be less amenable to detection by this method. However, on the basis of the results of this study, as there does appear to be *some* reflection of the detected positive staining TEM changes after 1 week of MTT incubation in galactose, in the 2 week galactose electron density changes, it is hypothesised that this X-ray diffraction technique would be sufficiently sensitive to detect the changes after 1 weeks incubation.

In conclusion, this project has employed three distinct experimental techniques to investigate biochemical and structural aspects of glycation of type I collagen from MTT. Using HPLC with fluorescence detection, it has been demonstrated that four different sugars – i.e. glucose, galactose, fructose and ribose – produce different profiles of glycation end-products, not all of which have been identified. Galactose's profile shows reasonable similarities to glucose, but produces significant AGEs that are absent from glucose's. Galactose could be a viable, relevant, faster-glycating alternative to glucose but only in the light of identification of these novel AGEs. Ribose has already been widely used in *in vitro* glycation studies. Its profile of glycation products seems to be smaller than the other sugars. That, plus its speed of glycation, make it a conveniently fast model, less complicated by a spectrum of different AGEs formed in a variety of ways (one could say "purer").

Positive staining TEM studies have produced evidence here of up to seven possible early-preferred glycation sites on MTT collagen fibrils, after 1 week of galactose

treatment. It has also produced evidence of a less location-specific mechanism of preferred binding of glycating sugars – and polyols – along certain regions of the fibril. The X-ray electron density data suggest that these latter are not significantly associated with increased electron density and therefore glycation events. Although interesting and worthy of further investigation, this phenomenon appears to interfere with use of positive staining TEM in isolation as a means of determining favoured glycation sites.

Using the X-ray diffraction with isomorphous replacement, amplitudes and phases of meridional reflections from native MTT could be used to construct a one-dimensional electron density map of the mouse type I collagen *D*-period for the first time, to a resolution of 2.2 nm. Using the same technique, significant glycation buffer-related changes to the electron density map were seen as compared with the native. The resolution was also lower, i.e. 3.0 nm, presumably reflecting some loss of order in the arrangement of collagen molecules within the microfibrils. Nonetheless changes in electron density resulting from glycation at preferred locations have been shown using this technique and these have enabled more complete interpretation of the positive staining TEM data.

7.2 Future work

This entire study has produced results and conclusions, which establish bases for further investigation.

It would be of interest to establish the identities and kinetics of formation of AGEs formed during these incubations. Certainly knowledge of these and especially glycation products unique to certain sugars, might give more information about the chemistry of the process. Of particular interest would be a study of the nature of the AGEs formed with galactose but not glucose. Is the reason for these differences related to the greater rate of glycation or the single, switched –H and –OH groups about a carbon atom, which is the only chemical difference between the two sugars, altering the chemistry of galactose post-Amadori rearragement?

Further use of positive staining TEM in the investigation of glycation along the type I collagen fibril would be useful. In view of the apparent buffer effects seen with the system used here, it would be useful to confirm that the week-on-week MOPS changes recorded were not artefacts of using samples from different individual mice (see Chapter 4, Figure 4.15 and section 4.4.1). It would also be interesting to conduct a

comparative study using other appropriate buffer systems, most especially phosphate buffer, to establish whether or not such effects show a similar specificity along the collagen *D*-period. It is hypothesised that in the case of phosphate buffer, as there is a greater concentration of negative charge on the phosphate ion - which also forms the basis of its buffering mechanism (unlike an organic tertiary amine group in the case of MOPS (Good et al., 1966; www.sigmaaldrich.com, 2013a)) – its avidity of binding to positively charged amino acids, and so therefore the rate and magnitude of its effects, might be greater than with MOPS. Conducting similar TEM studies to this one with phosphate buffer might shed light on the nature of its catalytic effects and its influence on preferred sites of glycation. In addition further investigation of the non-glycatory binding effect of sorbitol and the glycating sugars along the *D*-period, noted in this study, would be of interest. It appears that this effect is a manifestation of sugars interacting with and binding to fibrils in a non-covalent way, but as buffer interaction with the fibril also occurs concurrently, there is a likely buffer involvement in the binding pattern. It is hypothesised that substituting phosphate buffer for MOPS – and therefore a relatively small, non-organic buffering species for a larger, more hydrophobic zwitterionic buffering ion – would alter the pattern of non-glycatory sugar binding. Study of this might elucidate more information about the bases of both buffer effects and sugar binding.

It appears that use of the X-ray diffraction methodology described here is valuable for the study of glycation-related electron density changes. Further glycation incubations, especially corresponding to the treatments and time periods found to exhibit significant changes using the positive staining TEM, should be studied using X-ray diffraction in this way. More information about the sensitivity of this method and the relationship between sites of removal of positive charge and increased electron density could be further gleaned. The two techniques appear to be complementary and have great potential for use together in structural studies of glycation. Further *in vitro* work could be used to conduct comparative structural studies of AGE formation without a chelating agent and perhaps with added transition metal ions and oxidation potentiators such as H_2O_2 (see Fujimori (1989), Culbertson et al. (2003)), to study changes in the nature and site-specificity of the changes recorded.

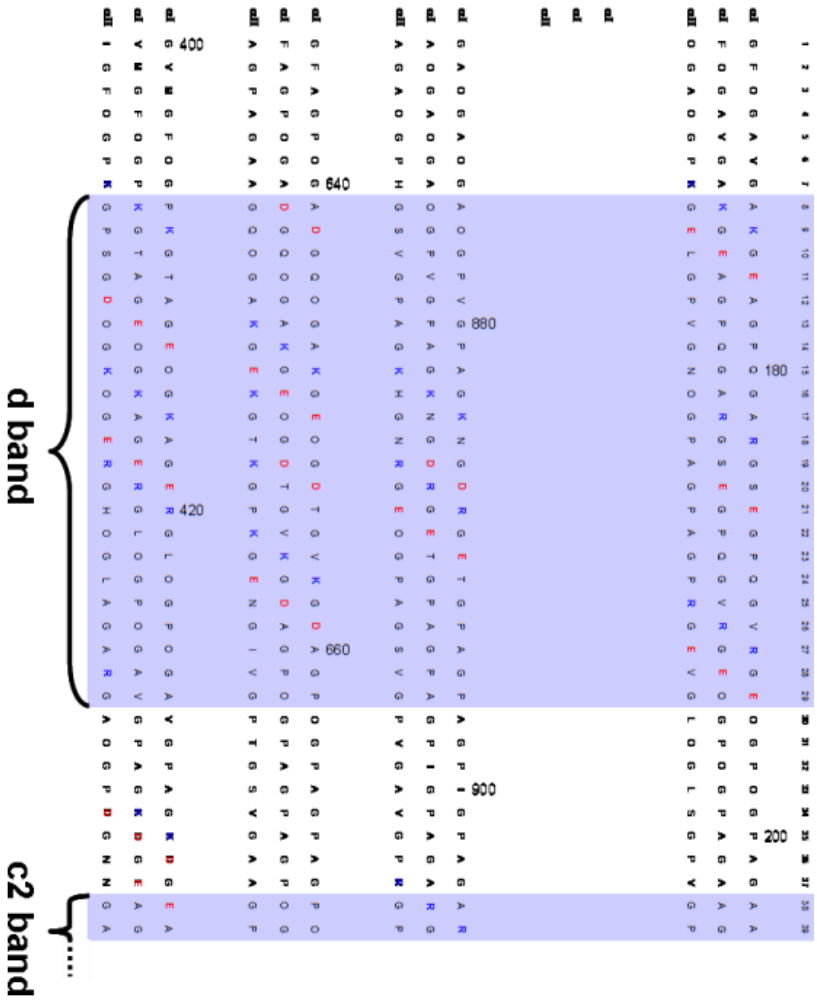
This investigation was carried out entirely *in vitro*. There are clearly ongoing applications for these techniques in the study of *in vivo* glycation models of ageing and diabetes. Certainly study of changes in collagen due to glycation *in vivo* would have the advantage of not being buffer-dependent although other variables such as inter-

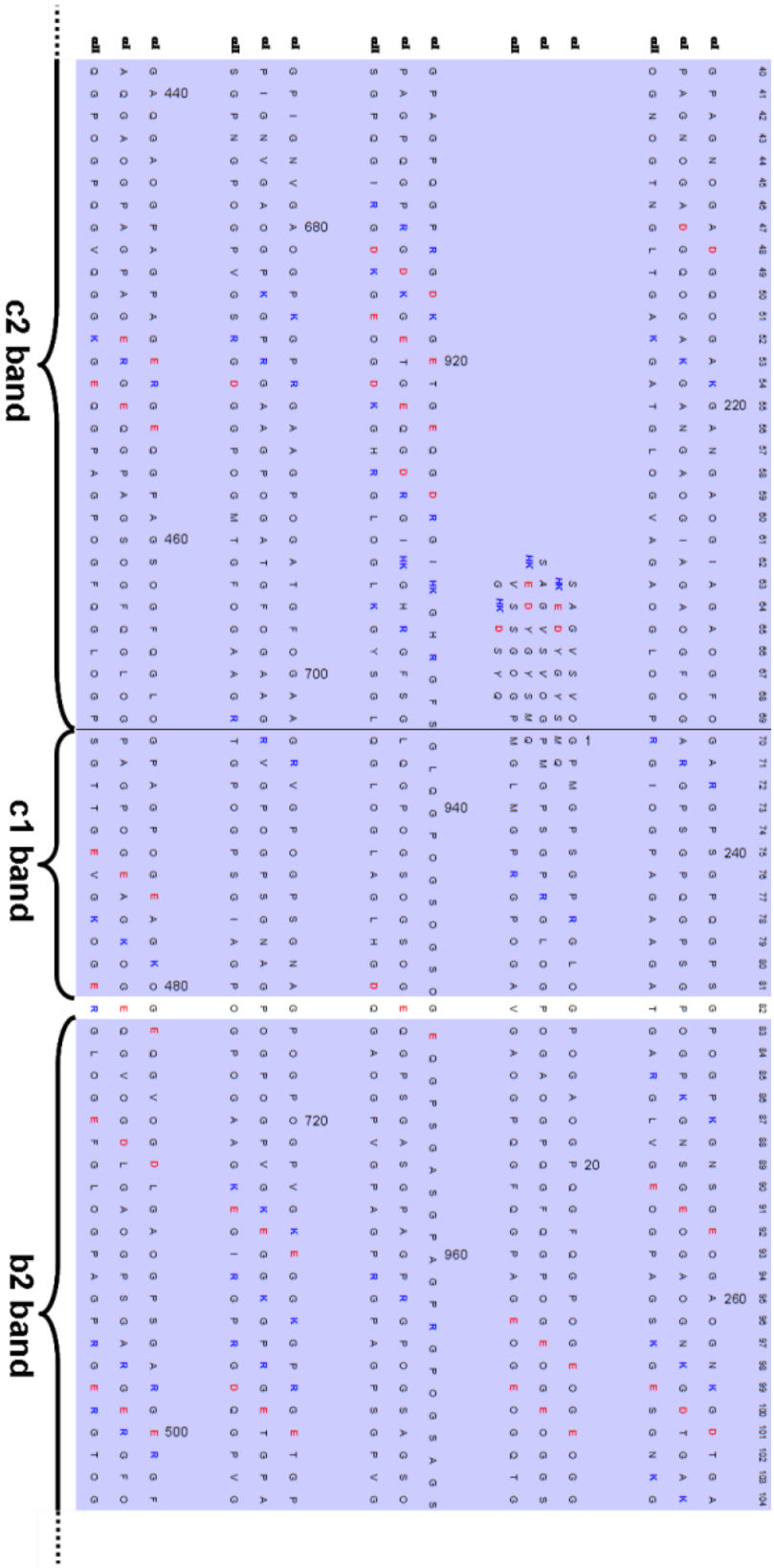
individual variation and metabolic and other stresses and factors dependent on the model being studied would have to be allowed for. In addition, as the physiological concentration of sugar (glucose) *in vivo* would be considerably lower than the 0.2M concentration studied here even in a diabetic model, the study timeframes would, of necessity, be longer.

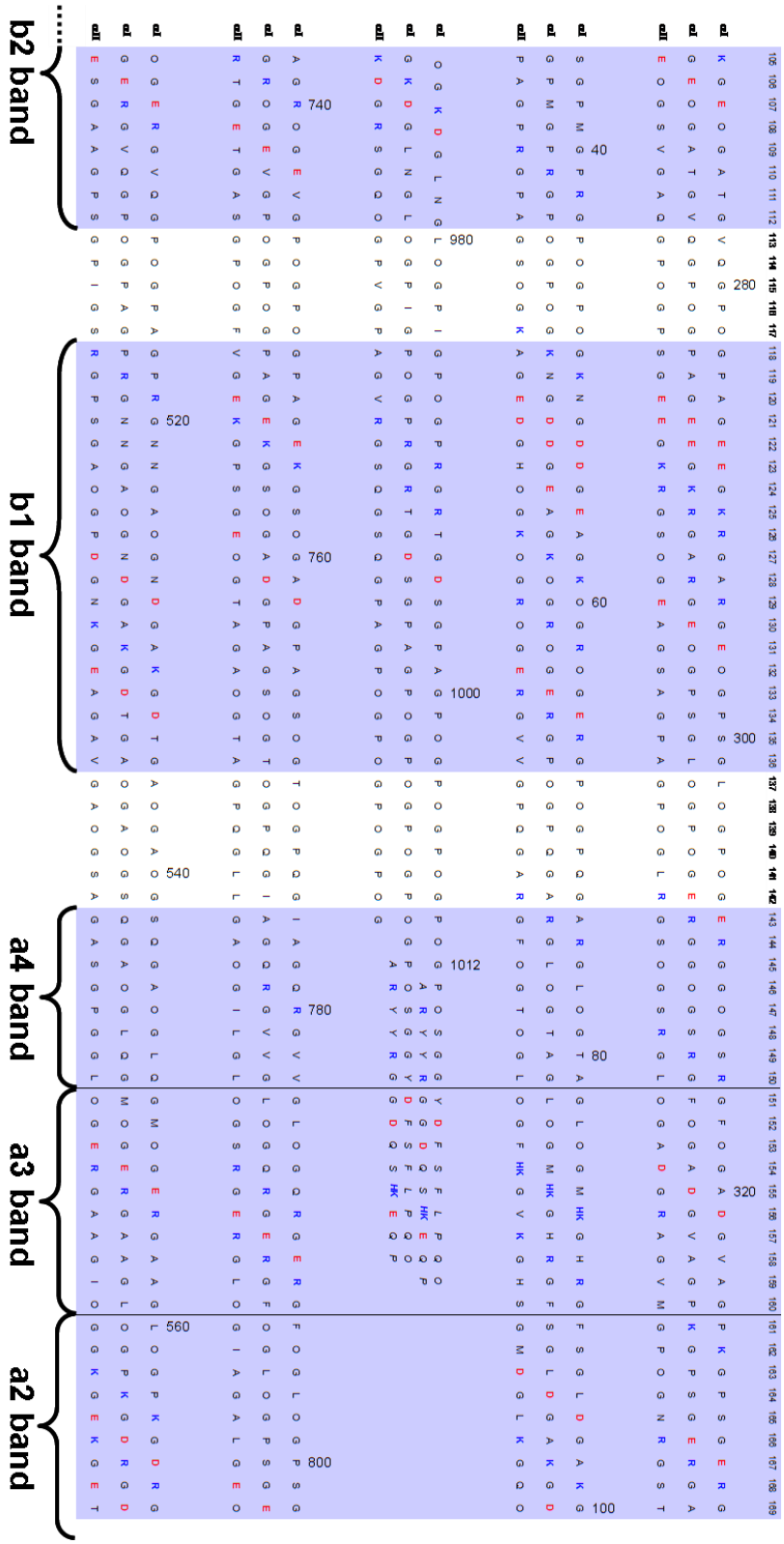
The use of mouse tendon in this study was in anticipation of an accelerated ageing model in mouse collagen for use in *in vivo* studies (Private communication from Vincent Monnier, Cape Western Reserve University, Cleveland, Ohio, USA). Therefore here, a baseline has been established using MTT tissue for continuing studies utilising this model. There is continuing scope for building on the one-dimensional electron density maps produced for MTT. In a manner analogous to that employed in RTT (Wess et al., 1998b; Orgel et al., 2001; Orgel et al., 2006), a three-dimensional map might be constructed and it is hoped that the locations and nature of the glycation related electron density changes might then be mapped onto that in the future.

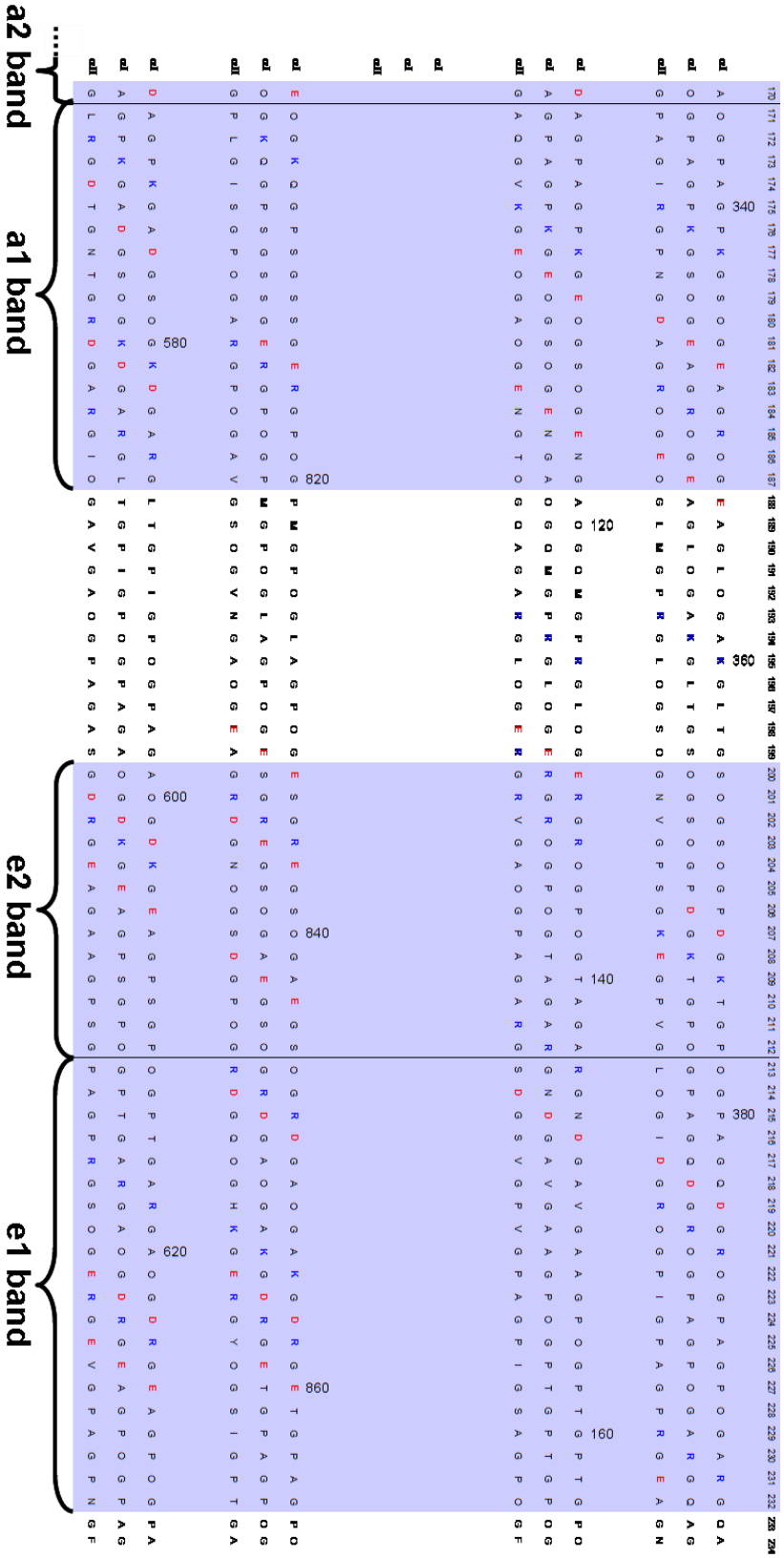
Appendix 1 – Simulated type I collagen fibril from *Mus musculus*

Entire structure of one *D*-period of the simulated fibril constructed as described in Chapter 2, section 2.5 (rotate 90^{o} left to view), based on amino acid sequence information (UniProt, 2012). Standard one letter amino acid notation used. Basic amino acids are highlighted in blue and acidic residues in red. O is hydroxyproline and HK shows possible hydroxylysine positions. Telopeptide amino acids are shown in italics. Blue-gray highlighted regions are positive staining bands, labelled at the foot of the *D*-period. Numbers by upper α1-chain amino acid positions denote triple helical locations. Numbers along the top show simulated fibril position. Continuous over this and the three following pages.




c2 band
c1 band
b2 band


b2 band
b1 band
a4 band
a3 band
a2 band


a2 band
a1 band
e2 band
e1 band

Appendix 2 – Table (split) of all derived phases for native, untreated MTT

Scaling group	11	12	13	14	15	16	21	22	23	24	25	26	31	32	33	34	35	36
Order																		
1			1.2371			4.6545		2.2660	4.9216	0.8609	2.3455	4.9347			1.0819		4.9912	4.4034
2									4.4702		0.2296				4.7240			
3												0.4687						0.4494
4			1.4469	5.9828		1.4173	0.1621	0.1782	1.1190	2.9881	0.1633	1.0886	0.4452			3.1682	1.8784	
5																2.4625		
6		2.5858	3.1699	1.7800	2.6220		3.0770	5.4657	2.0830	2.7728	5.3807		2.3669	2.4091	5.9187	2.4309	4.4941	6.1224
7					6.2564	5.5299			0.8659	2.2612	6.2247		3.2598	1.5895		5.8013		5.3628
8	3.8193	2.9176	5.1692		1.0549			2.8812	0.6167	4.5254	5.0977	2.3776	3.7285	5.4873	4.0546		5.4066	1.8685
9				5.5612	0.0035		3.6781		0.9340	3.5968	6.1789	3.7054	3.7919	1.7564	3.9484	3.6667	1.9560	
10		0.0038	5.9048									5.6629		3.5256				
11			4.1629								3.5200	5.7125					0.5765	
12									4.6258	4.9891	4.0769	1.7463	2.7938	1.7317			1.6561	
13			3.5288		0.7645	4.0957				4.6630	4.1150	4.1422	2.9339	4.1107	2.7337			
14																		
15		0.6178		4.7685			5.3536		2.8512	5.2541	5.7070	2.8151	2.5842	4.5618	3.1004		1.8830	5.9621
16						0.5743				3.5214		0.2313	1.7948	3.0619	1.1223		0.4126	3.8500
17													4.4974	4.6931				
18										1.2514	2.2258	5.1302	3.4039	4.4593				
19					3.4461		2.8882		4.5404		3.9431				1.9577		1.3395	
20		2.3084		5.0294		1.6947				5.4265		5.3768	3.2674	5.4066	3.5663			
21					2.3296			5.3960		1.2347	5.3748		1.9045	2.4398	2.4762		2.4350	
22			0.0636	4.1254		3.6208	0.3688		0.3408	0.4757	0.2609	3.9001			0.5684	5.2928	4.1994	4.0413
23		3.0998	3.9960		5.2293		0.2049		0.6670	0.5558	1.9949		3.3211	3.0745	3.3465		5.1075	3.1302
24		4.9274	4.3523	5.1877	4.7162	4.2876	5.7453	5.3179	5.0245	5.7608	5.2738	5.0531	2.8493	2.4866	2.2708	2.8434		
25		0.6007			1.5671	0.6894	3.7784	4.5120	1.2063	1.3468	4.6333	0.8775	4.1064	1.2940	4.1486	4.6478	1.5374	3.7191
26				4.9819	0.7842	5.6984	3.8587		0.9901	3.8358	0.4587		3.9323	5.0936	4.0358	4.0017	1.3231	6.1013
27			2.8389	4.9201	5.3431	5.5396	0.1008	0.2656	2.7031	0.0716	0.2623	0.0561	4.1965	0.4088	3.6769	0.5549	0.4148	0.0631
28			3.0631	5.5562	4.7313			0.6307	1.5072	5.7505		1.5290		1.7031	5.3582	5.1504	5.3642	1.7319
29			0.4687	5.4500	5.5299		6.1297		0.0112		5.6734	6.2719				4.6558	3.2057	1.3802
30		5.5354		1.3600	5.4494		5.0936		4.5062				1.8830		4.4925	1.7785		4.6449

Table (split between this and next page) of all individual phases determined for untreated MTT using the isomorphous replacement/Harker method, described in Chapter 5 (see Sections 5.1.1.2, 5.2.5.4 and 5.3.4. Numbers in black represent phases determined by precise three-way intersects, blue is near-precise and brown represents less clear-cut. No number indicates no phase could be determined with any confidence.

Scaling group heading numbers represent groups of difference Patterson scaling factors. The first digit represents a set of **gold** scaling factors, as follows:- **1**x: a=0.07, c=0.970; **2**x: a=0.09, c=0.970; **3**x: a=0.11, c=0.970

The second digit represents a set of **iodine** scaling factors, as follows:- x**1**: a=0.0450, c=0.90; x**2**: a=0.0575, c=0.90; x**3**: a=0.0700, c=0.90; x**4**: a=0.0450, c=0.92; x**5**: a=0.0575, c=0.92; x**6**: a=0.0700, c=0.92.

In ALL cases, b=4.3.

Scaling group	41	42	43	44	45	46	51	52	53	54	55	56	61	62	63	64	65	66
Order																		
1		1.7761	0.7974		4.6531		2.2857		0.9541		4.9039							2.7477
2	4.0357	5.4925	1.0865					2.5972							4.5090	5.5405		
3				1.1499	6.2389		1.7415					3.7514		4.2819	4.2945			
4	0.9849		1.8415				0.5662	2.6865			2.6089	0.6564	3.5017	6.0355	4.5491	0.0339	6.0189	4.5801
5															4.1985			4.2744
6	5.3882			5.4350	4.2884		4.7542	1.9659		4.9943	1.9238	3.6846		2.0596	5.9733	1.6175	1.9856	6.1575
7		0.9570	0.3718	1.7844	3.5376		0.4620	1.3286	3.5625			6.0023		4.6169		2.2607		2.2708
8	0.3421	5.7086		4.4906		5.7589		0.2569		4.6258	4.7840		1.1089	2.7367	4.4263	1.3231	4.6327	
9			4.1986	3.9948			3.0850		0.6670	3.0309		3.3073	3.1156	5.7467	3.9975	3.0619	5.7852	0.4518
10							0.9038							3.7220				
11	1.8613	5.6441	1.5878														1.0729	
12	5.7540		4.8061			1.8538	5.6228					5.0485			4.7308			1.8403
13			5.6959	4.6272	0.8183			0.9176			0.9714	4.1629				1.5968		0.9130
14				0.1280	3.1874		0.4285						3.8252					
15		4.1797	6.1926	4.9039	1.3407		5.0222		5.6480	1.4236	2.1806	5.6055	4.5847	0.4654		4.5047		
16	1.7097					0.4882		3.1606	1.6042	0.7270			1.2626	3.0625	4.5589	3.7843		0.6643
17	4.6439		1.2875				4.1445							1.0070				
18											2.2258	2.0024					2.2391	
19		3.6767						4.3532	4.4683	1.7832						5.3865		
20			0.5998	4.5503	0.3858					1.8712		1.7652		2.2031	6.2196			
21	4.7775			4.6540		3.4408					5.7812	0.1650		5.4381			5.5032	
22		3.1946			0.4895	1.3271	4.2876			4.3382	3.8629		3.3649		0.0941		4.2797	3.7327
23	6.1911	3.6392	0.1056		5.7782	6.2756	2.8294	0.2768	0.6017				3.0700	0.1015	0.3361	0.8978	2.2496	0.3757
24			1.3084	1.6547	1.4849	1.2031	0.0650					3.8241	5.5401	5.4190	1.6278	0.4340	1.7169	1.4300
25	0.7130				4.7970	0.7406		4.6706				0.7270		1.2080	4.4234	1.3451		0.6352
26					4.1424	5.6278	4.3015	1.2992	4.5641	4.3297	1.2418			5.5247	4.6258			
27		0.1739	4.4520	5.4752	0.2051	0.7860		1.2514		0.4485	1.2090	1.4981	4.4002	1.0953			1.1423	1.1256
28		0.7687	1.4686	2.4572		1.4424	1.1522		4.7622		4.9264	4.7086		1.6126	4.3639	2.4838	5.3431	4.2302
29		0.0227				1.6652	4.7694			2.0395					2.9233			6.1151
30	5.2268		0.8469	5.3192			4.7542			4.7086	0.8394	2.4045	2.1035			2.1433		4.4867

Table (split between this and previous page) of all individual phases determined for untreated MTT using the isomorphous replacement/Harker method, described in Chapter 5 (see Sections 5.1.1.2, 5.2.5.4 and 5.3.4. Numbers in black represent phases determined by precise three-way intersects, blue is near-precise and brown represents less clear-cut. No number indicates no phase could be determined with any confidence.

Scaling group heading numbers represent groups of difference Patterson scaling factors. The first digit represents a set of **gold** scaling factors, as follows:- **4**x: a=0.07, c=0.985; **5**x: a=0.09, c=0.985; **6**x: a=0.11, c=0.985

The second digit represents a set of **iodine** scaling factors, as follows:- x**1**: a=0.0450, c=0.90; x**2**: a=0.0575, c=0.90; x**3**: a=0.0700, c=0.90; x**4**: a=0.0450, c=0.92; x**5**: a=0.0575, c=0.92; x**6**: a=0.0700, c=0.92.

In ALL cases, b=4.3.

Appendix 3 – Table (split) of all derived phases for the 2 week buffer control MTT

Scaling group Order	11	12	13	14	15	16	21	22	23	24	25	26	31	32	33	34	35	36
1	4.7921		3.2057	4.8522	3.0550	1.4075	4.8532		3.5695	4.9510	3.1378	1.8735	3.5695	4.5309	2.9323		3.0243	1.0558
2	1.2467				0.4001					5.6374	0.1409	5.0510			0.2961		0.2550	
3	0.8299	0.9235	4.8072	5.1135		4.8288	0.1842			4.7384	2.6829	6.1499	4.8222	0.3801	4.8882	4.1759	2.7925	2.1910	4.4178
4		0.9221	3.5832	4.2436	0.9880		2.8793	1.5868	4.1655	5.3798	1.5671	4.1629	5.9354	1.0549	3.1137			2.5818
5	0.3784	4.4097		0.3812	4.4315	4.3144	1.7965				5.2413	4.6124			3.9165			1.1757
6	2.6497	4.7921	4.9766	2.7036	0.2788	3.1303	1.7254		2.1163	1.7667	0.5534	0.5420	6.0739	2.1578		1.5822	0.4653	2.9501
7		1.9968		3.9163			5.2930				3.2439		5.0596	1.3558		1.1281	4.0417	
8	3.6036	1.9864	5.4124			5.2682	3.8368			1.7816	1.1253			1.0623				4.8372
9	0.8277	1.1611		3.0150			3.9216	3.5712					3.4443	0.6079	5.6333	5.3283	3.0234	3.4681
10		0.1082	1.2797			1.3429			6.1434	2.2226	4.9454				1.6194	0.1008		1.6349
11	5.6433	1.4354		0.9025	1.3824	4.7087	6.1584		2.3089	4.3382	1.6120	0.9776	3.5214	4.3133	5.6795	4.1297	4.3270	4.9653
12	4.8980		5.1034		2.2660	1.3249		1.5600	5.0866	3.8416	5.3642	1.2900			5.3335	3.8300	4.3015	
13	3.3820	0.9479	1.7002	2.8151	1.8293	0.5455	0.5932	4.3409	4.2302	0.1755	5.5241	4.1380	0.5765	0.6591	4.3784	2.8793	1.9282	0.5582
14	3.8902	1.4998	3.5595		1.5067	3.7118	2.8085	5.3912		2.6165	5.3748				5.8405	1.9826		
15	1.3940	2.7674	3.6910	1.3926	2.7771	0.8171	3.9538	1.7403	2.7506	4.0123	1.7243		3.4225	1.2847	2.0092	3.4443		5.2779
16		5.1440		1.8784	5.0635	4.2270	3.6036	5.1642	2.6184		5.0588		6.2291		0.2562	2.2403	4.7162	4.1255
17	0.4305		0.8012	0.5765	5.6300		1.8662	0.4689	0.4380			5.7975	1.4584	2.7068	3.0775	4.2504	6.1457	
18	5.0866	0.3147		0.6591		2.1008	5.0628	0.4686		3.2409	5.1576	2.1138	5.0993	0.2882	3.3211	3.5932	1.7021	2.0632
19	2.3190	1.9407	6.2345	2.3186	1.9618		4.4628	3.3111	5.6897	0.6074	3.2974	3.0177	4.0181	2.7307	6.1509	1.6548	2.7199	3.5495
20	3.9968	3.7707	6.0739	4.0666	3.8146	6.1274	0.2882	1.9925		0.3194	1.9826	0.9250	4.5837	2.5947	0.2166	4.5150		0.3145
21	1.5214	0.3286				4.1965	4.8831	0.0859	2.4538		2.1699	4.5791	5.6795	0.4525		3.3450	4.9755	4.2802
22	2.8483	6.0189	5.1811	5.6149	2.5858		2.3670	4.7765	3.7224	5.8312	2.0280	0.1621	0.2003	5.6647	3.5465	0.1082	2.2180	0.6683

Table (split between this and next page) of all individual phases determined for the 2 week buffer control MTT using the Isomorphous replacement/Harker method (see Chapters 5 & 6, Sections 5.1.1.2, 5.2.5.4, 5.3.4 and 6.3.1.3). Numbers in black represent phases determined by precise three-way intersects, blue is near-precise and brown represents less clear-cut. No number indicates no phase could be determined with any confidence.

Scaling group heading numbers represent groups of difference Patterson scaling factors. The first digit represents a set of **gold** scaling factors, as follows:- **1**x: a=0.10, c=1.04; **2**x: a=0.12, c=1.04; **3**x: a=0.10, c=1.05.

The second digit represents a set of **iodine** scaling factors, as follows:- x**1**: a=0.140, c=1.08; x**2**: a=0.165, c=1.08; x**3**: a=0.190, c=1.08; x**4**: a=0.140, c=1.10; x**5**: a=0.165, c=1.10; x**6**: a=0.190, c=1.10.

In ALL cases, b=4.3.

Scaling group	41	42	43	44	45	46	51	52	53	54	55	56	61	62	63	64	65	66
Order																		
1	0.6246	3.7707	2.8222	5.3098	3.5732	1.1121	4.3075		2.5882	4.4441	2.9634	2.2996	1.0819	0.8659	0.2695	1.0579	2.8882	1.9993
2	0.6442		0.2688	3.0293	4.2471	0.4464	1.8925	1.7497		5.4505		5.3293	1.4517	3.2664	4.2930			1.4905
3		1.2992	1.5593	2.1744	5.1003	1.6493	2.6108			2.2060	6.0976		0.3609		1.2305		2.5342	1.1423
4	6.1199		5.5437	0.0403	1.3032		0.4884		5.2180		1.9256		3.1226	3.6171	3.4078		0.2108	3.3729
5	3.0785	2.0530		3.1001	2.0328						5.6208	4.3115		5.0503			5.1340	
6		5.1440		1.4842	3.5285	3.0550	5.2187		2.5170	4.9799	3.8898	2.6527	6.0355	1.5746	5.4715	1.4517	3.2282	0.6952
7			4.0651	3.6509	3.2674	0.6428	2.2921			2.6506	3.4514	3.0582	2.4792		5.0964	2.7969	3.3673	
8	3.0770					1.9212			0.9144				5.5346	3.3010		2.2708	3.0401	2.2060
9	0.9050	0.9653		3.0731		3.8086	1.2371	0.8226	5.1744	5.2667	3.2132		0.4718			2.6425	0.0797	0.6352
10		3.7223	5.3854		0.9356		5.5506		1.9688	5.5189	6.0446		1.8998	5.9090	1.7874	1.9256	2.3304	
11	3.1154	2.1075	2.5097	4.2504		4.5611	5.5687	4.7618	5.9893	0.8552	4.6935	3.9804	0.0727	1.6574	6.1846	0.6821	2.0426	4.4336
12	5.4214	3.9743	5.4514				5.2913			5.2346	5.2871	0.4285		5.1340		4.6183	5.1340	0.7801
13	3.4443	4.3646	1.5293		2.0395	0.7149	0.7958	4.3787	4.2200	0.0641	5.6680	4.1422	3.3893	0.7381	1.4138	2.6992	1.9309	0.6159
14	3.2038	2.1202	3.1007	5.7775	2.1266	0.4552	2.8491		0.5638	2.6896		0.5558		1.5147		0.0337	1.5218	1.1687
15	5.4214	3.5256	1.0278		0.5289	0.9818	2.0154		3.6304	2.0601	2.6695	0.3458	1.6947	2.2073		1.6574	2.2815	5.8714
16	0.5351		0.1173				0.2531	1.7726	2.9176		1.7711	0.3433	0.1212	4.8782	0.2214	2.4177	4.7996	4.0635
17	2.2485	3.3656	1.6731	2.2651		1.6424	2.0463	3.7390	1.8403	2.1590		1.8590	0.4819		2.8058	0.4853	4.8297	0.0716
18	1.6418	3.1529		0.5932	5.3240	2.0530	5.2375	0.3098	3.0800	3.4204	1.7180	4.8980	5.0936		0.7596	0.7227	1.7106	4.8819
19	3.3673	2.5922	0.1313	0.8872	2.5986	3.9007		3.4393	6.0501	2.9975	3.5805	3.4204	3.6820	4.3284		6.1708	4.3716	4.7990
20	0.7164	4.2331	1.2389	3.3838	1.4236	4.5031	5.5774	2.2544	1.4213	5.5725	2.1139	4.1605	3.8736	3.4420	1.7948	3.8515	3.4681	4.5716
21	5.1642	3.0097	3.2282	2.8939	4.9747	0.8495	1.2133	0.5597	2.7912	3.5034	5.0590	0.3312	5.0936		5.7313	3.1680	5.3656	
22	1.9264	0.1165	3.4849	5.2911	3.0318	0.5534	1.0652	5.3611	5.2515	4.5254	1.8351		1.3451	1.2160	3.9165	4.6860	1.2775	0.4520

Table (split between this and previous page) of all individual phases determined for the 2 week buffer control MTT using the Isomorphous replacement/Harker method (see Chapters 5 & 6, Sections 5.1.1.2, 5.2.5.4, 5.3.4 and 6.3.1.3). Numbers in black represent phases determined by precise three-way intersects, blue is near-precise and brown represents less clear-cut. No number indicates no phase could be determined with any confidence.

Scaling group heading numbers represent groups of difference Patterson scaling factors. The first digit represents a set of **gold** scaling factors, as follows:- **4**x: a=0.12, c=1.05; **5**x: a=0.10, c=1.06; **6**x: a=0.12, c=1.06

The second digit represents a set of **iodine** scaling factors, as follows:- x**1**: a=0.140, c=1.08; x**2**: a=0.165, c=1.08; x**3**: a=0.190, c=1.08; x**4**: a=0.140, c=1.10; x**5**: a=0.165, c=1.10; x**6**: a=0.190, c=1.10.

In ALL cases, b=4.3.

References

Ahmed M U, Thorpe S R, and Baynes J W (1986) Identification of N^{ε}-carboxymethyllysine as a degradation product of fructoselysine in glycated protein. *Journal of Biological Chemistry* 261: 4889-4894.

Ait-Belkacem D, Guilbert M, Roche M, Duboisset J, Ferrand P, Sockalingum G, Jeannesson P et al. (2012) Microscopic structural study of collagen aging in isolated fibrils using polarized second harmonic generation. *Journal of Biomedical Optics* 17.

Alexander R M, and Bennet-Clark H C (1977) Storage of elastic strain energy in muscle and other tissues. *Nature* 265: 114-117.

Andreassen T T, Seyer-Hansen K, and Bailey A J (1981) Thermal stability, mechanical properties and reducible cross-links of rat tail tendon in experimental diabetes. *Biochimica et Biophysica Acta* 677: 313-317.

Antipova O, and Orgel J P R O (2010) *In situ D*-periodic molecular structure of type II collagen. *Journal of Biological Chemistry* 285: 7087-7096.

Antipova O A (2008) *The studies of collagen type II macromolecular structure by X-ray fibre diffraction, electron microscopy, and computational techniques.* Illinois Institute of Technology. PhD thesis.

Avery N C, and Bailey A J (2005) Enzymic and non-enzymic cross-linking mechanisms in relation to turnover of collagen: Relevance to aging and exercise. *Scandinavian Journal of Medicine and Science in Sports* 15: 231-240.

Avery N C, Sims T J, and Bailey A J (2009) Quantitative determination of collagen cross-links. *Methods in Molecular Biology* 522: 103-121.

Bai P, Phua K, Hardt T, Cernadas M, and Brodsky B (1992) Glycation alters collagen fibril organization. *Connective Tissue Research* 28: 1-12.

Bailey A J (2001) Molecular mechanisms of ageing in connective tissues. *Mechanisms of Ageing and Development* 122: 735-755.

Bailey A J, Paul R G, and Knott L (1998) Mechanisms of maturation and ageing of collagen. *Mechanisms of Ageing and Development* 106: 1-56.

Bailey A J, Sims T J, Avery N C, and Halligan E P (1995) Non-enzymatic glycation of fibrous collagen - reaction-products of glucose and ribose. *Biochemical Journal* 305: 385-390.

Balguid A, Rubbens M P, Mol A, Bank R A, Bogers A J J C, Van Kats J P, De Mol B A J M et al. (2007) The role of collagen cross-links in biomechanical behavior of human aortic heart valve leaflets - Relevance for tissue engineering. *Tissue Engineering* 13: 1501-1511.

Bannister D W, and Burnes A B (1970) Adaptation of Bergman and Loxley technique for hydroxyproline determination to autoanalyzer and its use in determining plasma hydroxyproline in domestic fowl. *Analyst* 95: 596-600.

Banos C C, Thomas A H, and Kuo C K (2008) Collagen fibrillogenesis in tendon development: Current models and regulation of fibril assembly. *Birth Defects Research Part C - Embryo Today: Reviews* 84: 228-244.

Bartling B, Desole M, Rohrbach S, Silber R-E, and Simm A (2009) Age-associated changes of extracellular matrix collagen impair lung cancer cell migration. *FASEB Journal* 23: 1510-1520.

Bear R S (1944) X-ray diffraction studies on protein fibres. I. The large fibre-axis period of collagen. *Journal of the American Chemical Society* 66: 1297-1305.

Bensusan H B, Dixit S N, and McKnight S D (1971) An improved enzymic hydrolysis of collagen. *BBA - Protein Structure* 251: 100-108.

Bergman I, and Loxley R (1963) Two improved and simplified methods for spectrophotometric determination of hydroxyproline. *Analytical Chemistry* 35: 1961-1965.

Berzhitskaya V V, Timofeeva M V, Rebrova G A, Vasilevskii V K, Kho S S, Rebrov L B, and Bykov V A (2002) Structural modification of collagen during nonenzymatic glycosylation *in vitro*. *Bulletin of Experimental Biology and Medicine* 133: 444-447.

Biemel K M, Alexander Friedl D, and Lederer M O (2002) Identification and quantification of major Maillard cross-links in human serum albumin and lens protein: Evidence for glucosepane as the dominant compound. *Journal of Biological Chemistry* 277: 24907-24915.

Birk D E, Zycband E I, Winkelmann D A, and Trelstad R L (1990) Collagen fibrillogenesis *in situ*. Discontinuous segmental assembly in extracellular compartments. *Annals of the New York Academy of Sciences* 580: 176-194.

Bloemendal H, De Jong W, Jaenicke R, Lubsen N H, Slingsby C, and Tardieu A (2004) Ageing and vision: Structure, stability and function of lens crystallins. *Progress in Biophysics and Molecular Biology* 86: 407-485.

Bradshaw J P, Miller A, and Wess T J (1989) Phasing the meridional diffraction pattern of type I collagen using isomorphous derivatives. *Journal of Molecular Biology* 205: 685-694.

Brady J D, and Robins S P (2001) Structural characterization of pyrrolic cross-links in collagen using a biotinylated Ehrlich's Reagent. *Journal of Biological Chemistry* 276: 18812-18818.

Brennan M (1989a) Changes in solubility, non-enzymatic glycation, and fluorescence of collagen in tail tendons from diabetic rats. *Journal of Biological Chemistry* 264: 20947-20952.

Brennan M (1989b) Changes in the cross-linking of collagen from rat tail tendons due to diabetes. *Journal of Biological Chemistry* 264: 20953-20960.

Brodsky B, Eikenberry E F, Belbruno C, and Sterling K (1982) Variations in collagen fibril structure in tendons. *Biopolymers* 21: 935-951.

Brownlee M (1995) Advanced protein glycosylation in diabetes and aging. *Annual Review of Medicine* 46: 223-234.

Brownlee M, Pongor S, and Cerami A (1983) Covalant attachment of soluble proteins by nonenzymatically glycosylated collagen. Role in the *in situ* formation of immune complexes. *Journal of Experimental Medicine* 158: 1739-1744.

Buffa R, Betak J, Kettou S, Hermannova M, Pospisilova L, and Velebny V (2011) A novel DTPA cross-linking of hyaluronic acid and metal complexation thereof. *Carbohydrate Research* 346: 1909-1915.

Bunn H F, and Higgins P J (1981) Reaction of monosaccharides with proteins: Possible evolutionary significance. *Science* 213: 222-224.

Burger C, Zhou H W, Wang H, Sics I, Hsiao B S, Chu B, Graham L et al. (2008) Lateral packing of mineral crystals in bone collagen fibrils. *Biophysical Journal* 95: 1985-1992.

Cameron G J, Alberts I L, Laing J H, and Wess T J (2002) Structure of type I and type III heterotypic collagen fibrils: An X-ray diffraction study. *Journal of Structural Biology* 137: 15-22.

Cameron G J, Cairns D E, and Wess T J (2007) The variability in type I collagen helical pitch is reflected in the *D*-periodic fibrillar structure. *Journal of Molecular Biology* 372: 1097-1107.

Cameron I L, Lanctot A C, and Fullerton G D (2012) Maintenance of low sodium and high potassium levels in cells and in tendon/collagen. *Cell Biology International* 36: 503-509.

Cantor C R, and Schimmel P R (1980) X-Ray Crystallography in *Biophysical Chemistry, Part II: Techniques for the study of biological structure and function*. San Francisco: W. H. Freeman and Company, pp. 687-791.

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

Canty E G, Lu Y, Meadows R S, Shaw M K, Holmes D F, and Kadler K E (2004) Coalignment of plasma membrane channels and protrusions (fibropositors) specifies the parallelism of tendon. *Journal of Cell Biology* 165: 553-563.

Capote F P, and Sanchez J C (2009) Strategies for proteomic analysis of non-enzymatically glycated proteins. *Mass Spectrometry Reviews* 28: 135-146.

Cerami A, Vlassara H, and Brownlee M (1985) Protein glycosylation and the pathogenesis of atherosclerosis. *Metabolism - Clinical and Experimental* 34: 37-44.

Cerami A, Vlassara H, and Brownlee M (1986) Role of nonenzymatic glycosylation in atherogenesis. *Journal of Cellular Biochemistry* 30: 111-120.

Chandross R J, and Bear R S (1973) Improved profiles of electron density distribution along collagen fibrils. *Biophysical Journal* 13: 1030-1048.

Chang K C, Liang J T, Tsai P S, Wu M S, and Hsu K L (2009) Prevention of arterial stiffening by pyridoxamine in diabetes is associated with inhibition of the pathogenic glycation on aortic collagen. *British Journal of Pharmacology* 157: 1419-1426.

Chang K C, Tseng C D, Wu M S, Liang J T, Tsai M S, Cho Y L, and Tseng Y Z (2006) Aminoguanidine prevents arterial stiffening in a new rat model of type 2 diabetes. *European Journal of Clinical Investigation* 36: 528-535.

Chanut-Delalande H, Fichard A, Bernocco S, Garrone R, Hulmes D J S, and Ruggiero F (2001) Control of heterotypic fibril formation by collagen V is determined by chain stoichiometry. *Journal of Biological Chemistry* 276: 24352-24359.

Chanzy H, Franc J M, and Herbage D (1976) High angle electron diffraction of frozen hydrated collagen. *Biochemical Journal* 153: 139-140.

Chapman J A, Holmes D F, Meek K M, and Rattew C J (1981) Electron-optical studies of collagen fibril assembly. In: M. Balaban J L S, W. Traub and A. Yonath [ed.] *Structural Aspects of Recognition and Assembly in Biological Molecules*. Vol. 1. Rehovot and Philadelphia: Balaban International Science, pp. 387-401.

Chapman J A, Tzaphlidou M, Meek K M, and Kadler K E (1990) The collagen fibril-A model system for studying the staining and fixation of a protein. *Electron Microscopy Reviews* 3: 143-182.

Cho S J, Roman G, Yeboah F, and Konishi Y (2007) The road to advanced glycation end products: A mechanistic perspective. *Current Medicinal Chemistry* 14: 1653-1671.

Chothia C (1975) Structural invariants in protein folding. *Nature* 254: 304-308.

Culbertson S M, Vassilenko E I, Morrison L D, and Ingold K U (2003) Paradoxical impact of antioxidants on post-amadori glycoxidation. Counterintuitive increase in the yields of pentosidine and N^{ε} - carboxymethyllysine using a novel multifunctional pyridoxamine derivative. *Journal of Biological Chemistry* 278: 38384-38394.

Derwin K A, Soslowsky L J, Kimura J H, and Plaas A H (2001) Proteoglycans and glycosaminoglycan fine structure in the mouse tail tendon fascicle. *Journal of Orthopaedic Research* 19: 269-277.

Di Lullo G A, Sweeney S M, Körkkö J, Ala-Kokko L, and San Antonio J D (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *Journal of Biological Chemistry* 277: 4223-4231.

Duance V C, Restall D J, Beard H, Bourne F J, and Bailey A J (1977) The location of three collagen types in skeletal muscle. *FEBS Letters* 79: 248-252.

Eble A S, Thorpe S R, and Baynes J W (1983) Nonenzymatic glucosylation and glucose-dependent cross-linking of protein. *Journal of Biological Chemistry* 258: 9406-9412.

Eyre D R, Apon S, Wu J J, Ericsson L H, and Walsh K A (1987) Collagen type IX: Evidence for covalent linkages to type II collagen in cartilage. *FEBS Letters* 220: 337-341.

Fan X, Liu X, Potts B, Strauch C M, Nemet I, and Monnier V M (2011) Topical application of L-arginine blocks advanced glycation by ascorbic acid in the lens of hSVCT2 transgenic mice. *Molecular Vision* 17: 2221-2227.

Fan X, and Monnier V M (2008) Inhibition of crystallin ascorbylation by nucleophilic compounds in the hSVCT2 mouse model of lenticular aging. *Investigative Ophthalmology & Visual Science* 49: 4945-4952.

Farquharson C, and Robins S P (1989) Immunolocalization of collagen types I and III in the arterial wall of the rat. *Histochemical Journal* 21: 172-178.

Fernandes R J, Weis M, Scott M A, Seegmiller R E, and Eyre D R (2007) Collagen XI chain misassembly in cartilage of the chondrodysplasia (cho) mouse. *Matrix Biology* 26: 597-603.

Fessel G, and Snedeker J G (2009) Evidence against proteoglycan mediated collagen fibril load transmission and dynamic viscoelasticity in tendon. *Matrix Biology* 28: 503-510.

Fischetti R, Stepanov S, Rosenbaum G, Barrea R, Black E, Gore D, Heurich R et al. (2004) The BioCat undulator beamline 18ID: A facility for biological non-crystalline diffraction and X-ray absorption spectroscopy at the Advanced Photon Source. *Journal of Synchrotron Radiation* 11: 399-405.

Franchi M, Fini M, Quaranta M, De Pasquale V, Raspanti M, Giavaresi G, Ottani V et al. (2007a) Crimp morphology in relaxed and stretched rat Achilles tendon. *Journal of Anatomy* 210: 1-7.

Franchi M, Trirè A, Quaranta M, Orsini E, and Ottani V (2007b) Collagen structure of tendon relates to function. *TheScientificWorldJournal* 7: 404-420.

Francis-Sedlak M E, Uriel S, Larson J C, Greisler H P, Venerus D C, and Brey E M (2009) Characterization of type I collagen gels modified by glycation. *Biomaterials* 30: 1851-1856.

Fraser R D B, and Macrae T P (1981) Unit-cell and molecular connectivity in tendon collagen. *International Journal of Biological Macromolecules* 3: 193-200.

Fraser R D B, MacRae T P, and Miller A (1987) Molecular packing in type I collagen fibrils. *Journal of Molecular Biology* 193: 115-125.

Fujimori E (1989) Cross-linking and fluorescence changes of collagen by glycation and oxidation. *Biochimica et Biophysica Acta* 998: 105-110.

Fujimoto D, and Moriguchi T (1978) Pyridinoline, a non-reducible cross-link of collagen. Quantitative determination, distribution, and isolation of a cross-linked peptide. *Journal of Biochemistry* 83: 863-867.

Fulwood N J, and Meek K M (1994) An ultrastructural, time-resolved study of freezing in the corneal stroma. *Journal of Molecular Biology* 236: 749-758.

Gabbay K H (1975) Hyperglycemia, polyol metabolism, and complications of diabetes mellitus. *Annual Review of Medicine* 26: 521-536.

Gathercole L J, and Keller A (1991) Crimp morphology in the fibre-forming collagens. *Matrix* 11: 214-234.

George A, Malone J P, and Veis A (1999) The secondary structure of type I collagen N-telopeptide as demonstrated by Fourier transform IR spectroscopy and molecular modeling. *Proceedings of the Indian Academy of Sciences: Chemical Sciences* 111: 121-131.

Gil H, Salcedo D, and Romero R (2005) Effect of phosphate buffer on the kinetics of glycation of proteins. *Journal of Physical Organic Chemistry* 18: 183-186.

Glišić B, Rychlewska U, and Djuran M I (2012) Reactions and structural characterization of gold(III) complexes with amino acids, peptides and proteins. *Dalton Transactions* 41: 6887-6901.

Glusker J P, Lewis M, and Rossi M (1994) *Crystal Structure Analysis for Chemists and Biologists*. New York, USA: Wiley-VCH, Inc.

Gobeaux F, Mosser G, Anglo A, Panine P, Davidson P, Giraud-Guille M M, and Belamie E (2008) Fibrillogenesis in dense collagen solutions: A physicochemical study. *Journal of Molecular Biology* 376: 1509-1522.

Goh K L, Holmes D F, Lu H Y, Richardson S, Kadler K E, Purslow P P, and Wess T J (2008) Ageing changes in the tensile properties of tendons: Influence of collagen fibril volume fraction. *Journal of Biomechanical Engineering* 130.

Goh S Y, and Cooper M E (2008) The role of advanced glycation end products in progression and complications of diabetes. *Journal of Clinical Endocrinology and Metabolism* 93: 1143-1152.

Goldberg B D, and Burgeson R E (1982) Binding of soluble type I collagen to fibroblasts: Specificities for native collagen types, triple helical structure, telopeptides, propeptides, and cyanogen bromide-derived peptides. *Journal of Cell Biology* 95: 752-756.

Goldsack D E, and Chalifoux R C (1973) Contribution of the free energy of mixing of hydrophobic side chains to the stability of the tertiary structure of proteins. *Journal of Theoretical Biology* 39: 645-651.

Gong B, Sun J, Vargas G, Chang Q, Xu Y, Srivastava D, and Boor P J (2008) Nonlinear imaging study of extracellular matrix in chemical-induced, developmental dissecting aortic aneurysm: Evidence for defective collagen type III. *Birth Defects Research Part A - Clinical and Molecular Teratology* 82: 16-24.

Good N E, Winget G D, Winter W, Connolly T N, Izawa S, and Singh R M M (1966) Hydrogen ion buffers for biological research. *Biochemistry* 5: 467-&.

Gordon M K, and Hahn R A (2010) Collagens. *Cell and Tissue Research* 339: 247-257.

Grandhee S K, and Monnier V M (1991) Mechanism of formation of the Maillard protein cross-link pentosidine: Glucose, fructose, and ascorbate as pentosidine precursors. *Journal of Biological Chemistry* 266: 11649-11653.

Gross J, and Kirk D (1958) The heat precipitation of collagen from neutral salt solutions: Some rate-regulating factors. *The Journal of biological chemistry* 233: 355-360.

Hadley J, Malik N, and Meek K (2001) Collagen as a model system to investigate the use of aspirin as an inhibitor of protein glycation and cross-linking. *Micron* 32: 307-315.

Hadley J C, Meek K M, and Malik N S (1998) Glycation changes the charge distribution of type I collagen fibrils. *Glycoconjugate Journal* 15: 835-840.

Hammersley A (1997) FIT2D: An Introduction and Overview. *http://www.esrf.eu/computing/scientific/FIT2D/FIT2D_INTRO/fit2d.html* [Online]. ESRF, Grenoble, France. Last accessed: September, 2013.

Hanson D A, and Eyre D R (1996) Molecular site specificity of pyridinoline and pyrrole cross-links in type I collagen of human bone. *Journal of Biological Chemistry* 271: 26508-26516.

Harker D (1956) The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Crystallographica* 9: 1-9.

Harper D (2008) *http://www.etymonline.com* [Online]. Last accessed: September 2013.

Hayward L D, and Angyal S J (1977) A symmetry rule for the circular dichroism of reducing sugars, and the proportion of carbonyl forms in aqueous solutions thereof. *Carbohydrate Research* 53: 13-20.

Heinemeier K M, Schjerling P, Heinemeier J, Magnusson S P, and Kjaer M (2013) Lack of tissue renewal in human adult Achilles tendon is revealed by nuclear bomb ^{14}C. *Faseb Journal* 27: 2074-2079.

Hodge A J, and Petruska J A (1963) Recent studies with the electron microscope on ordered aggregates of the tropocollagen macromolecule.*G. N. Ramachandran, Editor. Aspects of Protein Structure. Proceedings of a Symposium*. pp. 289-300.

Hodge A J, and Schmitt F O (1960) The charge profile of the tropocollagen macromolecule and the packing arrangement in native-type collagen fibrils. *Proceedings of the National Academy of Sciences of the United States of America* 46: 186-197.

Hofmann H, Fietzek P P, and Kuehn K (1980) Comparative analysis of the sequences of the three collagen chains α1(I), α2 and α1(III). Functional and genetic aspects. *Journal of Molecular Biology* 141: 293-314.

https://beam.aps.anl.gov/pls/apsweb/beamline_display_pkg.display_beamline?p_beamline_num_c=25 (2012) *Beamline 18-ID-D: Sector 18 - Insertion Device Beamline* [Online]. Available at: [Accessed:

Huang T C, Toraya H, Blanton T N, and Wu Y (1993) X-ray-powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *Journal of Applied Crystallography* 26: 180-184.

Huby R, and Harding J J (1988) Non-enzymic glycosylation (glycation) of lens proteins by galactose and protection by aspirin and reduced glutathione. *Experimental Eye Research* 47: 53-59.

Hulmes D J, Miller A, White S W, and Brodsky Doyle B (1977) Interpretation of the meridional X-ray diffraction pattern from collagen fibres in terms of the known amino acid sequence. *Journal of Molecular Biology* 110: 643-666.

Hulmes D J S, Jesior J C, Miller A, Berthetcolominas C, and Wolff C (1981) Electron-microscopy shows periodic structure in collagen fibril cross-sections. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 78: 3567-3571.

Hulmes D J S, Miller A, and White S W (1980) Interpretation of the low-angle meridional neutron diffraction patterns from collagen fibres in terms of the amino acid sequence. *International Journal of Biological Macromolecules* 2: 338-346.

Humphries S M, Lu Y, Canty E G, and Kadler K E (2008) Active negative control of collagen fibrillogenesis *in vivo*: Intracellular cleavage of the type I procollagen propeptides in tendon fibroblasts without intracellular fibrils. *Journal of Biological Chemistry* 283: 12129-12135.

Iijima K, Murata M, Takahara H, Irie S, and Fujimoto D (2000) Identification of N(ω)-carboxymethylarginine as a novel acid-labile advanced glycation end product in collagen. *Biochemical Journal* 347: 23-27.

Iozzo R V, and Murdoch A D (1996) Proteoglycans of the extracellular environment: Clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB Journal* 10: 598-614.

Jakas A, Katić A, Bionda N, and Horvat S (2008) Glycation of a lysine-containing tetrapeptide by D-glucose and D-fructose-influence of different reaction conditions on the formation of Amadori/Heyns products. *Carbohydrate Research* 343: 2475-2480.

James V J, Delbridge L, McLennan S V, and Yue D K (1991) Use of X-ray diffraction in study of human diabetic and aging collagen. *Diabetes* 40: 391-394.

Jiang H, Ramunno-Johnson D, Song C, Amirbekian B, Kohmura Y, Nishino Y, Takahashi Y et al. (2008) Nanoscale imaging of mineral crystals inside biological composite materials using X-ray diffraction microscopy. *Physical Review Letters* 100.

Jimenez S A, and Bashey R I (1978) Solubilization of bovine heart-valve collagen. *Biochemical Journal* 173: 337-340.

Johansen M B, Kiemer L, and Brunak S (2006) Analysis and prediction of mammalian protein glycation. *Glycobiology* 16: 844-853.

Kadler K E, Baldock C, Bella J, and Boot-Handford R P (2007) Collagens at a glance. *Journal of Cell Science* 120: 1955-1958.

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

Kadler K E, Holmes D F, Trotter J A, and Chapman J A (1996) Collagen fibril formation. *Biochemical Journal* 316: 1-11.

Kannus P (2000) Structure of the tendon connective tissue. *Scandinavian Journal of Medicine and Science in Sports* 10: 312-320.

Kavitha O, and Thampan R V (2008) Factors influencing collagen biosynthesis. *Journal of Cellular Biochemistry* 104: 1150-1160.

Kawasaki Y, Fujii J, Miyazawa N, Hoshi A, Okado A, Tano Y, and Taniguchi N (1998) Specific detections of the early process of the glycation reaction by fructose and glucose in diabetic rat lens. *FEBS Letters* 441: 116-120.

Kitamura A, Matsui K, Konoki K, Matsumori N, Murata M, Kawakami T, and Aimoto S (2011) Lysine proximity significantly affects glycation of lysine-containing collagen model peptides. *Bioorganic and Medicinal Chemistry* 19: 2125-2129.

Kostyuk O, Nalovina O, Mubard T M, Regini J W, Meek K M, Quantock A J, Elliott G F et al. (2002) Transparency of the bovine corneal stroma at physiological hydration and its dependence on concentration of the ambient anion. *Journal of Physiology* 543: 633-642.

Kumar P A, Reddy P Y, Srinivas P N B S, and Reddy G B (2009) Delay of diabetic cataract in rats by the antiglycating potential of cumin through modulation of α-crystallin chaperone activity. *Journal of Nutritional Biochemistry* 20: 553-562.

Kuznetsova N, Chi S L, and Leikin S (1998) Sugars and polyols inhibit fibrillogenesis of type I collagen by disrupting hydrogen-bonded water bridges between the helices. *Biochemistry* 37: 11888-11895.

Lapolla A, Fedele D, Aronica R, Baldo L, D'Alpaos M, Seraglia R, and Traldi P (1996) The *in vitro* glycation of lysozyme and the influence of buffer concentration investigated by mass spectrometry. *Rapid Communications in Mass Spectrometry* 10: 1512-1518.

Lapolla A, Fedele D, Garbeglio M, Martano L, Tonani R, Seraglia R, Favretto D et al. (2000) Matrix-assisted laser desorption/ionization mass spectrometry, enzymatic digestion, and molecular modeling in the study of nonenzymatic glycation of IgG. *Journal of the American Society for Mass Spectrometry* 11: 153-159.

Laroque D, Inisan C, Berger C, Vouland E, Dufosse L, and Guerard F (2008) Kinetic study on the Maillard reaction. Consideration of sugar reactivity. *Food Chemistry* 111: 1032-1042.

Le Pape A, Guitton J D, and Muh J P (1984) DISTRIBUTION OF NON-Enzymatically bound glucose in *in vivo* and *in vitro* glycosylated type I collagen molecules. *FEBS Letters* 170: 23-27.

Ledesma-Osuna A I, Ramos-Clamont G, and Vazquez-Moreno L (2008) Characterization of bovine serum albumin glycated with glucose, galactose and lactose. *Acta Biochimica Polonica* 55: 491-497.

Li Y, Asadi A, Monroe M R, and Douglas E P (2009) pH effects on collagen fibrillogenesis *in vitro*: Electrostatic interactions and phosphate binding. *Materials Science and Engineering C* 29: 1643-1649.

Li Y, Fessel G, Georgiadis M, and Snedeker J G (2013) Advanced glycation end-products diminish tendon collagen fibre sliding. *Matrix Biology* 32: 169-177.

Liu X, Wu H, Byrne M, Krane S, and Jaenisch R (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proceedings of the National Academy of Sciences of the United States of America* 94: 1852-1856.

Lyon M, and Phelps C F (1981) A rapid purification of bovine testicular hyaluronidase by chromatography on dermatan sulphate-substituted 1,6-diaminohexane-Sepharose 4B. *Biochemical Journal* 199: 419-426.

MacRae T H (2000) Structure and function of small heat shock/α-crystallin proteins: Established concepts and emerging ideas. *Cellular and Molecular Life Sciences* 57: 899-913.

Madinaveitia J, and Stacey M (1944) Substrates for hyaluronidase. *Biochemical Journal* 38: 413-417.

Maillard L C (1912) Réaction générale des acides aminés sur les sucres. *Journal de Physiologie* 14: 813.

Malone J P, George A, and Veis A (2004) Type I collagen N-telopeptides adopt an ordered structure when docked to their helix receptor during fibrillogenesis. *Proteins: Structure, Function and Genetics* 54: 206-215.

Malone J P, and Veis A (2004) Heterotrimeric type I collagen C-telopeptide conformation as docked to its helix receptor. *Biochemistry* 43: 15358-15366.

Matsumoto S I, Katoh M, Saito S, Watanabe T, and Masuho Y (1997) Identification of soluble type of membrane-type matrix metalloproteinase-3 formed by alternatively spliced mRNA. *Biochimica et Biophysica Acta - Gene Structure and Expression* 1354: 159-170.

McBride Jr D J, Choe V, Shapiro J R, and Brodsky B (1997) Altered collagen structure in mouse tail tendon lacking the α2(I) chain. *Journal of Molecular Biology* 270: 275-284.

Mechanic G L, Katz E P, Henmi M, Noyes C, and Yamauchi M (1987) Locus of a histidine-based, stable trifunctional, helix to helix collagen cross-link: Stereospecific collagen structure of type I skin fibrils. *Biochemistry* 26: 3500-3509.

Meek K M, Chapman J A, and Hardcastle R A (1979) The staining pattern of collagen fibrils. Improved correlation with sequence data. *Journal of Biological Chemistry* 254: 10710-10714.

Meek K M, Elliott G F, Sayers Z, Whitburn S B, and Koch M H (1981) Interpretation of the meridional X-ray diffraction pattern from collagen fibrils in corneal stroma. *Journal of Molecular Biology* 149: 477-488.

Meek K M, and Fullwood N J (2001) Corneal and scleral collagens - A microscopist's perspective. *Micron* 32: 261-272.

Meek K M, and Holmes D F (1983) Interpretation of the electron microscopical appearance of collagen fibrils from corneal stroma. *International Journal of Biological Macromolecules* 5: 17-25.

Meek K M, and Leonard D W (1993) Ultrastructure of the corneal stroma: A comparative study. *Biophysical Journal* 64: 273-280.

Meek K M, and Quantock A J (2001) The use of X-ray scattering techniques to determine corneal ultrastructure. *Progress in Retinal and Eye Research* 20: 95-137.

Meli M, Granouillet R, Reynaud E, Chamson A, Frey J, and Perier C (2003) Changes in glycation of fibrous type I collagen during long-term *in vitro* incubation with glucose. *Journal of Protein Chemistry* 22: 521-525.

Mikulíková K, Eckhardt A, Pataridis S, and Mikšík I (2007) Study of posttranslational non-enzymatic modifications of collagen using capillary electrophoresis/mass spectrometry and high performance liquid chromatography/mass spectrometry. *Journal of Chromatography A* 1155: 125-133.

Miles C A, Avery N C, Rodin V V, and Bailey A J (2005) The increase in denaturation temperature following cross-linking of collagen is caused by dehydration of the fibres. *Journal of Molecular Biology* 346: 551-556.

Miller A, and Tocchetti D (1981) Calculated X-ray diffraction pattern from a quasi-hexagonal model for the molecular arrangement in collagen. *International Journal of Biological Macromolecules* 3: 9-18.

Miller A, and Wray J S (1971) Molecular packing in collagen. *Nature* 230: 437-439.

Miller E J, and Matukas V J (1969) Chick cartilage collagen: A new type of α-1 chain not present in bone or skin of the species. *Proceedings of the National Academy of Sciences of the United States of America* 64: 1264-1268.

Moncton D E (2002) Toward a fourth-generation X-ray source. *accelconf.web.cern.ch/accelconf/l98/PAPERS/FR2002.PDF* [Available online] pp. 1048-1051. Last accessed: September 2013.

Monnier V M (1990) Nonenzymatic glycosylation, the Maillard reaction and the aging process. *Journals of Gerontology* 45: B105-B111.

Monnier V M (2006) The fructosamine 3-kinase knockout mouse: a tool for testing the glycation hypothesis of intracellular protein damage in diabetes and aging. *The Biochemical journal.* 399: e11-13.

Monnier V M, Bautista O, Kenny D, Sell D R, Fogarty J, Dahms W, Cleary P A et al. (1999) Skin collagen glycation, glycoxidation, and crosslinking are lower in subjects with long-term intensive versus conventional therapy of type 1 diabetes: Relevance of glycated collagen products versus HbA(1c) as markers of diabetic complications. *Diabetes* 48: 870-880.

Mosler E, Folkhard W, and Knorzer E (1985) Stress-induced molecular rearrangement in tendon collagen. *Journal of Molecular Biology* 182: 589-596.

Myllyharju J (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biology* 22: 15-24.

Myllyharju J, and Kivirikko K I (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics* 20: 33-43.

Myllyla R, Majamaa K, and Gunzler V (1984) Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *Journal of Biological Chemistry* 259: 5403-5405.

Myllyla R, Wang C, Heikkinen J, Juffer A, Lampela O, Risteli M, Ruotsalainen H et al. (2007) Expanding the lysyl hydroxylase toolbox: New insights into the localization and activities of lysyl hydroxylase 3 (LH3). *Journal of Cellular Physiology* 212: 323-329.

Nemet I, Strauch C M, and Monnier V M (2011) Favored and disfavored pathways of protein crosslinking by glucose: Glucose lysine dimer (GLUCOLD) and crossline versus glucosepane. *Amino Acids* 40: 167-181.

Nozaki Y, and Tanford C (1971) Solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions: Establishment of a hydrophobicity scale. *Journal of Biological Chemistry* 246: 2211-&.

Odani H, Shinzato T, Usami J, Matsumoto Y, Frye E B, Baynes J W, and Maeda K (1998) Imidazolium crosslinks derived from reaction of lysine with glyoxal and methylglyoxal are increased in serum proteins of uremic patients: Evidence for increased oxidative stress in uremia. *FEBS Letters* 427: 381-385.

Oimomi M, Maeda Y, Hata F, Kitamura Y, Matsumoto S, Baba S, Iga T et al. (1988) Glycation of cataractous lens in non-diabetic senile subjects and in diabetic patients. *Experimental Eye Research* 46: 415-420.

Orgel J P, Wess T J, and Miller A (2000) The *in situ* conformation and axial location of the intermolecular cross- linked non-helical telopeptides of type I collagen. *Structure* 8: 137-142.

Orgel J P R O (2000) *The Molecular Structure of Collagen.* University of stirling. PhD thesis.

Orgel J P R O, Irving T C, Miller A, and Wess T J (2006) Microfibrillar structure of type I collagen *in situ*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9001-9005.

Orgel J P R O, Miller A, Irving T C, Fischetti R F, Hammersley A P, and Wess T J (2001) The *in situ* supermolecular structure of type I collagen. *Structure* 9: 1061-1069.

Ortolani F, Giordano M, and Marchini M (2000) A model for type II collagen fibrils: Distinctive *D*-band patterns in native and reconstituted fibrils compared with sequence data for helix and telopeptide domains. *Biopolymers* 54: 448-463.

Pageon H, and Asselineau D (2005) An *in vitro* approach to the chronological aging of skin by glycation of the collagen: The biological effect of glycation on the reconstructed skin model *Annals of the New York Academy of Sciences* 1043: 529-532.

Pageon H, Bakala H, Monnier V M, and Asselineau D (2007) Collagen glycation triggers the formation of aged skin *in vitro*. *European Journal of Dermatology* 17: 12-20.

Paul R G, Avery N C, Slatter D A, Sims T J, and Bailey A J (1998) Isolation and characterization of advanced glycation end products derived from the *in vitro* reaction of ribose and collagen. *Biochemical Journal* 330: 1241-1248.

Paul R G, and Bailey A J (1996) Glycation of collagen: The basis of its central role in the late complications of ageing and diabetes. *International Journal of Biochemistry and Cell Biology* 28: 1297-1310.

Perumal S, Antipova O, and Orgel J P R O (2008) Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* 105: 2824-2829.

Perutz M F, Rossmann M G, Cullis A F, Muirhead H, Will G, and North A C T (1960) Structure of hæmoglobin: A three-dimensional fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* 185: 416-422.

Pins G D, Christiansen D L, Patel R, and Silver F H (1997) Self-assembly of collagen fibres. Influence of fibrillar alignment and decorin on mechanical properties. *Biophysical Journal* 73: 2164-2172.

Pirttiniemi P, Kantomaa T, Salo L, and Tuominen M (1996) Effect of reduced articular function on deposition of type I and type II collagens in the mandibular condylar cartilage of the rat. *Archives of Oral Biology* 41: 127-131.

Ramshaw J A M, Shah N K, and Brodsky B (1998) Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *Journal of Structural Biology* 122: 86-91.

Rasband W (2011) ImageJ. *http://imagej.nih.gov/ij* [Online]. National Institutes of Health, USA. Last accessed: September 2013.

Reddy G K (2004) Cross-linking in collagen by nonenzymatic glycation increases the matrix stiffness in rabbit Achilles tendon. *Experimental Diabesity Research* 5: 143-153.

Reddy G K, Stehno-Bittel L, and Enwemeka C S (2002) Glycation-induced matrix stability in the rabbit achilles tendon. *Archives of Biochemistry and Biophysics* 399: 174-180.

Rees S G, Dent C M, and Caterson B (2009) Metabolism of proteoglycans in tendon. *Scandinavian Journal of Medicine and Science in Sports* 19: 470-478.

Regini J W, Elliott G F, and Hodson S A (2004) The ordering of corneal collagen fibrils with increasing ionic strength. *Journal of Molecular Biology* 336: 179-186.

Reigle K L, Di Lullo G, Turner K R, Last J A, Chervoneva I, Birk D E, Funderburgh J L et al. (2008) Non-enzymatic glycation of type I collagen diminishes collagen-proteoglycan binding and weakens cell adhesion. *Journal of Cellular Biochemistry* 104: 1684-1698.

Reiser K, McCormick R J, and Rucker R B (1992a) Enzymatic and nonenzymatic cross-linking of collagen and elastin. *FASEB Journal* 6: 2439-2449.

Reiser K M, Amigable M A, and Last J A (1992b) Nonenzymatic glycation of type I collagen. The effects of aging on preferential glycation sites. *Journal of Biological Chemistry* 267: 24207-24216.

Roberts T J, Marsh R L, Weyand P G, and Taylor C R (1997) Muscular force in running turkeys: The economy of minimizing work. *Science* 275: 1113-1115.

Robins S P, Shimokomaki M, and Bailey A J (1973) The chemistry of the collagen cross-links. Age-related changes in the reducible components of intact bovine collagen fibres. *Biochemical Journal* 131: 771-780.

Rodda N (2000) *Protein cross-linking in the Maillard reaction.* The Open University. PhD thesis.

Ronziere M C, Herbage B, Herbage D, and Bernengo J C (1998) Fourier analysis of electron micrographs of positively stained collagen fibrils: Application to type I and II collagen typing. *International Journal of Biological Macromolecules* 23: 207-213.

Sasaki N, and Odajima S (1996a) Elongation mechanism of collagen fibrils and force-strain relations of tendon at each level of structural hierarchy. *Journal of Biomechanics* 29: 1131-1136.

Sasaki N, and Odajima S (1996b) Stress-strain curve and Young's modulus of a collagen molecule as determined by the X-ray diffraction technique. *Journal of Biomechanics* 29: 655-658.

Schnider S L, and Kohn R R (1980) Glucosylation of human collagen in aging and diabetes mellitus. *Journal of Clinical Investigation* 66: 1179-1181.

Schonherr E, Witschprehm P, Harrach B, Robenek H, Rauterberg J, and Kresse H (1995) Interaction of biglycan with type I collagen. *Journal of Biological Chemistry* 270: 2776-2783.

Scott J E (1996) Proteodermatan and proteokeratan sulphate (decorin, lumican/fibromodulin) proteins are horseshoe shaped. Implications for their interactions with collagen. *Biochemistry* 35: 8795-8799.

Scott J E, Orford C R, and Hughes E W (1981) Proteoglycan-collagen arrangements in developing rat tail tendon. An electron-microscopical and biochemical investigation. *Biochemical Journal* 195: 573-581.

Screen H R C, Bader D L, Lee D A, and Shelton J C (2004) Local strain measurement within tendon. *Strain* 40: 157-163.

Sell D R, Biemel K M, Reihl O, Lederer M O, Strauch C M, and Monnier V M (2005) Glucosepane is a major protein cross-link of the senescent human extracellular matrix: Relationship with diabetes. *Journal of Biological Chemistry* 280: 12310-12315.

Sell D R, and Monnier V M (1989) Structure elucidation of a senescence cross-link from human extracellular matrix. Implication of pentoses in the aging process. *Journal of Biological Chemistry* 264: 21597-21602.

Sell D R, and Monnier V M (2004) Conversion of arginine into ornithine by advanced glycation in senescent human collagen and lens crystallins. *Journal of Biological Chemistry* 279: 54173-54184.

Sell D R, Nemet I, and Monnier V M (2010) Partial characterization of the molecular nature of collagen-linked fluorescence: Role of diabetes and end-stage renal disease. *Archives of Biochemistry and Biophysics* 493: 192-206.

Shapiro R, McManus M J, Zalut C, and Bunn H F (1980) Sites of non-enzymatic glycosylation of human hæmoglobin A. *Journal of Biological Chemistry* 255: 3120-3127.

Shoulders M D, and Raines R T (2009) Collagen structure and stability. *Annual Review of Biochemistry* 78: 929-958.

Silver F H, and Birk D E (1984) Molecular structure of collagen in solution: Comparison of types I, II, III and V. *International Journal of Biological Macromolecules* 6: 125-132.

Silver F H, Freeman J W, and Bradica G (2006) Structure and Function of Ligaments, Tendons, and Joint Capsule. In: Walsh W R [ed.] *Repair and Regeneration of Ligaments, Tendons, and Joint Capsule*
Totowa, New Jersey: Humana Press, pp. 15-47.

Silver F H, Freeman J W, Horvath I, and Landis W J (2001) Molecular basis for elastic energy storage in mineralized tendon. *Biomacromolecules* 2: 750-756.

Silver F H, Freeman J W, and Seehra G P (2003) Collagen self-assembly and the development of tendon mechanical properties. *Journal of Biomechanics* 36: 1529-1553.

Silver F H, Horvarth I, and Foran D J (2002) Mechanical implications of the domain structure of fibre-forming collagens: Comparison of the molecular and fibrillar flexibilities of the α1-chains found in types I-III collagen. *Journal of Theoretical Biology* 216: 243-254.

Singh R, Barden A, Mori T, and Beilin L (2001) Advanced glycation end-products: A review. *Diabetologia* 44: 129-146.

Slatter D A, Avery N C, and Bailey A J (2004) Identification of a new cross-link and unique histidine adduct from bovine serum albumin incubated with malondialdehyde. *Journal of Biological Chemistry* 279: 61-69.

Slatter D A, Avery N C, and Bailey A J (2008) Collagen in its fibrillar state is protected from glycation. *International Journal of Biochemistry and Cell Biology* 40: 2253-2263.

Starborg T, Lu Y, Huffman A, Holmes D F, and Kadler K E (2009) Electron microscope 3D reconstruction of branched collagen fibrils *in vivo. Scandinavian Journal of Medicine and Science in Sports* 19: 547-552.

Stryer L (1981) Connective-Tissue Proteins.*Biochemistry*. (Second edn.) San Francisco: W. H. Freeman and Company, pp. 184-204.

Suarez G, Rajaram R, Oronsky A L, and Gawinowicz M A (1989) Nonenzymatic glycation of bovine serum albumin by fructose (fructation). Comparison with the Maillard reaction initiated by glucose. *Journal of Biological Chemistry* 264: 3674-3679.

Svensson R B, Mulder H, Kovanen V, and Magnusson S P (2013) Fracture mechanics of collagen fibrils: Influence of natural cross-links. *Biophysical Journal* 104: 2476-2484.

Sweeney S M, Orgel J P, Fertala A, McAuliffe J D, Turner K R, Di Lullo G A, Chen S et al. (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *Journal of Biological Chemistry* 283: 21187-21197.

Syrový I (1994) Glycation of albumin: Reaction with glucose, fructose, galactose, ribose or glyceraldehyde measured using four methods. *Journal of Biochemical and Biophysical Methods* 28: 115-121.

Tanaka S, Avigad G, Brodsky B, and Eikenberry E F (1988a) Glycation induces expansion of the molecular packing of collagen. *Journal of Molecular Biology* 203: 495-505.

Tanaka S, Avigad G, Eikenberry E F, and Brodsky B (1988b) Isolation and partial characterization of collagen chains dimerized by sugar-derived cross-links. *Journal of Biological Chemistry* 263: 17650-17657.

Thornalley P J, Langborg A, and Minhas H S (1999) Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochemical Journal* 344: 109-116.

Tsomides T J, and Eisen H N (1993) Stoichiometric labeling of peptides by iodination on tyrosyl or histidyl residues. *Analytical Biochemistry* 210: 129-135.

Tzaphlidou M, Chapman J A, and Alsamman M H (1982a) A study of positive staining for electron-microscopy using collagen as a model system. 2. Staining by uranyl ions. *Micron* 13: 133-145.

Tzaphlidou M, Chapman J A, and Meek K M (1982b) A study of positive staining for electron-microscopy using collagen as a model system. 1. Staining by phosphotungstate and tungstate ions. *Micron* 13: 119-131.

Tzaphlidou M, and Hardcastle R A (1984) A computer method for the comparison or the quantitative similarity between normal and abnormal collagen or between sequence data and experimental data. *International Journal of Bio-Medical Computing* 15: 113-120.

Ulrich P, and Cerami A (2001) Protein glycation, diabetes, and aging. *Recent Progress in Hormone Research* 56: 1-21.

UniProt (2012) *www.uniprot.org* [Online] UniProt Consortium. Last accessed: September 2013.

Usha R, Raman S S, Subramanian V, and Ramasami T (2006) Role of polyols (erythritol, xylitol and sorbitol) on the structural stabilization of collagen. *Chemical Physics Letters* 430: 391-396.

Valentine R C, and Horne R W (1962) An assessment of negative staining techniques for revealing ultrastructure. In: Harris R J C [ed.] *The Interpretation of Ultrastructure*. New York & London: Academic Press Inc., pp. 263-278.

Van der Rest M, and Garrone R (1991) Collagen family of proteins. *FASEB Journal* 5: 2814-2823.

Veit G, Kobbe B, Keene D R, Paulsson M, Koch M, and Wagener R (2006) Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *Journal of Biological Chemistry* 281: 3494-3504.

Venkatraman J, Aggarwal K, and Balaram P (2001) Helical peptide models for protein glycation: Proximity effects in catalysis of the Amadori rearrangement. *Chemistry and Biology* 8: 611-625.

Verzijl N, DeGroot J, Thorpe S R, Bank R A, Shaw J N, Lyons T J, Bijlsma J W J et al. (2000) Effect of collagen turnover on the accumulation of advanced glycation end products. *Journal of Biological Chemistry* 275: 39027-39031.

Vitagliano L, Nemethy G, Zagari A, and Scheraga H A (1995) Structure of the type I collagen molecule based on conformational energy computations: The triple-stranded helix and the N-terminal telopeptide. *Journal of Molecular Biology* 247: 69-80.

Vlassara H, Striker L J, Teichberg S, Fuh H, Li Y M, and Steffes M (1994) Advanced glycation end products induce glomerular sclerosis and albuminuria in normal rats. *Proceedings of the National Academy of Sciences of the United States of America* 91: 11704-11708.

Vogel K G (2004) What happens when tendons bend and twist? Proteoglycans. *Journal of Musculoskeletal Neuronal Interactions* 4: 202-203.

Vrdoljak A, Trescec A, Benko B, Hecimovic D, and Simic M (2004) *In vitro* glycation of human immunoglobulin G. *Clinica Chimica Acta* 345: 105-111.

Watkins N G, Neglia-Fisher C I, Dyer D G, Thorpe S R, and Baynes J W (1987) Effect of phosphate on the kinetics and specificity of glycation of protein. *Journal of Biological Chemistry* 262: 7207-7212.

Wenstrup R J, Florer J B, Brunskill E W, Bell S M, Chervoneva I, and Birk D E (2004) Type V collagen controls the initiation of collagen fibril assembly. *Journal of Biological Chemistry* 279: 53331-53337.

Wenstrup R J, Florer J B, Davidson J M, Phillips C L, Pfeiffer B J, Menezes D W, Chervoneva I et al. (2006) Murine model of the Ehlers-Danlos syndrome: col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. *Journal of Biological Chemistry* 281: 12888-12895.

Wess T J, Hammersley A, Wess L, and Miller A (1995) Type I collagen packing. Conformation of the triclinic unit cell. *Journal of Molecular Biology* 248: 487-493.

Wess T J, Hammersley A P, Wess L, and Miller A (1998a) A consensus model for molecular packing of type I collagen. *Journal of Structural Biology* 122: 92-100.

Wess T J, Hammersley A P, Wess L, and Miller A (1998b) Molecular packing of type I collagen in tendon. *Journal of Molecular Biology* 275: 255-267.

Wess T J, Miller A, and Bradshaw J P (1990) Cross-linkage sites in type I collagen fibrils studied by neutron diffraction. *Journal of Molecular Biology* 213: 1-5.

Wess T J, Wess L, Miller A, Lindsay R M, and Baird J D (1993) The *in vivo* glycation of diabetic tendon collagen studied by neutron diffraction. *Journal of Molecular Biology* 230: 1297-1303.

Willett T, Labow R, Avery N, and Lee J (2007) Increased proteolysis of collagen in an *in vitro* tensile overload tendon model. *Annals of Biomedical Engineering* 35: 1961-1972.

Willett T L, Labow R S, Aldous I G, Avery N C, and Lee J M (2010) Changes in collagen with aging maintain molecular stability after overload: Evidence from an *in vitro* tendon model. *Journal of Biomechanical Engineering* 132.

www.sigmaaldrich.com (2013a) [Online] Sigma-Aldrich product page: MOPS. Last accessed: September 2013.

www.sigmaaldrich.com (2013b) [Online] Sigma-Aldrich product page: Protease inhibitor cocktail. Last accessed: September 2013.

Yamauchi M, London R E, Guenat C, Hashimoto F, and Mechanic G L (1987) Structure and formation of a stable histidine-based trifunctional cross-link in skin collagen. *Journal of Biological Chemistry* 262: 11428-11434.

Yoon J H, and Halper J (2005) Tendon proteoglycans: Biochemistry and function. *Journal of Musculoskeletal Neuronal Interactions* 5: 22-34.

Yu Q Y, Kandegedara A, Xu Y P, and Rorabacher D B (1997) Avoiding interferences from Good's buffers: A contiguous series of noncomplexing tertiary amine buffers covering the entire range of pH 3-11. *Analytical Biochemistry* 253: 50-56.

Zhao G H, and Chasteen N D (2006) Oxidation of Good's buffers by hydrogen peroxide. *Analytical Biochemistry* 349: 262-267.

Zilva J F, Pannall P R, and Mayne P D (1991) Carbohydrate metabolism and its interrelationships.*Clinical Chemistry in Diagnosis and Treatment*. (5 edn.) London: Edwrad Arnold (Hodder & Stoughton), pp. 200 - 230.

Zyzak D V, Richardson J M, Thorpe S R, and Baynes J W (1995) Formation of reactive intermediates from Amadori compounds under physiological conditions. *Archives of Biochemistry and Biophysics* 316: 547-554.