ASPECTS OF THE STRUCTURE AND DYNAMICS OF COLLAGEN

TIMOTHY JAMES WESS

THESIS SUBMITTED FOR THE DEGREE OF PhD

UNIVERSITY OF EDINBURGH

1989

PREFACE

The work reported in this thesis was carried out between the 1st October 1985 and the 30th September 1988, under the supervision of Professor Andrew Miller at the Department of Biochemistry, University of Edinburgh. All material presented in this thesis unless otherwise stated, is the sole work of the author, as is the composition.

ABSTRACT

This thesis describes the study of the type I collagen molecule and fibril as they occur in rat tail tendon.

The first three chapters are concerned with the background to the study. Chapter 1 is a review of connective tissue structure, occurrence and function. Chapter 2 explains the theory of X-ray and neutron diffraction by fibres, whilst chapter 3 describes the sources of X-rays and neutrons and the natures of cameras and sample cells used.

Chapter 4 presents the one dimensional structure of rat tail tendon as determined by X-ray diffraction. The method of phase determination used is isomorphous addition of heavy atoms. The chapter describes the production of suitable isomorphous derivatives, data collection, data correction and structural refinement. The position of the heavy atoms in the fibrils of tendon are discussed with respect to amino acid sequence of the collagen molecule.

Chapter 5 presents the one dimensional structure of rat tail tendon fibrils as determined by neutron diffraction. The method of phase determination employed was the isomorphous addition of deuterium. The chapter describes the possible sites of reduction by sodium

borodeuteride in tendon fibrils, the production of suitable isomorphous derivatives, data collection, data correction and structural refinememt.

The positions of deuterated sites are discussed with respect to the position of reducible crosslinks as determined by chemical methods.

Chapter 6 compares and contrasts the results of the techniques used, and their potential in further studies is discussed.

ACKNOWLEDGEMENTS

I gratefully acknowledge the financial support provided by the Science and Engineering Research Council for research during my three years at the Department of Biochemistry, University of Edinburgh.

I would like to express my gratitude to the following friends and colleagues who have helped me in the work for this thesis:

Professor A Miller and Dr J Bradshaw for their supervision, support and advice throughout the course of this work.

Dr R J Greenall, Dr L Townes, Dr W Bras and Dr C Nave for assistance with the fibre diffraction cameras at the Daresbury Synchrotron Radiation Source.

Dr P A Timmins for assistance with the neutron cameras at the Institut Laue Langevin.

Mr K C Duff and Miss L Hanlon for preparation of samples and assistance in data collection.

Mr J Burgess and Mr M McGregor for their expert workmanship and technical support.

The medical photography department for preparation of the photographs.

The entire crew of room 302 not already mentioned past and present who have provided many fruitful discussions and a great deal of entertainment.

Finally I would like to thank my parents and my girlfriend Linda for their unfailing care and attention.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF PLATES

LIST OF TABLES

TABLE OF ABBREVIATIONS

A	amplitude of incident radiation
a*, b*, c*	unit cell dimensions of reciprocal space
B	temperature factor
b...	neutron scattering length
b	lattice spacing
c	velocity of electromagnetic radiation
coh	coherent
D	displacement in reciprocal space (chap2)
D	collagen repeat period
d	displacement in real space
E	energy
e	charge of one electron
F	Fourier transform
F...	diffraction structure factor terms
f...	X-ray scattering lengths
$\|F_{hkl}\|$	Structure factor amplitude
h,k,l	Miller indicies
h	Planck's constant
I	spin of a nucleus
inc	incoherent
i	square root of -1
L	distance between origin and diffraction sampling point
O_1, O_2..	point scatterers
P	Patterson function
R	radius

r	vector joining point scatterers
s	X-ray scattering vector
S	unit vector in direction of scattered beam
S_0	unit vector in direction of incident beam
t	time
u	Patterson space vector
Vx	volume in real space
Vm	total volume of neutron scattering components
Y	displacement of electromagnetic field in wave of radiation
z(x)	crystal lattice function
$\eta..$	amplitude of scattered radiation
λ	wavelength
α_s	scattering phase shift
$\rho(x)$	electron density distribution
$\rho c(x)$	crystal electron density distribution
θ_c	critical angle
φhkl	phase angle of structure factor
υ	frequency of the radiation
$\sigma(s)$	Fourier transform of lattice function
ψn, ψs	incident and scattered neutron waveforms
σ	neutron scattering cross section
ξ	neutron scattering cross section (coherent)
δ	neutron scattering cross section (incoherent)
δ	delta function (section 2.2.2 only)

CHAPTER 1

INTRODUCTION TO CONNECTIVE TISSUES

1.1 OVERVIEW OF CONNECTIVE TISSUES

Connective tissues are essential for spatial organisation in most forms of multicellular life. In plants and insects cellulose and chitin respectively play major roles. This is not the case in animals where collagen is found to be the most abundant protein, and is the main component of connective tissues i.e. the extracellular matrix. Collagen thus occupies a key position in the molecular architecture of higher animals, guaranteeing the structural integrity of the organism by forming a scaffold to which cells or organs may attach. In various forms it has the ability to confer structural and mechanical support as well as influence differentiation of cells.

The importance of collagen amongst mammalian body proteins was shown quantitatively by Harkness et al, (1958) where it was found to constitute about one sixth of total body protein (wet weight) in mice. A more qualitative importance of collagen can be seen in reference to the diseases associated with collagen, as they often manifest themselves as biomechanical disorders of the body (Pope & Nicholls 1984).

Collagen also has a commercial value. Initial studies made on collagen were through its use as a glue, in the form of gelatin and in the process of tanning in the production of leather. Collagen is also used in the food and medical supplies industries. The general

importance of collagen has thus led to many types of investigation in varying morphological and histological situations. From these studies two general points have been found to be fundamental to an understanding of the role of collagen in the extracellular matrix.

(1) It has been established that a number of genetically distinct collagens exist, thus the term collagen applies to a family of proteins with similar characteristics.

(2) The stoichiometric relationship of collagen with other components within the extracellular matrix varies, leading to different morphology of the tissues.

The potential created by using different types of collagen in a structure, as well as varying the composition and amount of complementary material, is vast. It permits the production of a number of tissues ideally suited to serve their biological roles, that can be made from relatively few types of macromolecules.

1.2 WHAT DEFINES A COLLAGEN ?

There are a number of criteria that define a molecule as a member of the collagen family. The most important of these is the amino acid sequence, the occurrence of glycyl residues in position 1 of a repeated tripeptide -GLY-X-Y- where X and Y are variable amino & imino acids but often proline or hydroxyproline. The result of this in physical terms is the formation of the characterstic rope like structure of the collagen triple helix.

1.2.1 THE GENETICALLY DIFFERENT TYPES OF COLLAGEN

The first conclusive evidence that different types of collagen exist within a given organism was produced by Miller E.J. & Matukas (1969). On using procedures to characterise the soluble collagen from chicks, it was found that a collagen could be characterised that was different from that found previously in skin and bone. This marked the advent of discovery of a number of distinct collagens which, although containing the central characteristics attributable to the collagens, are genetically, chemically and immunologically different.

Current classification allows for eleven different collagens, and these are given Roman numerals I-XI. Types I-III are often referred to as the fibre forming collagens. The molecules tend to show similar superficial features, each being approximately 300nm long extents of collagen triple helix with non-helical domains

termed telopeptides at the N and C termini.

Type I collagen is the major component of tendon, and is to be the subject of this thesis. It was the first collagen to be identified and is thought to be the most abundant. As well as being the major constituent of tendon, it is also found in large amounts in skin, bone, cornea, annulus fibrosus and placental tissue. It is often found associated with other type I molecules in a supramolecular organisation as fibrils, which exhibits a characteristic banding pattern that can be seen in the electron microscope.

Type II collagen is the major structural protein of cartilage, it is also known to exhibit a characteristic supramolecular banding pattern, this has best been observed in the notochord sheath of the lamprey (Eikenberry et al, 1984). Type II has also been observed in annulus fibrosus, nucleus pulposus and the vitreous body of the eye.

Type III collagen occurs in various tissues notably skin, blood vessels and the submucosal layer of the colon. It produces a fibrous banded structure, but it is generally less well ordered than some structures formed by type I collagen. These first three types of collagen are most relevant to the scope of this thesis. A review of fibrous forming collagen types is Miller E.J. (1976).

The nature of types IV - X of collagen and their location is less well defined, see Table 1.1. A more recent review of collagen types including these is by

Miller E.J. & Gay (1987). The importance of these less well characterised collagens is beginning to be recognised, especially with respect to the interactions of these molecules with the fibrous forms of collagen in various tissues, for example see Birk et al, (1988).

Table 1.1 collagen type, form and location

Collagen Type	Form	Location (preferred)
I	fibrous	tendon, bone
II	fibrous	cartilage,notochord
III	fibrous	skin, embryo
IV	network	basement membrane, corneal stroma.
V	fibrous?	placenta, bone, skin
VI	microfibrillar	cartilage
VII	?	placenta
VIII	network?	Descemet's membrane cultured epithelium.
IX	fibril associated	cartilage
X	?	calcifying cartilage
XI	fibril associated	cartilage

1.2.2 COMPOSITES

The extracellular matrix does not only consist of collagens. The requirement for structural integrity demands the ability of various tissues in the body to exhibit elasticity, withstand compression and in special cases to act in cell-cell recognition and morphogenesis. With the collagens there can also be found proteoglycans, fibronectins, laminins, elastin and water. It is the proportion and arrangement of these macromolecules that lead to the differing biomechanical and physiological roles of specialised extracellular tissues. From this level of organisation, a structure function relationship can be seen.

Tendons consist mainly of fibres of collagen with small amounts of proteoglycan. Their function is to transmit force and thus they can be regarded as tissues built to withstand tension along their axis. Much of this mechanical strength can be attributed to the fibrous nature of the collagen within the tendon and its axial ordering.

Cartilage is designed to resist compression and often acts as a biological shock absorber at the end of bony structures. To fulfill this role, cartilage contains large amounts of highly hydrated proteoglycans which are able to absorb mechanical stress.

Tissues such as the aorta contain elastin, collagen and proteoglycan. These act to produce a structure that is able to be extended without rupture, in order to

accommodate the strong pulsatile flow of blood from the left ventricle of the heart in systole.

The composition of the tissue is not the only factor that confers the biomechanical properties. In many cases the diameter of collagen fibrils varies between tissues, and a range of fibre sizes can be characteristic of certain tissue types. Tendon has been found which exhibits a range of fibril diameters (Parry et al, 1978a), whereas the fibril diameter in corneal stroma is uniform and accounts for the precise packing order and thus the high transparency associated with the tissue (Hart and Farrell 1969)(Craig & Parry 1981). This particular example emphasises how the organisation of the constituent components can produce such differing structures as cornea and tendon. The structure of connective tissue is therefore far more intricate than is initially envisaged, and than that suggested by the name connective tissue.

1.3 THE COMPONENTS

Little has been said so far about the relation of the component molecules' structure to the function of connective tissue. This thesis primarily describes an investigation of the structure of collagen within one type of connective tissue, tendon, and therefore a review of type I collagen structure is most relevant and will take priority in the following section.

1.3.1 COLLAGEN

A vast amount of research has been concerned with collagen, this short survey of structure will be limited to type I collagen of rat tail tendon unless specified.

1.3.1.1 Primary structure.

The determination of the type I collagen sequence has proved to be a difficult task. In the years before the full primary structure was elucidated, a number of chemical composition features were revealed that aided an understanding of collagen chemistry.

Bowes & Kenten (1948) showed that the protein had a high glycine content in relation to globular proteins; it also contained little tyrosine, tryptophan and no cysteine. This analysis was the first to account for essentially the entire weight of the sample and its nitrogen content.

Collagen was also found to be rich in the imino acids proline and hydroxyproline the latter being a rare amino acid not often found in proteins outwith the extracellular matrix.

Piez et al, (1963) revealed type I collagen to be a heterotrimer containing two types of chain called α1 and α2 in a ratio of 2:1. An amino acid sequence of the α1 chain was collated from published data by Hulmes et al, 1973. A sequence determination of the α1 chain (mostly calf skin) was also conducted (Fietzek and Kühn 1975). A larger number of sequences have now been obtained for α1 and α2 chains in different species, the sequence of type I collagen α1 chain within an organism is the same in different tissues (Bornstein 1969). Sequence data can be found in Weiss & Jayson (1984). A greater degree of homology was found between type I α1 chains between species than that between α1 and α2 chains in the same organism.

Both the α1 and α2 chains have 3 distinct regions, each has an N terminal region that contains no internal periodicity, a core region of 1014 amino acid residues containing the repeating tripeptide unit -GLY-X-Y- as described earlier, and is capped at the C terminus by another aperiodic sequence. The length of the α1 and α2 N termini regions are 16 and 9 residues respectively and those of the C termini being 25 and 6 residues respectively, see Figure 1.1. Various features can be related from the primary structure to the conformation

of the molecule and its ability to interact with other macromolecules.

The -GLY-X-Y- motif is mandatory for formation of the triple helical structure (Rich & Crick 1955). Conformational restriction imposed by imino acids occurring frequently in this sequence, is essential for formation of a triple helical molecule as opposed to allowing a more entropically stable state to occur. The proportion of imino acids in a given collagen has been related to the thermal stability of the triple helix (Josse & Harrington 1964).

In the repeating tripeptide, the occupancy of positions X or Y is not random for certain amino acids (Fietzek & Kühn 1975) (Hofmann et al, 1980). This has been found to apply to the imino acids, and the position of amino acids in the tripeptide is also thought to affect the thermal stability of the collagen. Along the core region of the molecule the proportion of different tripeptide types to others varies.

The core region adjacent to the C terminal region has a relatively high proportion of the tripeptide -GLY-PRO-HYP-. This is believed to be a tripeptide unit that confers a great deal of stability to the triple helix. Other regions of the core contain few if any imino acids in the triplet sequences. These regions are thus thought to be less stable, (Fietzek & Kühn 1975).

The reduced transition temperature of a non hydroxylated form of collagen provides evidence of hydroxyproline playing

a role in the stabilisation of the triple helical conformation (Berg and Prockup 1973).

The distribution of polar amino acids along the α1 and α2 chains was first revealed by the electron microscope. It varies throughout the length of the molecule and is thought to be related to the ability of individual collagen molecules to assemble and form an organised fibrous system. Very few loci have been found at which a positive or negative charge predominates, however stretches of charged groups are identifiable along the axis of both types of chain. It is these that can be directly related to the staining bands found in collagen fibrils examined by electron microscopy. The polar and hydrophobic residues are also found in discrete bands in the sequence. These are postulated to contribute in the interactions between laterally adjacent collagen molecules.

Full sequence information of collagen type I also reveals it to contain relatively small amounts of methionine, histidine and tyrosine, three residues that are able to bind heavy atom stains with a high degree of specificity. Tyrosine is of special interest since it is only found in the telopeptide regions.

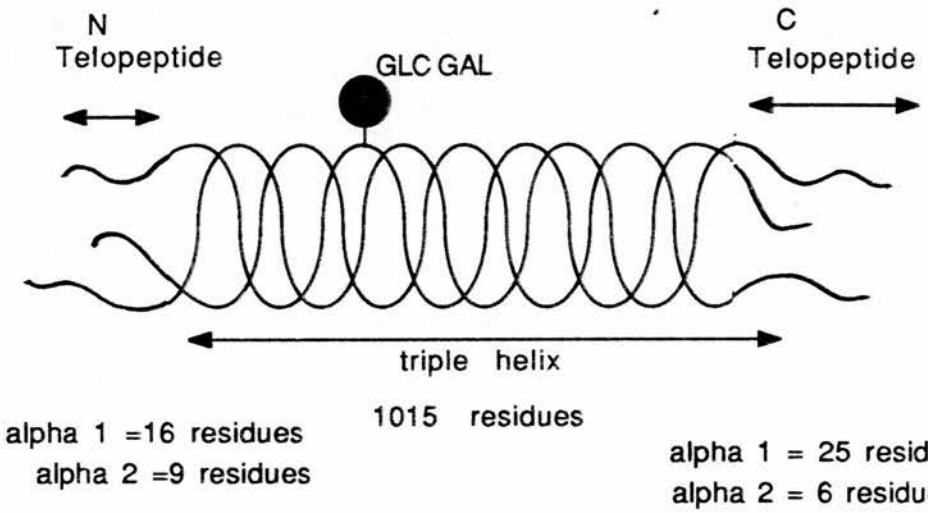
Collagen is also noted for the presence of crosslinks derived from lysine residues. They contribute to the mechanical strength of the tendon. This is easily demonstrable by the condition lathyrism in which the

crosslinks have been unable to form. This results in weak tendons.

Crosslink formation is initiated by the oxidative deamination of lysine or hydroxylysine residues in the molecule by lysyl oxidase. This produces allysine and hydroxyallysine respectively. These can spontaneously form crosslinks with a number of related species as shown in the Figure 1.2. A recent review of the crosslinking process is Eyre (1987). The exact nature of some crosslinks is uncertain, as is the position of the crosslinks. A process of maturation is thought to occur resulting in the lack of acid solubility of collagen. However the nature of this maturation process is not fully understood (Barnard et al, 1987); this is discussed in greater detail in chapter 5.

Collagen also contains specific glycosylation sites. The amount and position of glycosylation varies between tissues. Spiro (1969) investigated the nature and quantity of hydroxylysine linked carbohydrates. There is thought to be one primary glycosylation site on type I $\alpha 1$ chains of rat tail tendon. It has been suggested that the position of this bulky sidegroup plays a role in the packing of collagen molecules.

The supramolecular arrangement of the collagen may be directed by the carbohydrate moiety requiring a particular fit (Morgan et al, 1970). Non-enzymatic glycosylation has been observed in collagen (Rosenberg et al, 1979). The degree of glycosylation is thought to be related to the aging of the tendon and to the raised interstitial sugar concentrations found in diabetes.



The size of the nonhelical telopeptides is exaggerated

(GLC GAL) = O Glycosidic linked disaccharide,

Figure 1.1 the triple helical type I collagen molecule.

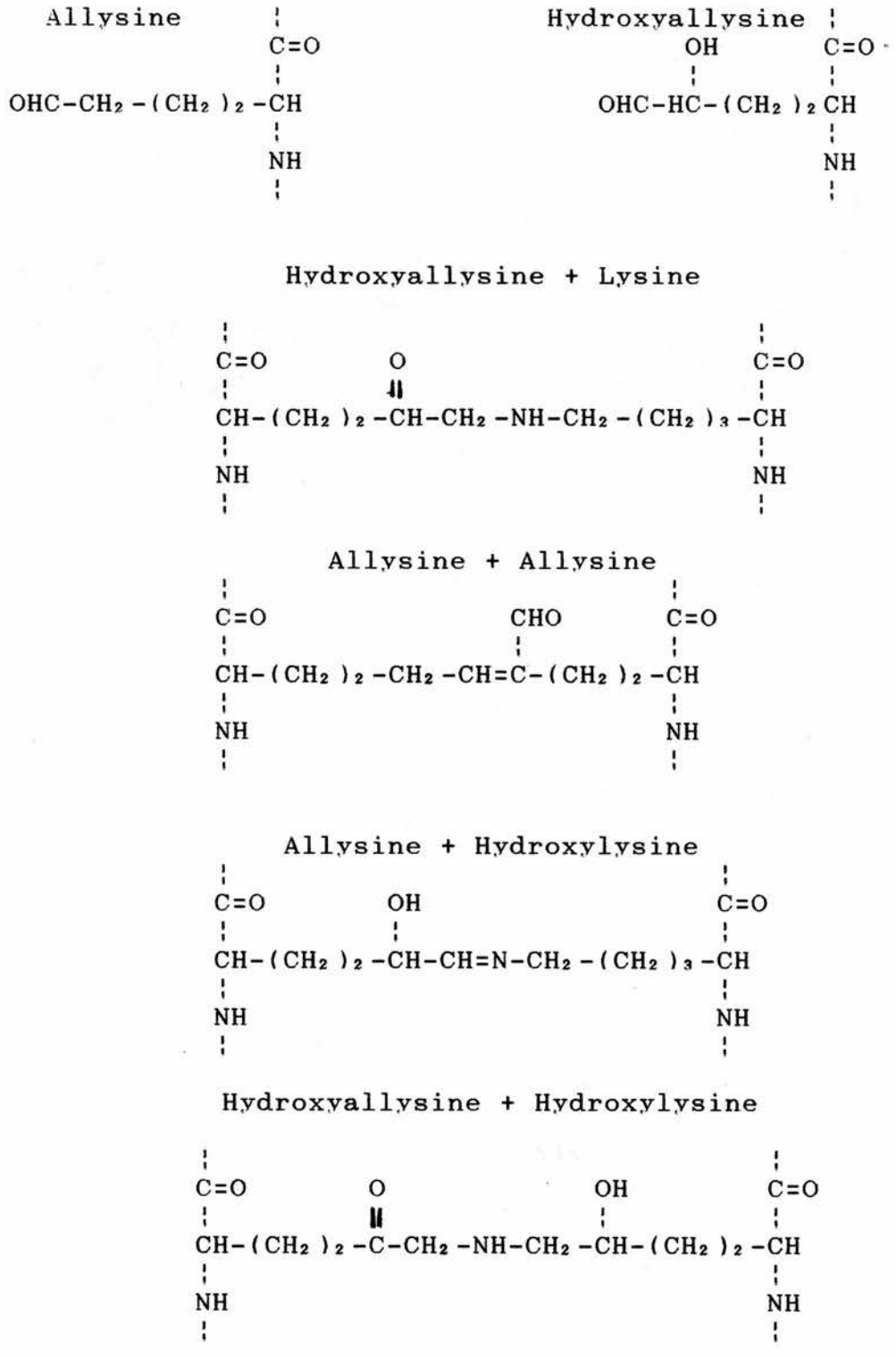


Figure 1.2 crosslinks commonly found in collagen

1.3.1.2 Physical properties of collagen

The complete collagen triplex can be described as a 285kDal protein, approximately 300nm long and 1.5nm in diameter. It is a heteropolymer containing two identical $\alpha 1$ chains and one $\alpha 2$ chain. This ratio was initially deduced by statistical evaluation of collagen composition (Piez et al, 1961). It was demonstrated to be in this ratio for individual triple helices of type I collagen, by the introduction of intramolecular crosslinks, the subsequent separation of complexes formed and their amino acid analysis (Nold et al, 1970).

Physical methods used to study the molecule showed that in solution it behaved as a rigid rod (Boedker & Doty 1956). More detailed studies have subsequently been made (Gelman and Piez 1980), suggesting that the molecule is semiflexible. A flexible site in the triple helical region of type I collagen was found by analysis of electron micrographs. This corresponds to a section of the sequence in $\alpha 1$ chains that lacks proline residues (Hofmann et al, 1984). A direct visualisation of the individual collagen molecule's shape was first obtained by Hall (1958). The occurence of type I collagen as a heteropolymer may be due to the need for specific interactions between triple helices, and also the stability of the triple helix. The stability of artificial collagen molecules has been investigated using molecules comprising of three $\alpha 1$ chains and three $\alpha 2$ chains. An $(\alpha 1)_3$ molecule has about the same thermal

denaturation properties as that of a native type I molecule. That is, the triple helical structure of the molecule is lost at about the same temperature in both cases. The $(\alpha 2)_3$ molecule is found to be markedly less stable as judged by renaturation (Tkocz & Kühn 1969).

1.3.1.3 Structure of the triple helix

The helical nature of collagen was elucidated by a combination of techniques. The unusual amino acid sequence of the helical domain, with its repeating triplet structure, had to be accounted for in any proposed molecular structure. High angle X-ray diffraction and model building also served as valuable tools in the elucidation of the structural parameters. Interpretation of high angle X-ray diffraction patterns of collagen has been conducted for over 70 years. A comprehensive list of early studies has been given by Astbury (1933).

Significant improvements were made to the quality of diffraction patterns by the sample being stretched by 10% (Cowan et al, 1953). This produced patterns of sufficient quality to permit helical parameters to be determined (see Plate 1.1).

The nature of helical diffraction is such that direct structure determination from film fibre diagrams proves impossible. Structure determination therefore results from model building, using information from the

fibre diagram and the constraints imposed by torsion angles of the polypeptide backbone. The theory of helical diffraction (Cochran et al, 1952) led to an initial indexing of major layer lines. The indexing of the fibre diagram layer lines, resulted from a proposed structure that contained 10 equivalent scattering units in three turns of a helix (Cohen and Bear 1953). This proposal however was difficult to relate to possible molecular conformations.

A significant breakthrough in the proposed structure of collagen was made by Ramachandran & Kartha (1954). They proposed that collagen was essentially a triple helical structure, each chain having a helical nature as well as the triple helix exhibiting a coiled coil conformation, see Figure 1.3. This model was able to explain the presence of glycine at every third residue being necessary to allow the correct conformation to occur. Each glycine in the triple helix is located at the same radius near the centre of the triple helix. Glycine is the only amino acid whose side chain is small enough to fit in the space allowed. In considering the conformational properties of other collagen constituents, the allowed $\phi.\psi$ torsion angles of the imino acids are more restricted than that of the other amino acids.
This acts as a limiting factor on the number of possible structures that can be modelled to fit the collagen diffraction pattern. The left handed nature of the helical turn in each α chain is thought to be due to the

presence of proline. Polypeptides resembling collagen have been made, such as homopolymers of proline. These have been shown to have a left handed helical nature, for a review see Bornstein & Traub (1979). The X-ray diffraction study of synthetic polypeptides such as poly -GLY-PRO-PRO- gave diffraction patterns similar to those of collagen (Yonath & Traub 1969). Hydroxyproline residues are thought to play a role in stabilising the triple helix by forming a system of hydrogen bonds.

Advances on the model proposed by Ramachandran were made by Rich & Crick (1955). Two structures were proposed whose main differences involved the nature of the hydrogen bonding system. These have subsequently been known as the collagen I and the collagen II structures.

Both structures contain one hydrogen bond per three residues. Of the two, collagen II was found to be the most compatible with data obtained from X-ray analysis of collagen and that of related synthetic triple helical molecules. The model has been refined by a linked atom least square procedure, using quantitative X-ray data obtained from kangaroo tail tendon (Fraser et al, 1979). They proposed the model of a collagen molecule to be a three start helix. The introduction of a rope like twist allows the molecule to assume a supercoil conformation. The hydrogen bonding systems occur between the peptide NH group of a glycine residue and the peptide C=O group of a residue in the X position of a triplet in another chain.

The amide group of a residue at position Y may form hydrogen bonds with water.

In globular proteins the role of water surrounding the molecule in a "cage" is recognised as being significant to its stability and conformational properties. This is also thought to apply to collagen. (Fraser et al, 1987).

The observed unit height of the helix was determined to be 0.2983nm i.e. the displacement of an amino acid parallel to the axis of the helix. The unit twist of a residue in the helix is 107.1°$\pm$ 0.6°. This corresponds to a helix containing ten units in three turns. These values can only be regarded as averages, the occupancy of the positions X and Y by residues other than imino acids may lead to deviations in the localised helical parameters. The pitch of the superhelix associated with the triple helix has been difficult to measure by X-ray analysis. Estimations have been made in the region of 9nm.

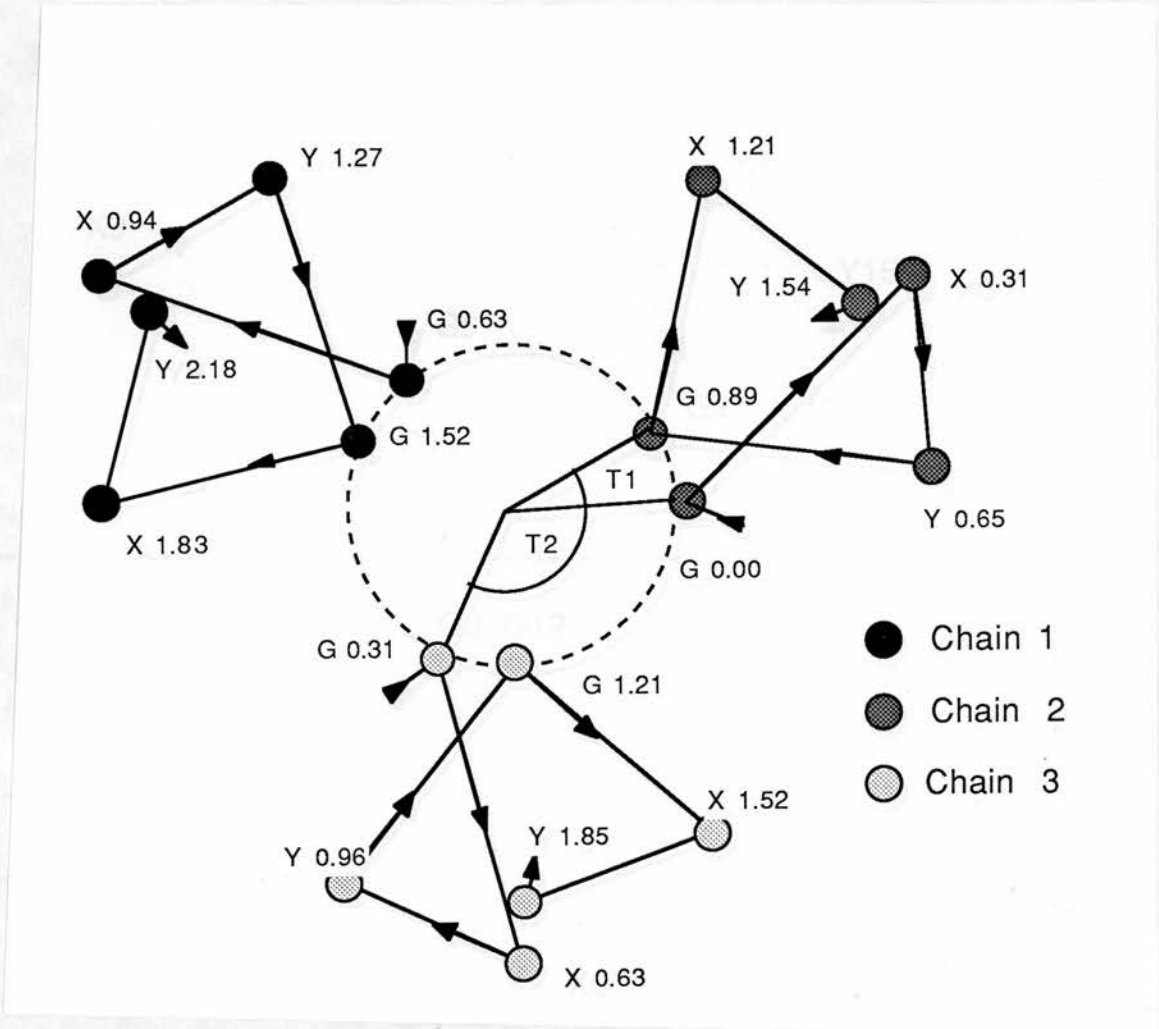


Figure 1.3 The geometry of the collagen molecule.

The circles represent the alpha carbons of each polypeptide chain, and the numbers the approximate height in nanometres. The projection is from the C terminus to the N terminus. The left handed helix of each chain is supercoiled by an angle of T1 per tripeptide. The angle T2 is therefore not equal to 120° but 120°-(T1/3)°.
The molecular helix contains 3.3 units per turn.
The coordinates used are the Rich & Crick II, one bonded structure.

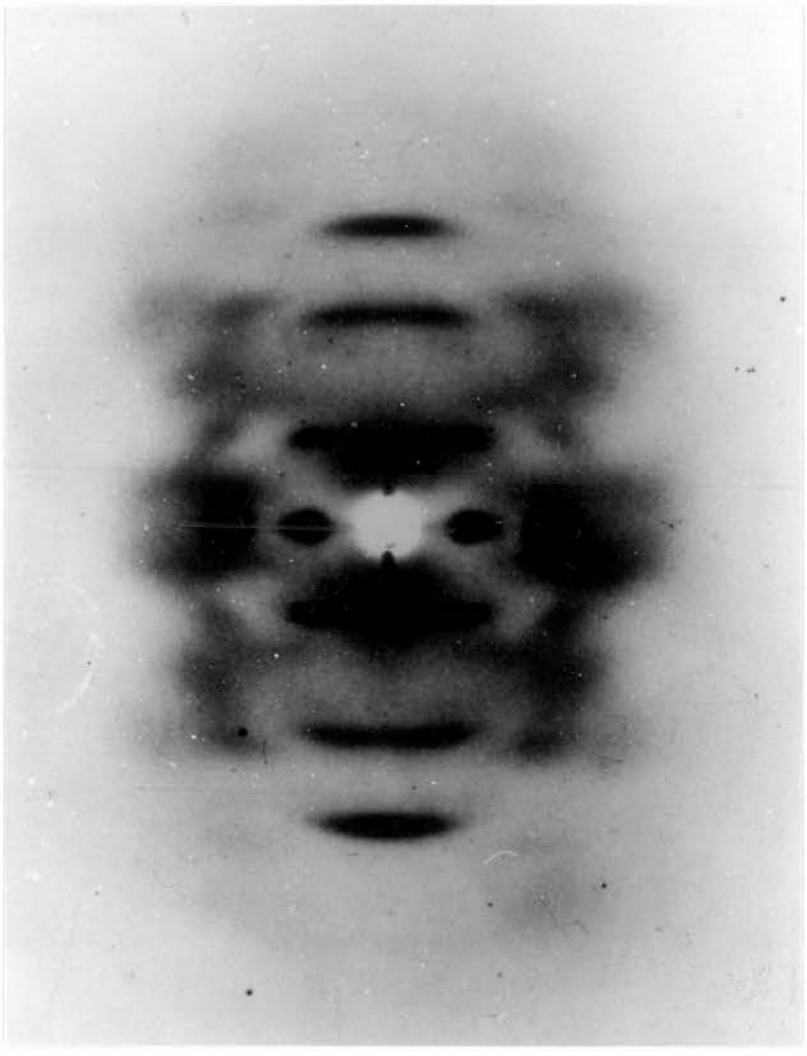


Plate 1.1 The high angle diffraction pattern of collagen (by A.C.T. North)

1.3.1.4 The Telopeptides

Only small parts of the collagen molecule do not contain the -GLY-X-Y- concatenation. These parts cannot form the distinctive triple helix and are found at the ends of both ∝1 and ∝2 chains as described earlier. These regions of the molecule have been postulated to be essential in fibril assembly (Liebovich & Weiss 1970) and are thought to be involved in crosslink formation, see for example Davison & Brennan (1982), Piez (1976).

The structure of the telopeptides is a matter of controversy. Information on telopeptide structure has been obtained from X-ray diffraction (Hulmes et al, 1977), structural prediction algorithms (Helseth & Veis 1979), (Jones & Miller A. 1987), (Dion & Myers 1987) and nuclear magnetic resonance (Otter et al, 1988). These techniques have not however produced much agreement as to the structure.

1.3.1.5 Fibrillar nature of collagen

The nature of collagen as a biomaterial makes the understanding and investigation of the interaction of this molecule with others essential. One of the most easily understood systems is that of type I collagen in tendon. Structural investigation of the fibrillar nature has been conducted mainly by the use of electron microscopy, neutron diffraction and X-ray diffraction.

These techniques have allowed the elucidation of a number of parameters concerning the native axial and lateral structure of packing in tendon.

1.3.1.6 Axial structure

The axial structure refers to the molecular structure of the fibril projected onto the fibril axis. It disregards the lateral packing of molecules in the other two directions. Historically, the first visualisation of axial fibrillar structure was obtained by Schmitt (1942) using electron microscopy.

The electron micrographs had a repeat periodicity (D) of 64nm, that was far less than that of the length of an individual collagen molecule (L). Schmitt in 1955 thought that L=4D and proposed the quarter stagger model. This suggested that the molecules are arranged parallel to the fibril axis and staggered by 64nm, or multiples of this, with respect to one another.

Negative staining of collagen fibrils revealed bands of dark staining regions in electron micrographs. The process of negative staining involves the addition of electron dense atoms in the form of a stain, which is not thoroughly removed by washing, thus the stain remains and is able to fill holes between the molecules.

A non-integral relationship between L and D was proposed that could explain the appearance of the fibril when negatively stained. This results in the axial structure exhibiting a step like function of lower and higher electron densities, respectively known as the gap and overlap regions. This is the basis of the Hodge Petruska model of fibril structure (Hodge and Petruska 1963).

They proposed that a collagen molecule in a stained and dried fibril is 4.4D long, giving rise to a 0.4D overlap and a 0.6D gap region, the predicted mass ratio being 5:4 overlap to gap, see Figure 1.4.

From this information it could be considered that the collagen molecule can be divided into 5 units, 1-5 each having length D, with the exception of 5, which has a length of approximately 0.4D. In projection onto the fibril axis, each D length contains a full complement of units 1-5, thus axially the one dimensional structure is periodic containing one gap overlap function, see Figure 1.4.

Positive staining of collagen differs from negative staining of collagen in that much of the stain is washed off, leaving only those molecules of stain that have found sites with which they can interact electrostatically (Nemetschek et al, 1955). This work revealed the aperiodic nature of staining within the D repeat and also showed the fibre to have polarity in the axial direction.

Analysis of the positive staining pattern of the fibril in conjunction with information obtained from sequence data has been conducted. The molecule is approximated by a linear chain of equally spaced side chains and it is assumed that the stain has only adhered to charged residues. The correlation between this and a predicted staining pattern is strong. The technique was first applied to a cyanogen bromide digest of collagen (von der Mark 1970), and has been conducted a number of times on the whole of the molecule e.g. Hulmes et al, (1973) and Meek et al, (1979) who conducted a more detailed study. Both found that the number of residues in a D spacing was 234.2 $\pm$ 0.5.

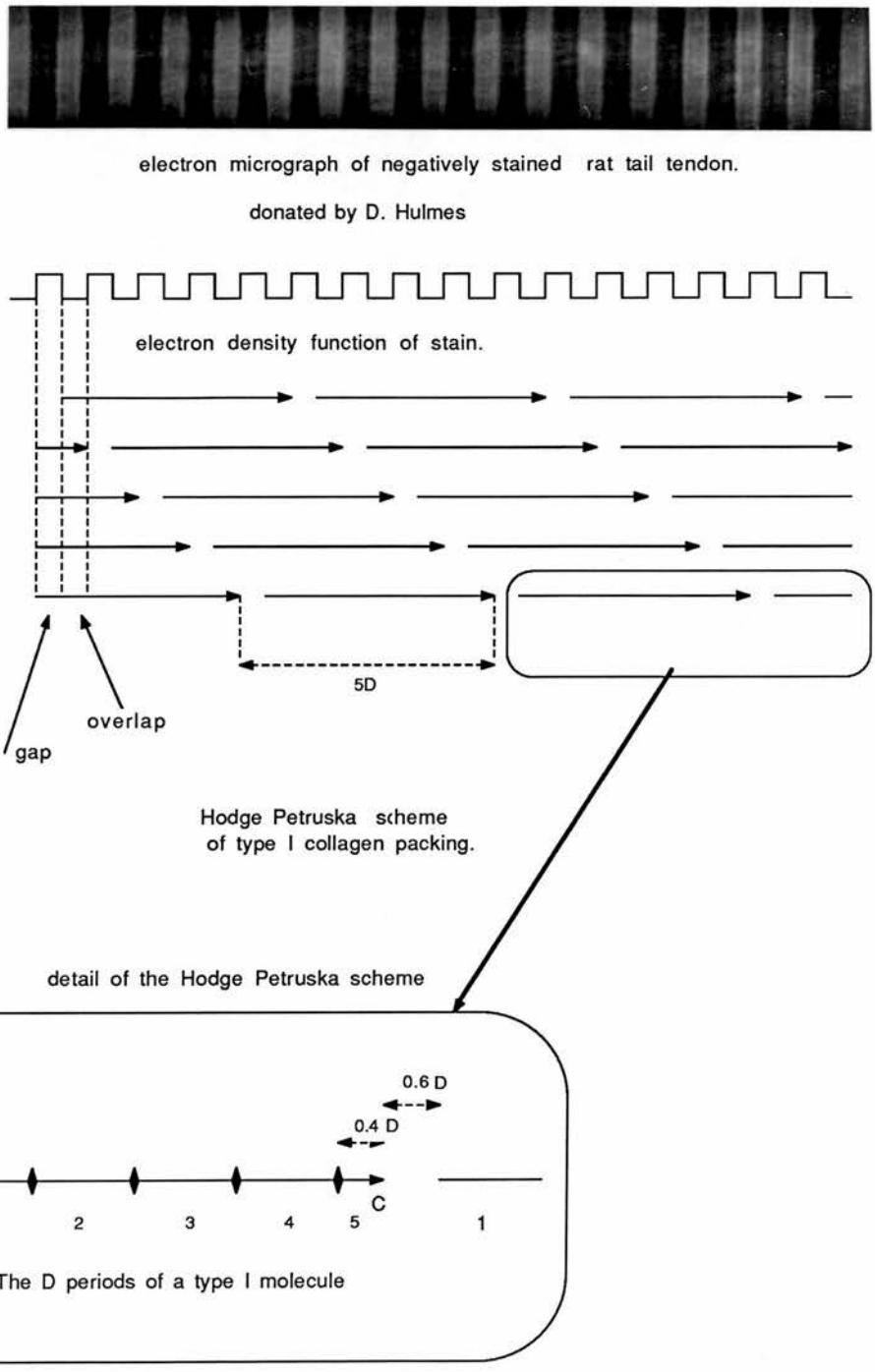


Figure 1.4 the axial structure of collagen in tendon.

The discussion so far has been limited to analysis of the electron microscopy data. However, X-ray analysis of collagen has contributed a great deal to the understanding of the axial structure, and is the main subject of this thesis. A detailed discussion of the method of X-ray analysis relevant to collagen will be given later.

Low angle X-ray analysis suggested an axial periodicity of 64nm in dried samples (Bear 1942,1944). Tomlin & Worthington (1957), using X-ray diffraction, also predicted that a gap overlap of electron density existed in native collagen fibres. The advantage of X-ray studies is that it allows the fibres to be studied in a state pertaining to that _in vivo_. This is almost impossible in electron microscopy. It also allows a more detailed investigation of the axial structure since the resolution which can be obtained using the electron microscope to study biological specimens is less than that obtained in X-ray analysis.

The "phase problem" associated with X-ray diffraction data interpretation has however hampered the ability to determine an absolute axial conformation. Hulmes _et al_, (1977,1980) used model building techniques to overcome this problem and showed that X-ray and neutron diffraction data did in fact correlate with a number of the previously determined parameters for axial collagen structure. However significant differences were found in the gap:overlap ratio and in the size of the D period since the fibre was in a wet state.

The contribution of X-ray diffraction and neutron diffraction will be studied in more detail in chapters 4 and 5.

Amino acid sequence data has been used to determine a value for D, and thus the gap overlap function, by evaluating the possible interactions between hypothetical one dimensional models as a function of stagger between them. Hulmes et al, (1973) scored two identical but staggered amino acid sequences for possible hydrophobic and electrostatic interactions. This resulted in distinct maxima in the scores of a number of interactions when the stagger was D=nx234. Thus a number of techniques have combined to support the theories of axial structure which are now generally agreed upon.

1.3.1.7 Lateral structure

Compared to the axial structure found in tendon, very little can be said with certainty about the lateral packing. This is mainly due to the limited amount of information obtained so far, which is relevant to the three dimensional structure and also our ability to interpret it.

The electron microscope reveals little information involving the three dimensional molecular structure of tendon, its resolution is insufficient to image the lateral organisation of molecules. X-ray diffraction analysis was able to demonstrate the presence of organised lateral packing in some tendons. This was first established in 1954 (North et al, 1954). From this point, the speculation of the way in which molecules packed, began in earnest.

Several theories were proposed, each having its own adherents. One of the most significant of these was the microfibril model proposed by Smith (1968). He proposed the structural unit of packing was a group of 5 molecules closely packed with helical symmetry. These were discrete rope-like structures and called microfibrils. These microfibrils were then thought to be packed together into fibrils.

The way in which each theory of packing could be tested, was to use the predicted X-ray diffraction pattern of a three dimensional model and compare it to the medium angle X-ray diffraction data (10 to 1nm resolution).

Improvements in the quality of diffraction data and the interpretation of these, led to the proposal that extents of the tendon exist as crystalline regions based on quasi hexagonal molecular packing without micro-fibrillar substructures (Hulmes & Miller A. 1979).

Advancements in structure determination were made in 1983 (Fraser et al, 1983). Here X-ray data from tendon specimens that had been stained with phosphotungstic acid were used, the diffraction peaks obtained were best accounted for by the molecular packing resulting in a triclinic unit cell.

This work also proposed that the structure of the molecular segments in the gap region was less rigid than that of the overlap region. Further theoretical interpretation of data resulted in proposals of three dimensional packing for the gap and overlap regions (Fraser et al, 1987).

It must be stressed that the observation of X-ray data reveals that the diffraction peaks corresponding to crystalline order are on an "equatorial fan" of diffuse intensity, implying a certain amount of inherent disorder within the tendon. Also the existence of crystallinity of any kind in many collagenous structures cannot be found, and seems to be limited to a number of specialised tissues.

1.3.2 OTHER COMPONENTS OF THE EXTRACELLULAR MATRIX

The molecules discussed here contribute well defined structural properties to the complex.

1.3.2.1 Mineral

This phase of mature bone contributes 65% by weight. Its function is to provide resistance to compression and torsion in the tissue. The principal components of the mineral phase are Ca^{2+} and Pi (inorganic phosphate). Although the chemical composition of bone has been known for many years, research has been relatively unsuccessful in revealing the molecular structure of bone and the nature of interaction between mineral and collagen.

1.3.2.2 Elastin

This represents a group of fibrous proteins that are found associated with other components of the extracellular matrix. They are found mainly in large arteries and specific parts of ligaments where they provide elasticity and high mechanical strength in the tissue. In these tissues they can be up to 50% of dry weight.

Electron microscopic examination of elastic fibres has revealed them to comprise two morphologically distinguishable components. The amorphous component (this does not usually posess any regular repeating structure) is quantitatively the largest proportion, composing upwards of 90% of the mature fibre.

The microfibrillar component, as its name implies, is composed of small fibrils 10-12nm in diameter. These are found primarily around the amorphous component.

The primary structure of elastin is remarkable in that over 90% of the residues are hydrophobic. The protein also contains the crosslink groups desmosine and isodesmosine. These are formed from lysine residues in neighbouring molecules. Elastin contains a high proportion of glycine residues, however sequence work has failed to show the characteristic tripeptide (Foster et al, 1973). The X-ray diffraction pattern also differs from that of collagen (Bigi et al, 1977). A review of elastin structure is by (Rosenbloom 1987).

1.3.2.3 Glycosaminoglycans/Proteoglycans

These are the major carbohydrate containing molecules in the connective tissue matrix. These molecules are as diverse as their roles and occurrence. The common motif of a proteoglycan is a protein core with long chains of glycosaminoglycans which are O-glycosidically linked. It is the diversity of the glycosaminoglycans and the size of the proteoglycans that give each its physical properties.

Glycosaminoglycans are linear polymers of repeated disaccharides containing, hexosamines, O-sulphated hexosamines and uronic acid (Hascall & Kimura 1982). The number of repeat disaccharide values varies, but typical values are of the order of 50.

The glycosaminoglycans are thus frequently polyanionic and therefore, when associated in the form of a proteoglycan, give it a very extended structure occupying a large volume of space. There are several types of glycosaminoglycan, some of which are described in Table 1.2.

The properties of proteoglycans can be summed as:

(1)Their ability to bind water by osmotic forces is regarded as important in the resistance to compression, as found in cartilage (Ogston 1970).

(2)They form dense networks, conferring high resistance to the flow of water and ions (Preston et al, 1972).

(3)They exhibit rheological properties, especially viscoelasticity, particularly noteworthy is hyaluronic acid.

(4)It has been suggested that the location of a dermatan sulphate rich proteoglycan in the gap region of type I collagen fibrils may form at least part of the physiological inhibition of calcification as it normally occurs, (Scott & Haigh 1985).

(5)In tendon, the contribution of proteoglycan in the composite is that individual collagen fibrils invariably seem to be surrounded by a layer of gel. This would limit the transmission of fractures to individual fibrils and help to maintain the tensile strength of the structure (Hukins 1984).

Table 1.2 Composition of glycosaminoglycans found in vertebrate collagen

Type	Tissue	Repeating disaccharide	mol wt (kDal)
hyaluronate	vitreous humour, skin, aorta	D-glucuronic D-glucosamine	4-8,000
chondroitin 4-sulphate	cartilage cornea skin	D-glucuronic D-galactosamine	5-50
chondroitin 6-sulphate	cartilage skin, aorta	D-galactosamine D-glucosamine	5-50
dermatan sulphate	skin, tendon aorta	L-iduronic acid D-galactosamine	15-40
skeletal keratan sulphate	cartilage annulus fib.	D-galactose D-glucosamine	8-12
heparin	lung, liver skin	D-glucuronic D-glucosamine	7-16

1.4 THE ANATOMY OF TENDON

A tendon is a band of fibrous tissue through which the tension on a muscle is transmitted to a part of the skeleton. The main body of the tendon serves only to transmit tensile stress and has the capacity for repair and growth.

The tendon consists of a hierarchy, the basic unit being the fibril. These associate together in a proteoglycan matrix as a primary bundle. The primary bundles are delineated from each other by a thin layer of fibroblasts. A group of primary bundles can be surrounded by a sheath-like structure, the paratenon, this structural unit has been defined as the fascicle (Rowe 1985). The paratenon consists of collagen, fibroblasts and ground substance. The groups of fascicles are encased in the epitenon which is thought to contain a system of interlacing collagen fibrils. This is continuous with the endotenon which delineates the tendon, see Figure 1.5.

The diameter of fibril varies between tendons in an organism and between species. Distribution of fibril diameter has been related to the age of tendon (Parry et al, 1978b), and its biochemical role (Parry et al, 1978a). Structures containing fibrils with a large diameter are usually stress bearing, whereas those containing fibrils with a smaller diameter are held to exhibit more elastic properties.

The fibrils in the tendon tend to display iridescence when viewed under a light microscope. This is due to the occurence of a crimp of approximately 200 microns along the length of the filament structure. The crimp is visible when no stress is applied to the tendon and is lost when sufficient stress is imposed to stretch the tendon by about 4% (Kastelic et al, 1978).

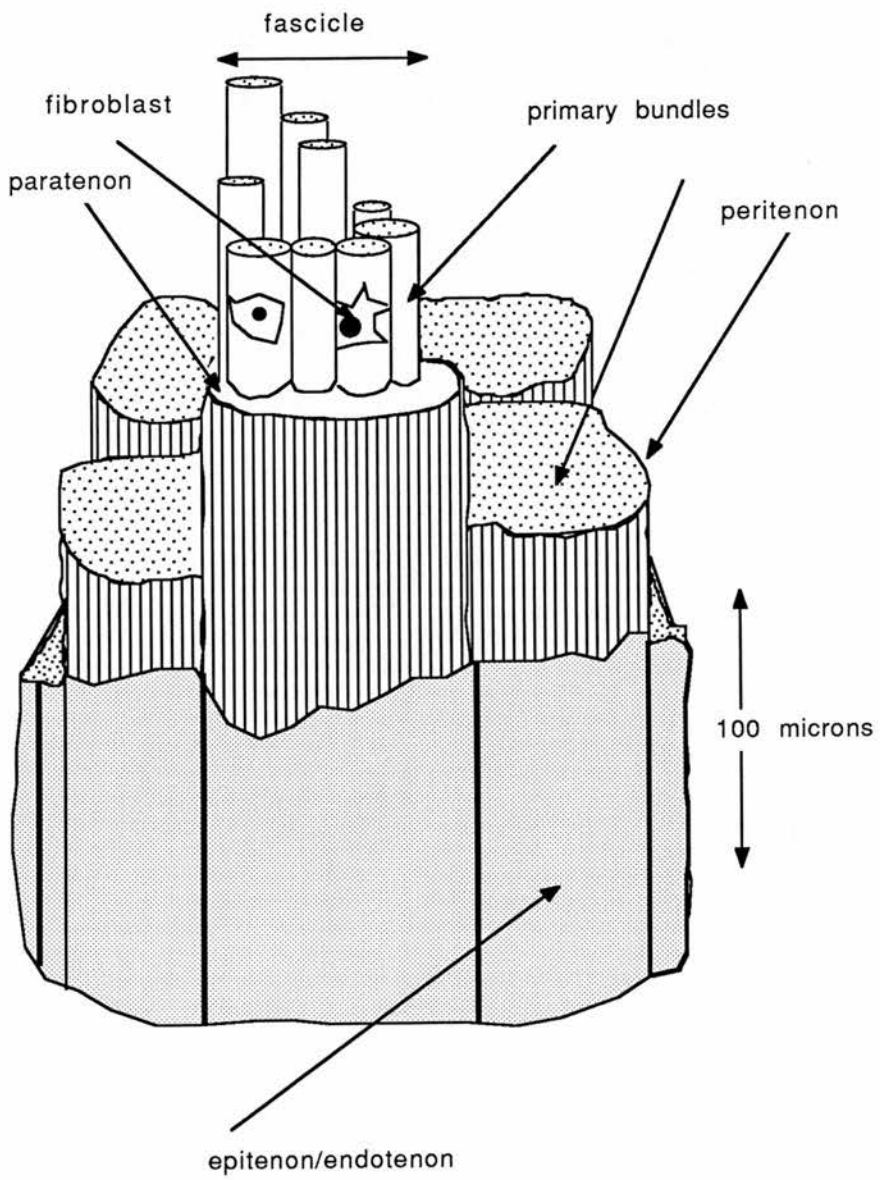


Figure 1.5, the anatomical structure of tendon.

(after Rowe 1985)

1.5 CONCLUSIONS

The information given in this chapter, relates to the historical background and currently held views on connective tissue structural research. Despite the wealth of research conducted on collagen containing structures, a large number of questions remain to be answered, and the answers given to some improved, (the result of answering, or attempting to answer these questions, inevitably results in the uncovering of more subjects for conjecture). The nature of many of these questions pertain to the structure of collagen and its interaction with other molecules. The answers require information relating to:

(1) The axial structure at high resolution.
(2) The conformation of telopeptides.
(3) The axial position of crosslinks and other reducible components.
(4) The various aspects of chemical reactivity of collagen.

It is the purpose of this thesis to reveal information relevant to the above areas of interest. In the following chapters, it is described how the powerful analytical tools of X-ray analysis, neutron diffraction and protein chemistry are employed together to advance an understanding of collagen related structure.

The result of this is:

(1) The presentation of axial structures of type I collagen derived from the methods of isomorphous addition phase determination. This structure has been resolved to a higher resolution than previously obtained, and does not employ a model building basis as used by Hulmes et al, (1977).

(2) The structure of the axial repeat in tendon was determined using a similar method to (1) but used neutron diffraction, this also supporting the X-ray structures, and also providing information appertaining to the position of crosslinks, and carbohydrates.

A more detailed understanding of collagen structure, can be used in the development of connective tissue research as a whole.

CHAPTER 2

THEORY OF DIFFRACTION

The majority of data presented in this thesis has been obtained using diffraction techniques. An understanding of the phenomena involved in the process of diffraction is necessary to evaluate the results of an experiment. The purpose of this chapter is to give a concise introduction to the theory of diffraction and techniques used to determine structure. A full description of scattering and diffraction is given in Woolfson (1970).

2.1. X-RAY DIFFRACTION SCATTERING PROCESS

2.1.1 THE TWO TYPES OF SCATTERING

A number of properties are displayed by electromagnetic radiation. The dichotomy of electromagnetic radiation wave-particle duality allows the interpretation of scattering phenomena to be attributed to the wave or particulate nature of radiation. The two types of X-ray scattering by electrons emphasise these two aspects of electromagnetic radiation.

1) Thomson Scattering

The interaction of a fluctuating electromagnetic field with a free electron, forces the electron into oscillations of the same frequency as the incident electromagnetic field. The oscillating charge is able to produce electromagnetic radiation of the same wavelength,

but differing in phase from the incident waveform by π radians.

Since all of the secondary wavefronts produced by this oscillating electron have the same phase difference from that of the incident wavefront, the scattering produced is coherent. Thomson scattering emphasises the wave nature of electromagnetic radiation.

2) Compton Scattering

This is also known as incoherent scattering and emphasises the particulate nature of X-rays. The effect of an incident photon interacting with an electron is that the photon is deflected from its path and experiences a loss of energy. This is also reflected in the change of wavelength of the incident waveform after deflection. For the purpose of X-ray diffraction the contribution of Compton scattering to the total amount of scattering is negligible except at high angles. The cooperative coherent scattering of many electrons in a crystalline array is significantly greater than that of the summed incoherent contributions.

2.1.2 SCATTERING FROM ONE CENTRE

Figure 2.1 represents scattering of an incident beam of X-rays at a centre O. The incident beam is in the form of a parallel monochromatic radiation. Scattered radiation emanates from this centre in all directions, however the intensity of scattered X-rays observable varies in relation to the position of the incident beam.

If the wavelength of incident radiation is monochromatic, the variation of displacement Y at a point in the waveform with respect to time can be described as:-

$$Y=A\cos 2\pi\nu t \qquad 2.1$$

Where A is the amplitude (maximum displacement) of the waveform

and ν is the frequency of the radiation.

Thus the scattered wave observed at the point P of Figure 2.1 will have a displacement that can be determined by taking into consideration:

1)The phase difference produced by the length of O-P = (L) not being an integer multiple of the incident wavelength.

The introduced phase shift is:

$$-2\pi L/\lambda$$

This can also be described as:

$$-2\pi L\,(\nu/c)$$

(λ and c being the wavelength and velocity of the incident radiation respectively) in relation to the scattered wave at point O.

2)The scattering process itself results in the introduction of a phase shift. The scattered wave at O is retarded with respect to the incident wave at O.
This can be quantified as the scattering phase shift, α_s.

3)The inverse square law of intensity applies in this case. The amplitude of the scattered waveform decreases as the distance from the scattering source increases.

The displacement of the wavefront at the point P relative to that of the incident wavefront at O can thus be determined by the equation:-

$$Y_{(2\theta,L,t)} = f_{2\theta}\,(A/L)\cos[2\pi\nu(t-L/c)-\alpha_s] \qquad 2.2$$

$f_{2\theta}$ is a constant of proportionality (the scattering length) a property associated with the scattering centre and is a function of the scattering angle 2θ.
The amplitude of $f_{2\theta}$ decreases as the angle of diffraction increases.

Interpretation of X-ray diffraction is aided if complex number notation can be used. Therefore the formula of a progressive wave can be described as:

$$Y = A\exp[2\pi i\nu(t-L/c)]$$
$$= A\cos 2\pi\nu(t-L/c) + iA\sin 2\pi\nu(t-L/c) \qquad 2.3$$

In complex number format the equation can be regarded as having a real and imaginary part. The real component of the equation represents the displacement, whereas the ratio of imaginary to real parts of the equation is the tangent of the phase associated with wave motion at (L,t) with respect to that at the origin. The time dependence of variation of displacement at a point can thus be given as:

$$Y_{(2\theta,L,t)} = f_{2\theta}\,(A/L)\exp[2\pi i\nu(t-L/c)-i\alpha_s] \qquad 2.4$$

Amplitude associated with the scattering at P can be described as:

$$\eta_{(2\theta,l)} = f_{2\theta}\,(A/L) \qquad 2.5$$

The phase lag of the scattered radiation at P relative to that at the origin can be described as:

$$\alpha_{op} = 2\pi\nu\, L/c+\alpha_s \qquad 2.6$$

2.1.3 SCATTERING FROM A PAIR OF CENTRES

The scattering from two identical centres can be described as follows. The incident radiation interacting with two points O_1 and O_2 can be scattered and the resultant scattering sampled at a point P, where the distance from P to O_1 is very much greater than that between O_1 and O_2. This is described graphically in Figure 2.2.

The radiation scattered is received at P and has been scattered through an angle of 2θ from both centres. In this case the scattering centres are identical, and therefore so is the phase shift α_s. The phase difference at a point P of radiation emanating from O_2 relative to O_1 is given by:

$$\alpha_{O_1 O_2} = -(2\pi/\lambda)(CO_2+O_2D) \qquad 2.7$$

Modular vector arithmetic can now be used to represent path differences. Two unit vectors S_o and S can be defined as being parallel to the incident and scattered beams respectively. If the vector joining O_1 to O_2 is defined as r then it can be seen that:

$$CO_2 = \mathbf{r}.\mathbf{S_o}, \quad O_2D = -\mathbf{r}.\mathbf{S}$$

Substituting into equation 2.7 gives,

$$\alpha_{O_1 O_2} = -2\pi/\lambda.\mathbf{r}(\mathbf{S}-\mathbf{S_o}) \qquad 2.8$$

The bracketed quantity in 2.8 may be simplified by replacing it with an equivalent vector s.

$$\mathbf{s}=(\mathbf{S}-\mathbf{S_0})/\lambda \qquad 2.9$$

giving:

$$\alpha_{O_1O_2} = 2\pi \mathbf{r}.\mathbf{s} \qquad 2.10$$

All scattered beams can be described in terms of s, and its reciprocal dimensionality often leads to s space being referred to as reciprocal space. The important relation of vectors s, S_o and S can be seen in Figure 2.3.

The unit vectors $\mathbf{S_o}/\lambda$ in the incident direction and $\mathbf{S}/\lambda$ in the scattered direction have equal magnitude of $1/\lambda$. It can be seen that the vector s is perpendicular to the bisector of $\mathbf{S_o}$ and $\mathbf{S}$ and has a magnitude of:-

$$\mathbf{s}= (2\sin\theta)/\lambda \qquad 2.11$$

The scattering from one centre described in the previous section can be combined with the effects from scattering from two centres to produce an equation representing the resultant disturbance at P. If the displacement due to incident radiation is described by equation 2.1, then displacement at a point P at a distance L from the point O_1 can be given as:

$$Y_{(2\theta,L,t)} = f_{2\theta}\, A/L \left\{\exp[2\pi i\nu(t-L/c)-i\alpha_s] + \exp[2\pi i\nu(t-L/c)-i\alpha_s+2\pi i\mathbf{r}.\mathbf{s}\right\} \qquad 2.12$$

$$= f_{2\theta}\, A/L[\exp 2\pi i\nu(t-L/c)-i\alpha_s].(1+\exp 2\pi i\mathbf{r}.\mathbf{s}) \qquad 2.13$$

The amplitude of this can be described as:

$$\eta_{2\theta}(2\theta,L) = f_{2\theta}\, A/L\, (1+\exp 2\pi i\mathbf{r}.\mathbf{s}) \qquad 2.14$$

Using equation 2.5, this may be described in terms of scattering amplitude from a single unit.

$$\eta_{2\theta}(2\theta,L) = \eta(2\theta,L)\, (1+\exp 2\pi i\mathbf{r}.\mathbf{s}) \qquad 2.15$$

The process of scattering has however so far been described using the scattering point O_1. To further the description of the scattering process it is necessary to make the point from which phases are determined to be an arbitary one, O.

The relative positions of O_1 and O_2 can be described by the vectors r_1 and r_2 respectively, this makes equation 2.12 appear as:

$$\eta_{2(2\theta,L)} = \eta_{(2\theta,L)} (\exp 2\pi i \mathbf{r}_1.\mathbf{s} + \exp 2\pi i \mathbf{r}_2.\mathbf{s}) \qquad 2.16$$

2.1.4 SCATTERING FROM MANY CENTRES

An expansion of the description of scattering from two centres allows us to consider the scattering from many identical scatterers $O_1, O_2, O_3 \ldots\ldots O_n$. The scattered rays can be observed at a point P which is a long distance from the identical scatterers. Each point of scattering can be described as having a vector displacement r_j from an arbitary origin O. The resultant amplitude of these scatterers can be described by expanding equation 2.16 to accommodate a number of points.

$$\eta_{n(2\theta,L)} = \eta_{(2\theta,L)} \sum_{j=1}^{n} \exp 2\pi i \mathbf{r}_j.\mathbf{s} \qquad 2.17$$

This equation only applies to a system of identical scattering points. To accomodate scatterers that are not identical the scattering length of each must be within the summation in the formula.

This still makes the assumption that the phase shift α_s is the same for each scatterer.

$$\eta_n(2\theta,L) = A/L \sum_{(f_{2\theta})j} \exp 2\pi i \mathbf{r}_j.\mathbf{s} \qquad 2.18$$

In X-ray scattering the electrons are the secondary emitters and it is convenient to describe their distribution as a density function $\rho(x)$. Since the function is continuous, the summation becomes an integration and can be described as:

$$F(\mathbf{s}) = \int \rho(\mathbf{x}) \exp(2\pi i \mathbf{s}.\mathbf{x}) dV\mathbf{x} \qquad 2.19$$

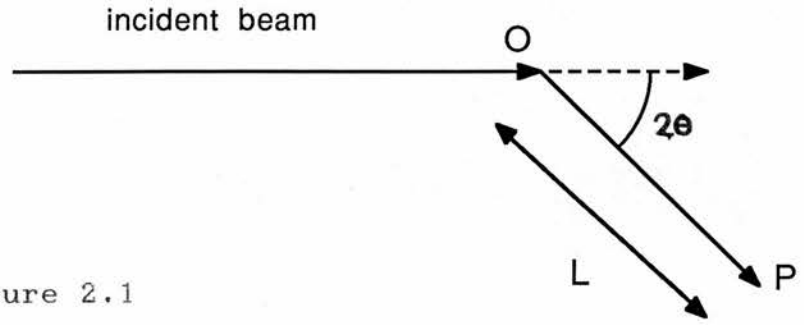


Figure 2.1

scattering from one centre.

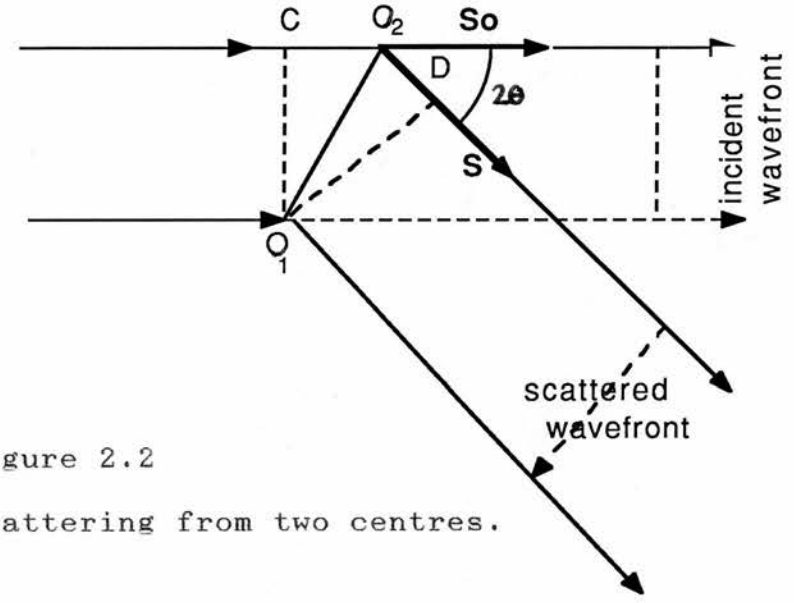


Figure 2.2

scattering from two centres.

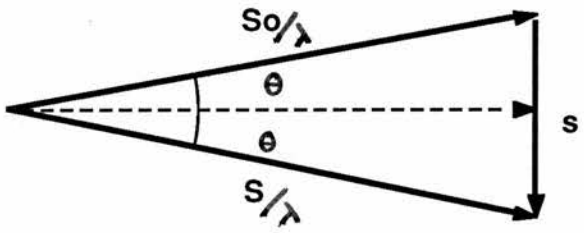


Figure 2.3 the relationship between **So**, **S** and **s**.

2.2 DIFFRACTION

2.2.1 THE RELATION OF THE FOURIER TRANSFORM TO DIFFRACTION

The continuous function of electron density can be constructed from a series of sine and cosine waves. The Fourier transform is a complex function as is the general scattering equation. The equation describing a Fourier transform is:

$$F(s)= \int C(\mathbf{x}) \exp(2\pi i \mathbf{s}.\mathbf{x}) d\mathbf{x} \qquad 2.20$$

Where F(s) is the Fourier transform.

It describes a continuous function as its fundamental and harmonic wave components. It can also be proved that the transform of the wave components is the original function (Woolfson 1970).

$$C(\mathbf{x})= \int F(\mathbf{s}) \exp(-2\pi i \mathbf{s}.\mathbf{x}) d\mathbf{s} \qquad 2.21$$

The direct similarity of the general scattering equation with that of the Fourier transform, shows that the scattering from a continuous electron density function, is the Fourier transform of the continuous electron density function and vice versa. The use of the Fourier transform to relate scattering and structure has been central to the elucidation of atomic and molecular structure using X-rays.

2.2.2 DIFFRACTION AND CRYSTAL STRUCTURE.

The interaction of X-rays with matter is weak, compared to the interaction of light with matter. This makes the observation of scattering due to an individual molecule impossible. Hence it is only possible to observe the scattering effects produced if the effects are amplified. This can be achieved if a large number of identical molecules are arranged in a regular array, a crystal. The effect of scattering from one molecule is then reinforced by that of the others in the process of diffraction.

The arrangement of scattering centres in a regular array leads to the interference of scattered radiation. This has the effect of increasing the total amount of coherent scattering to levels that can be recorded experimentally. However the effect of diffraction also results in the continuous scattering due to one molecule being resolved into a series of discrete peaks. This is known as the sampling of the molecular transform by the lattice transform.

Mathematically a crystal can be described as the convolution of a molecule with a lattice.

$$\rho_c(\mathbf{x}) = \rho(\mathbf{x}) * z(\mathbf{x}) \qquad 2.22$$

where $\rho_c(\mathbf{x})$ is the crystal electron density distribution

$\rho(\mathbf{x})$ is the electron density distribution in the molecule

and $z(\mathbf{x})$ is the crystal lattice function.

The Fourier transform of a crystal is given by as:

$$F(\mathbf{s})= \int \rho_c (\mathbf{x}) \exp(2\pi i \mathbf{s}.\mathbf{x}) dV \qquad 2.23$$

The "Convolution Theorem" states that the Fourier transform of two convoluted functions is equal to the multiplication product of the Fourier transforms of the individual functions, (Blundell and Johnson 1976).
Thus expanding the above gives:

$$F(s)= \int \rho(\mathbf{x})^* z(\mathbf{x}) \exp(2\pi i \mathbf{s}.\mathbf{x}) dV \qquad 2.24$$

The transform of the element $\rho(x)$ - the molecular transform - has already been considered, the transform of the lattice function can be expressed in a one dimensional case as follows:-

$$\sigma(\mathbf{s})=\int z(\mathbf{x}) \exp(2\pi i \mathbf{s}.\mathbf{x}) dx \qquad 2.25$$

where $\sigma(s)$ is the lattice transform.

The lattice may be termed as a regularly spaced array of delta functions. In one dimension this may be represented as:

$$z(\mathbf{x}) = \Sigma \delta(\mathbf{x}-n\mathbf{b}) \qquad 2.26$$

where **n** represents an integer

and **b** represents the lattice spacing.

The Fourier transform of this $\sigma(\mathbf{s})$ is given by the equation:

$$\sigma(\mathbf{s})= \sum_{n}\int\delta(\mathbf{x}-n\mathbf{b})\exp(2\pi i\mathbf{s}.\mathbf{x})d\mathbf{x} \qquad 2.27$$

The delta function is only non-zero at points where $\mathbf{x}=n\mathbf{b}$ and thus the integration reduces to a summation giving:-

$$\sigma(\mathbf{s})= \sum_{n}\exp(2\pi i\mathbf{s}.n\mathbf{b})$$
$$=\sin(N\pi\mathbf{s}.\mathbf{b})/\sin\pi\mathbf{s}.\mathbf{b} \qquad 2.28$$

This function is equal to zero except for peaks of height **N** and width **1/N** which are located at values for which (**s.b**) is an integer.

This is when:

$$\mathbf{s}= 1/\mathbf{b},2/\mathbf{b},3/\mathbf{b}.....n/\mathbf{b}$$

For a crystal, n is very large and the resulting Fourier transform of the lattice function has, sharp peaks at these values in **s** space. The multiplication of this function with the Fourier transform of the individual molecule results in the Fourier transform of the crystal. This can be described as the sampling of the molecular transform at the non zero points of the crystal lattice transform.

These non-zero points of the reciprocal lattice extend in three dimensions of reciprocal space occurring at:

$$s = ha^* + kb^* + lc^*$$

where h,k and l are integers known as the Miller indices, and a*,b* and c* represent the unit cell dimensions of the reciprocal lattice.

2.2.3 FOURIER SYNTHESIS

It was stated earlier, and shown in equation 2.18 that the Fourier transform of an electron density can itself undergo a Fourier transform to produce the electron density again. This when applied to a crystal gives:

$$\rho(x,y,z) = 1/v \sum\sum\sum F_{hkl} \exp[-2\pi i(hx+ky+lz)] \qquad 2.29$$

x,y and z are fractional coordinates in the real space cell of volume V.

Each non zero point in the crystal transform has a reciprocal coordinate value (h,k,l) and a value F_{hkl} the structure factor, corresponding to a wave of electron density. The structure factor is complex and can be described as:

$$F_{hkl} = |F_{hkl}| \exp(i\Phi_{hkl}) \qquad 2.30$$

F_{hkl} can be measured from an X-ray diffraction

pattern of a crystal. It is the square root of the intensity associated with a **hkl** coordinate. Using this however does not allow us to use equations 2.29 and 2.30 to calculate the electron density of the unit cell in real space since the imaginary part of the equation has still to be determined. The ratio of the real to the imaginary parts of this equation is the tangent of the phase associated with each structure factor.

It is essential that the phase associated with each structure factor be determined since without this information the electron density cannot be found. This information cannot be recorded on film since the phase of each diffracted ray cycles through 2π radians with the frequency of the incident radiation.

No detector has so far been constructed that is able to record events cycling at such a high frequency. The determination of a phase for each **hkl** reflection to be used in a Fourier synthesis has posed a problem at the core of X-ray diffraction analysis, the so called phase problem.

2.3 TACKLING OF THE PHASE PROBLEM

A variety of procedures have been proposed to determine the phase information necessary to determine the structure of a molecule. The merits of various methods are discussed in Blundell & Johnson (1976). In the context of this thesis, the methods of phase determination by model building and multiple isomorphous derivative production are most relevant and are discussed here.

2.3.1 MODEL BUILDING

This method of phase determination relies on the ability to model the electron density of a structure, so that the Fourier transform of this produces structure factor amplitudes that match those of the diffraction pattern. In the resulting Fourier transform the phase associated with each structure factor will be given.

In construction of such an electron density or model, information relating to the structure obtained from other analytical techniques can be used, for example electron microscopy and amino acid sequence data. Using model building in this way only allows the X-ray data to test whether a model is correct or not. If a process of refinement is used, then the model parameters may be altered slightly until a best fit is obtained against the observed diffraction data.

Model building techniques are regarded as time

consuming, and were initially only employed in the structure determination of small molecules (up to about 25 atoms). It has been used to some success in determining a set of phases for the axial structure of rat tail tendon (Hulmes et al, 1977). Here sequence information was combined with estimated values of D periodicity to produce axial electron density models. The Fourier transform of these could be calculated and the resultant amplitudes compared with collected data. Although this technique has proved useful, it must be regarded as being a trial and error procedure.

2.3.2 ISOMORPHOUS REPLACEMENT/ADDITION

The method of phase determination by isomorphous heavy atom labelling was first successfully used by Green et al, (1954) to determine the structure of horse myoglobin.

The technique can be regarded as consisting of two closely related procedures.

1)Isomorphous addition; this involves the addition of a heavy atom to a structure.

2)Isomorphous replacement; this involves the replacement of an ion, atom or molecule by a similar group which is different in electron density than the original.

Apart from this difference in the labelling of the molecule, the two methods can be regarded as identical in terms of the ability to aid structural determination.

The method chosen to introduce heavy atoms to the structure often depends on the suitability of the protein to be labelled more successfully by one of the techniques. Both methods have the criterion that the structure of the molecule under observation remains unchanged, and thus the only difference in electron density between the isomorphous derivative and the native structure is a peak at the position occupied by each heavy atom.

The effect of introducing heavy atoms into a protein structure at specific locations results in an alteration of X-ray diffraction peak intensity, without altering the position of each peak. The intensities of the native and derivative X-ray diffraction peaks can be used to solve the phase problem. Figure 2.4 shows a vectorial representation (Argand diagram) of the relationship between the native, derivative and label only structure factors $\mathbf{F_p}$, $\mathbf{F_{ph}}$ and $\mathbf{F_h}$.

It can be seen from this diagram that the derivative structure factor consists of two components, the native structure factor and the label structure factor.

If the position of the heavy atom(s) in the unit cell is known, then it is possible to determine their contribution to each derivative structure factor. The calculated contribution of the heavy atoms in the derivative can be combined with the derivative and native amplitudes, obtained from observed intensities. From this information the phase of the structure factor for the native molecule unit cell can be determined. In order to

do this, it is convenient to employ what is known as the Harker construction (Harker 1956). This is illustrated in Figure 2.5.

The vector $-\mathbf{F_h}$ is drawn from the origin O. A circle of radius equal to $|\mathbf{F_{ph}}|$ is drawn centred on the end of the vector at **Q**. The other circle of magnitude $|\mathbf{F_p}|$ is centred at the origin. The intersection of the circles usually occurs at two points **G** & **H**. The vectors **OG** and **OH** represent two alternative structure factors for the native molecule unit cell. Both these structure factors have the same amplitude, but different phases.

To determine which of these is the correct phase, another isomorphous derivative is required. This second derivative must label at sufficiently different locations so as to provide X-ray derivative intensities and also heavy atom structure factors that are sufficiently different from the first derivative. The superposition of the Harker construction for a second derivative removes the ambiguity associated with phase determination, since the intersection of the three circles only occurs at one point, corresponding to the correct native phase.

There are cases where at low resolution, or in certain crystal planes, the unit cell can exhibit centrosymmetry. This means that the vectors **Fh** and **Fp** are colinear in the Harker construction and therefore the intersection of the circles occurs only at one point and thus the phase can be determined using only one derivative.

The essential part of phase determination using this method is the production of derivative molecules containing heavy atoms. The introduction of the heavy atoms must not disrupt the molecular or crystal structure, and it must also be possible to determine the position of heavy atoms within the unit cell.

2.3.3 THE PATTERSON FUNCTION

The Patterson function can be employed in the analysis of data where only the intensities of the diffraction maxima are available. It is therefore often used in the initial stages of phase determination by isomorphous methods, since it can be used as an aid to finding the distances between, and thus eventually the position of heavy atoms in a unit cell.

The function in one dimension is described as:-

$$P(u)= \int 1/a \ \rho(\mathbf{x})\ \rho(\mathbf{x}+\mathbf{u})dx \qquad 2.31$$

"a" is the length of the periodic repeat in the structure

Thus the value of the function at a point v(u) is calculated from the density at points x and x+u, this is summed over the whole unit cell. The vector **v** and the density product combined with it, can be termed as the constituents of Patterson space.

The significance of this function is that the value of P will be high where the length of the vector **v** is that corresponding to the distance between groups of high

electron density in the structure. A Patterson map therefore is a representation of the electron density associated with each end of a vector, where the length of the vector is varied to analyse all spacings in the unit cell.

The Patterson map corresponding to a molecule is more complex than that of the original molecular structure, since if the structure contains N atoms, then the number of vectors in the Patterson map is N^2. N of these vectors are self-self vectors i.e. at the origin of the Patterson map there are a number of vectors of zero length. The remaining vectors N(N-1) are distributed throughout the unit cell. The crowded nature of Patterson space and the centrosymmetry imposed by the function makes the interpretation difficult. However the use of Patterson function analysis combined with amino acid sequence information, provides a powerful tool to initiate phase determination by locating the positions of heavy atom labelling within the unit cell.

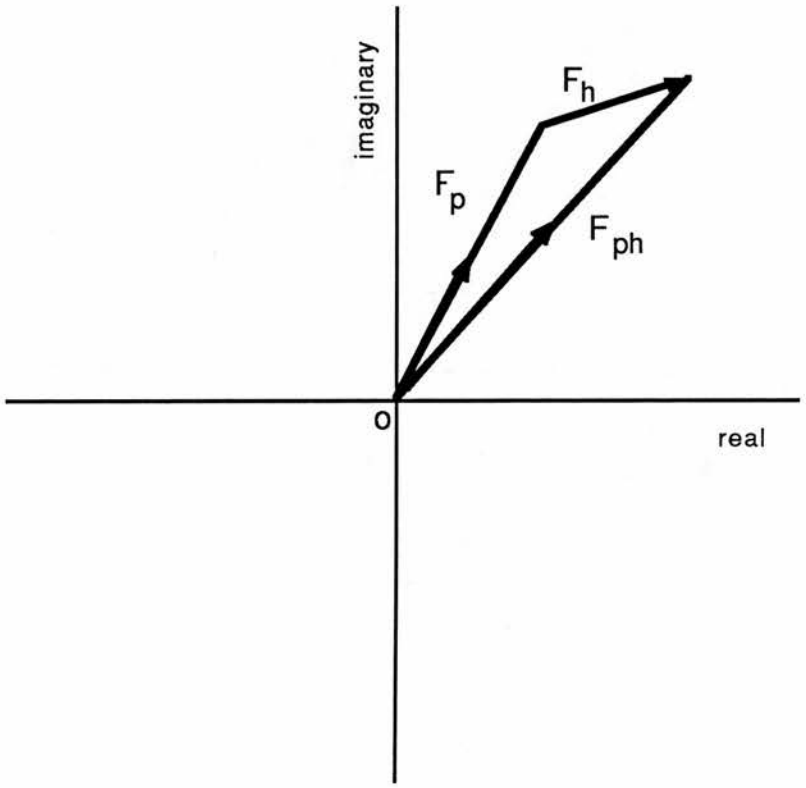


Figure 2.4. the native protein (F_p) and heavy atom (F_h) contribution to the structure factor (F_{ph}).

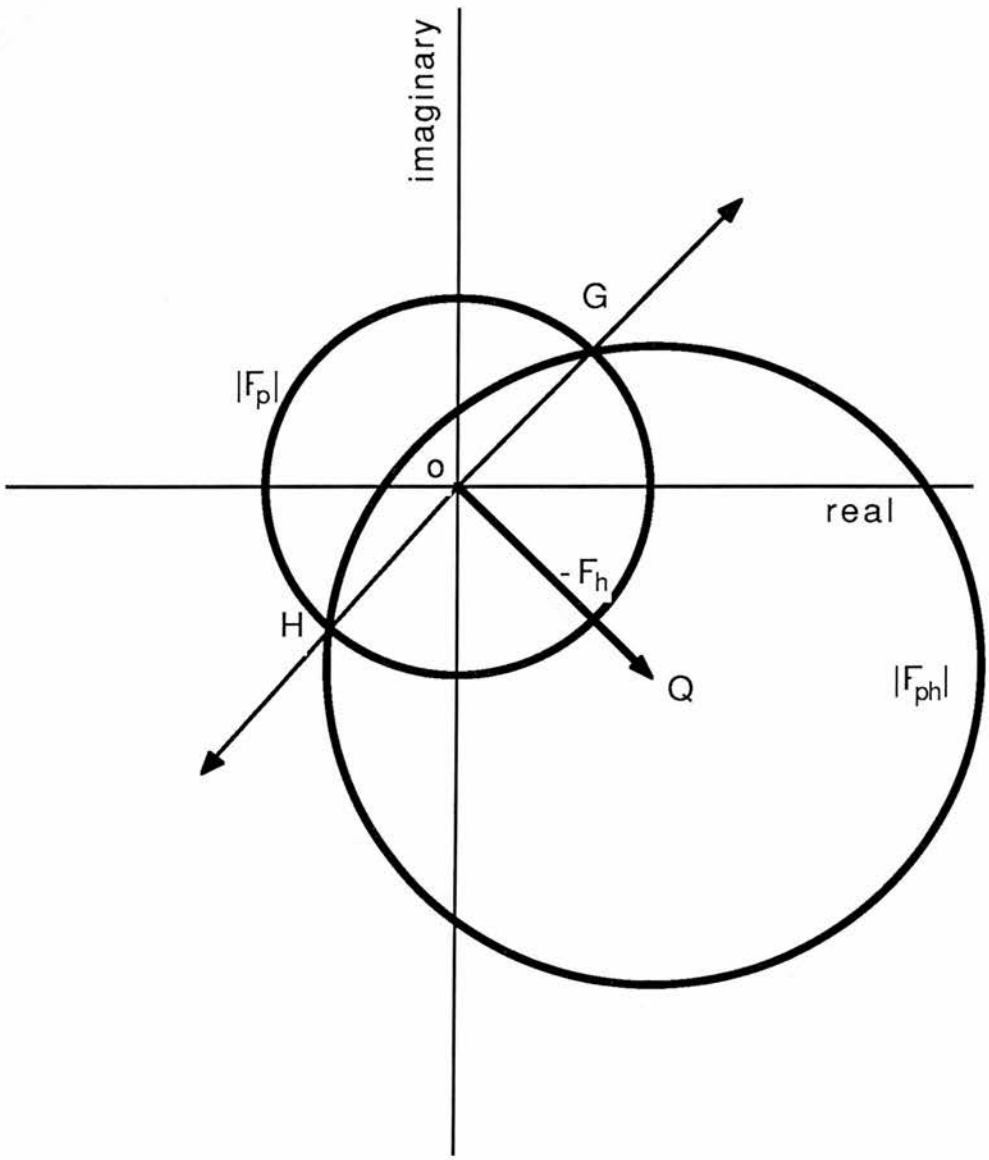


Figure 2.5. the Harker construction.

2.3.4 THE EWALD CONSTRUCTION

An aid to the description of diffraction phenomena has been the Ewald construction after Ewald (1921). This is a mathematical description of diffraction that can be represented graphically allowing the prediction of constructive interference and thus diffraction in reciprocal space. In Figure 2.6 a sphere is drawn with centre C at the crystal, and this has a radius of $1/\lambda$. The origin of the reciprocal lattice is the point O. This represents the intersection of the unreflected beam and the surface of the sphere.

The condition of diffraction, for example the production of the diffracted ray along CB is when the point B is a non zero point of the reciprocal lattice which is also lying on the surface of the sphere. Therefore any reciprocal lattice point can be brought into a position to allow diffraction by tilting the crystal until the particular reciprocal lattice point lies on the sphere of reflection. To record diffraction maxima corresponding to each F_{hkl} value, the crystal has to be rotated until all corresponding reciprocal lattice coordinates have passed through the Ewald sphere.

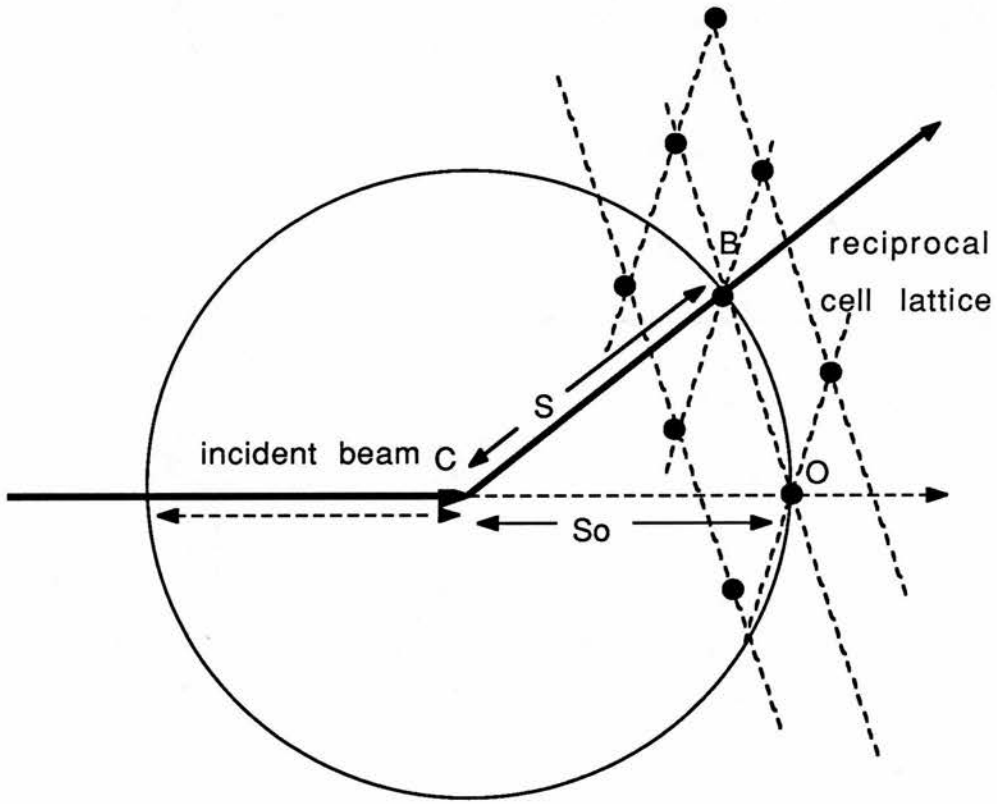


Figure 2.6. the Ewald construction.

2.3.5 THE NATURE OF FIBRES

The description of diffraction phenomena has concentrated on diffraction from crystals. The crystalline nature of a solid may be considered as the repetition of the molecular electron density at a series of 3-dimensional lattice points. The nature of various polymers is such that they are unable to form such ordered systems, but are able to form fibres.

In a fibre the degree of ordering is found to be anisotropic. The axial order parallel to the fibre axis is greater than that found perpendicular to this in the lateral direction. The extent of lateral order is displayed over a smaller range, and often includes a number of distortions. The nature of these distortions can be described as follows.

1) A sample may be composed of microcrystalline domains which have a common axial direction.
The orientation of the domains around the axis is random and thus the fibre diagram resembles the diffraction pattern of a crystal rotated about a principal axis.

2) A sample may consist of molecules that are only able to associate laterally in a quasi crystalline lattice with a regular axial structure.

3) Non crystalline fibres exist where lateral disorder is so great that the lattice effects are negligible and only the molecular transform may be observed.

4) In oriented gels, the colloidal solution of long rod like molecules can show spontaneous birefringence indicating that the molecules have orientation parallel to each other. The degree of angular spread around the common axis can be as little as 1°.

The fibrous nature of tendon is exhibited by the highly ordered axial displacement with a less ordered lateral array. The off meridional rowlines, of X-ray diffraction patterns, indicate some degree of crystallinity in tendon (Fraser et al, 1983). However diffuse scatter on X-ray diffraction patterns, indicates an inherent degree of disorder.

2.3.6 THE PROJECTION THEOREM

The projection theorem states that the two dimensional transform of a plane projection of a three dimensional electron density function, is identical with the central section of the three dimensional transform normal to the direction of projection. This theorem generalises to that of a line projection (Woolfson 1970), and has a central role in the interpretation of fibre diagrams. Order along the fibre axis is known to be greater than that found laterally. The correspondingly well defined meridional X-ray diffraction intensities, can therefore be seen to be the transform of an axial projection of the electron density onto the fibre axis.

The meridional reflections can be regarded as a one dimensional crystal diffraction pattern. Therefore the equations applicable to diffraction from crystalline forms, can be applied to the meridional diffraction intensities of tendon if they are expressed in a one dimensional form.

2.4 NEUTRON DIFFRACTION THEORY

The general theory of diffraction as described for X-rays can be applied to diffraction using neutrons. There are however some fundamental differences in the scattering process. These are described here since they account for the differences between neutron diffraction and X-ray diffraction.

2.4.1 NEUTRON & X-RAY SCATTERING

X-rays are scattered principally by electrons; neutrons however are scattered by interactions with the atomic nuclei. The diameter of the electron cloud around the nucleus is comparable to the wavelength of X-rays, this can explain the decrease in scattering amplitude as the scattering angle decreases. The scattered X-rays from an atom are built up from the contributions of extra nuclear electrons.

In the direction of the incident beam path, these contributions will be in phase with each other and the resultant amplitude is proportional to the atomic number. As the angle between the incident and the scattered wave increases, the phase differences between contributing waveforms increase. Thus the amplitude decreases.

The situation with neutrons is different, the size of the nuclei is small compared to the wavelength of the incident radiation and thus the nucleus acts as a point

scatterer. The effect of this is that the scattering of neutrons is effectively equal in all directions.

The nature of the interaction between a neutron and atomic nucleus and subsequent scattering length are dependent on a number of factors other than the atomic number of each type of atom. Within a factor of 2-3 most atoms scatter equally well, this is in contrast to the rapid increase of scattering length with atomic number associated with X-ray scattering.

2.4.2 NEUTRON SCATTERING

The following is a description of the wave nature associated with neutrons and their interaction with matter. A more extensive treatment is by Bacon (1975)

The principle of wave particle duality implies that neutrons of speed υ and mass M_N have an associated wavelength of:

$$h/M_N v = \lambda \qquad 2.32$$

a description of this as a wave function is:

$$\Psi n = \exp(-2\pi i v/\lambda) \qquad 2.33$$

This wave function when scattered by a neutron produces a spherical waveform:

$$\Psi s = -b/r \exp(2\pi i v/\lambda) \qquad 2.34$$

Here **b** is the scattering length of the nucleus and is analagous to $f_{2\theta}$ for X-rays; **r** is the distance of the point of measurement from the origin at which the nucleus is assumed to be rigidly fixed.

For the case of scattering that produces diffraction, only the coherent (elastic) scattering is considered, and λ remains invariant. Another parameter associated with this form of scattering is the scattering cross section, this is defined as:

$$\sigma = 4\pi b^2 \qquad 2.35$$

The size of this factor is determined by the size of the nucleus and spin associated with the nucleus. For a nucleus with zero spin, this will be defined mainly by the potential scattering. This in turn depends on the radius of the nucleus and the mass number. If the nuclear spin is not zero, then the resultant scattering is of two types (Halpern & Johnson 1939). The formation of a compound nucleus consisting of the spin nucleus and neutron can result in two spin states $I+^1/_2$ and $I-^1/_2$ these have a scattering length b^+ and b^- respectively. Therefore in a nucleus, the total scattering can be considered as the sum of two terms:

$$\sigma = \zeta + \delta$$

ζ is the coherent scattering, i.e scattering that is coherent with that by other nuclei and which can therefore produce interference.

The other factor δ is the cross-section for incoherent scattering. This can be significant compared to the case in X-rays and therefore the effect of incoherent scattering is to produce a high background count. In the case of most atoms, the contribution of coherent scattering to total scattering is significant (>75%). This is not the case with Hydrogen, where the coherent scattering is only approximately 2% of total.

The process of neutron scattering can therefore be seen to be different from that of X-ray scattering. The resultant diffraction phenomena can be regarded as being analagous as is the use of the Fourier transform and other functions in structure determination.

2.4.3 APPLICATIONS OF NEUTRON DIFFRACTION

The scattering process of neutron diffraction is sufficiently different from that of X-ray diffraction and these differences can be used to aid structural determination that could not be obtained by X-rays alone.

In the case of X-ray diffraction the scattering length associated with an atom that has a large electronic configuration, is accordingly large. In the case of Pt, the X-ray scattering length is 10 times that of an atom normally found in proteins. This provides the basis for isomorphous addition/replacement phase determination. In terms of neutron scattering scattering length, that of Pt is only approximately 1.4 that of C and therefore the

use of heavy atoms in neutron isomorphous replacement does not usually apply, although some exceptions are possible e.g. Dy^{164}.

In neutron diffraction the less conventional technique employable is to exploit the scattering length differences of deuterium and hydrogen. The scattering length of hydrogen is negative whereas that of deuterium is positive and comparable to that of other atoms found in proteins. Isomorphous derivatives of a protein may be produced if hydrogen atoms in the original structure can be specifically replaced by deuterium atoms. This has been conducted in a number of novel ways including the use of macromolecules from organisms grown in D_2O.
The production of deuterated collagen is discussed in chapter 5.

Another application of neutron diffraction is that at sufficiently low resolution, a large structure can be assigned an average neutron scattering length density.

$$\rho_{b\ (av)} = \Sigma b_j / V_m \qquad 2.36$$

b_j is the scattering length of the components.

V_m is the total volume of the components.

At low resolution, the effect of diffraction can be described as being due to a difference in the average scattering length of a molecule and its background medium. Table 2.1 shows H to have a negative scattering length and thus it can be seen that water has a lower scattering length than that of protein and D_2O.

In a bi- or multiphase system, where the components have sufficiently different scattering lengths, it is possible to contrast out various components and study the diffraction contribution from the other(s). This has been used in the investigation of calcification of collagen (White 1977).

Table 2.1

Neutron and X-ray scattering data for relevant atoms

	Neutrons			scattering amplitude X-rays
Element	b	coh	tot	sin θ\ λ
	10^{-12} cm	(barns)	(barns)	(Å^{-1}) =0.5Å^{-1}
H^1	-0.374	1.76	81.5	0.02
H^2	0.667	5.59	7.6	0.02
C	0.665	5.56	5.51	0.48
N	0.94	11.1	11.4	0.53
O	0.58	4.23	4.24	0.62
P	0.51	3.27	3.6	1.83
Pt	0.95	11.34	12	12.1
I	0.53	3.58	3.8	7.7
Au	0.76	7.26	9	12.3

b is the coherent scattering amplitude of neutrons
coh is the coherent scattering cross-section
tot is the total scattering cross-section

CHAPTER 3

INSTRUMENTATION AND DATA COLLECTION

3.1 INTRODUCTION

This chapter describes the way in which diffraction experiments using X-rays and neutrons are performed. Therefore it encompasses a wide range of topics from beam generation to sample preparation and methods of data collection and correction.

3.1.1 X-RAY PRODUCTION

The X-ray diffraction data collected and presented in this thesis was produced using the synchrotron source at the S.E.R.C., S.R.S. (Synchrotron Radiation Source), Daresbury. The synchrotron consists essentially of a beam of charged particles, in this case electrons, orbiting an evacuated ring at relativistic speed. Electrons are provided by an injection system, then accelerated and focussed by magnetic fields. The effect of angular acceleration of the electron around the ring results in the production of a wide wavelength band of electromagnetic radiation at a tangent to the electron path.

The S.R.S. was built between 1975 and 1980. It is a 2GeV electron storage ring dedicated to production of synchrotron radiation. The importance of spectral brilliance (photons per unit area per unit solid angle) in experimental investigation necessitated the upgrading of the synchrotron so as to reduce the horizontal

emission of radiation from 1.6×10^{-6} mrad to 1.1×10^{-7} mrad. This also resulted in a 15 fold increase in brilliance. This was facilitated by the installation of the High Brightness Lattice (H.B.L.) in the main storage ring. The installation took place from October 1986 to mid June 1987 (right in the middle of the research period for this thesis!)

The synchrotron consists of:-

1)LINAC - the linear accelerator.

This delivers electrons for the synchrotron, and usually operates at 12MeV. Electrons are "boiled off" from a heated cathode and accelerated by an anode to a high potential, firstly across a D.C. potential in the electron gun, and then to 12MeV in a S band accelerator system.

2)Booster synchrotron.

The electrons from the accelerator are further boosted by the use of a small ring synchrotron. This provides the beam for injection into the synchrotron storage ring at 600MeV. A transfer path delivers the electrons to the storage ring.

3)Storage ring.

This controls the electron beam to a small cross sectional area, and allows the release of electromagnetic radiation along beamlines. It accumulates electrons from the injection system, accelerates them to a peak energy of 2GeV and keeps them in an orbit for many hours.

Quadrupole magnets are used to bend the electrons into their nearly circular orbit.

A high power radio frequency accelerating system is required to accelerate the electrons from 600MeV to 2GeV and restore the energy lost by the electrons as synchrotron radiation, see Figure 3.1.

The resultant radiation from the synchrotron has a number of properties:

1)High flux, an electron of charge e moving with energy E GeV in an orbit of radius R radiates:

$$\Delta E = 88.5e^4/R$$

Where ΔE is the energy lost as radiation

This indicates that the intensity of radiation from the synchrotron is $10\text{-}10^7$ orders of magnitude brighter than conventional rotating anode X-ray sources (Perkins 1988).

2)The angular divergence of radiation produced is small and comparable to that of a laser. The beam of electromagnetic radiation provided is thus of high flux per unit area cross-section.

3)The electromagnetic radiation produced is polychromatic and includes wavelengths of 1000-0.1Å, that is the U.V. to X-ray regions.

4)The beam is highly polarised with 100% polarisation in the plane of the storage ring.

5)The electron density in the ring has a precise time structure. This results in the radiation consisting of pulses, and can be utilised in time resolved experimentation.

The high flux of electromagnetic radiation provided by the synchrotron enables the diffraction pattern of samples that scatter weakly to be viewed in minutes as opposed to a matter of days using a conventional source. Since the effect of beam damage on a sample is proportional to the time spent exposed to the X-ray beam as well as the amount of flux passing through the sample, the relation of the diffraction pattern of a sample to the *in vivo* conformation of the molecule is greater than that found using a conventional source Mandelkow *et al*. (1982).

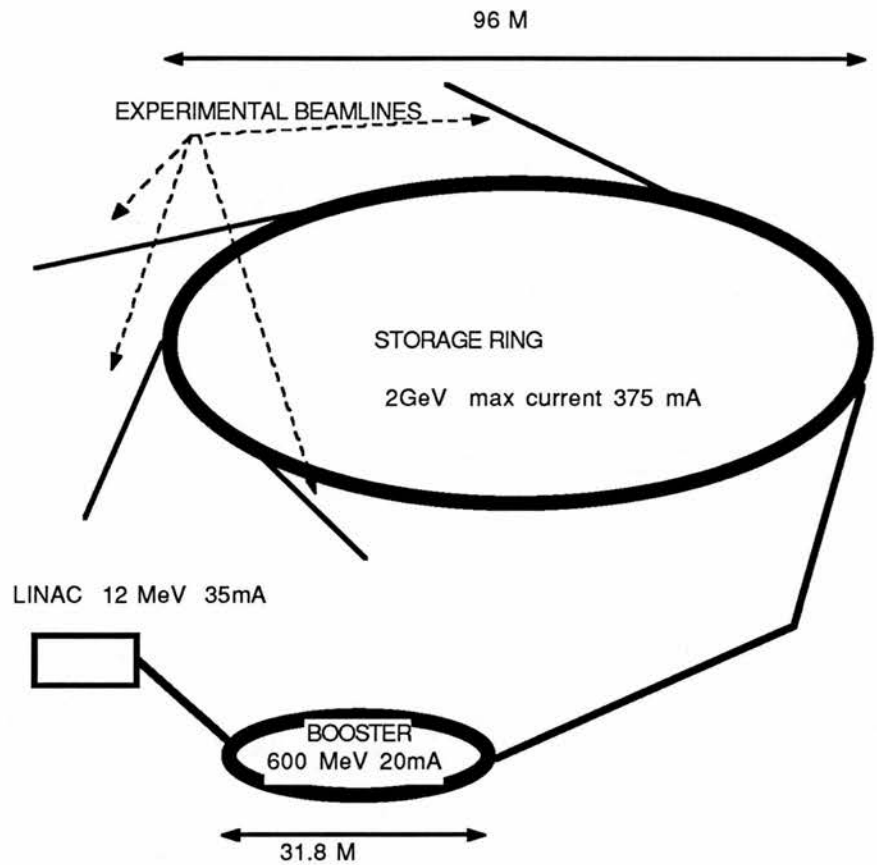


Figure 3.1. the S.E.R.C. synchrotron, Daresbury.

3.1.2 BEAMLINE AND CAMERA REQUIREMENTS

It is important to control a number of factors in the experimental process of X-ray diffraction, in order to optimise the quality and interpretability of data. The way in which this is attained is controlled by the beamline and camera design.

Two important factors affecting the quality of the diffraction pattern are beam size and beam scatter.

3.1.2.1 Beam size

Beam size has a profound effect on the resolution of the diffraction pattern, every reciprocal lattice point being smeared over an area approximately equal to the beam cross-section.

Large molecules such as collagen require high resolution to separate diffraction maxima that are located very close together in reciprocal space.

A high flux is required for samples that scatter weakly, and a small beam size is required for the reasons described above. These two requirements do not seem compatible since a reduction in beam cross-sectional area implies a reduction in flux, however one solution to this problem is to focus the X-ray beam onto the film plane. This can be achieved using a mirror and monochromator.

The monochromator usually consists of a single crystal cut at an angle to a specific set of crystal planes so that the incident radiation is diffracted by these crystal planes. The diffracted beam is effectively

monochromatic, since only the radiation of a specific wavelength can be diffracted by the crystal planes selected. The monochromator can be bent to produce a degree of focussing. The effect of this is to reduce the beamsize, flux, and radiation other than the wavelength selected.

The mirror consists of glass, it may be covered with a thin layer of metal (such as gold, nickel or platinum) to help reflectivity. It reflects the X-rays at a glancing angle smaller than the critical angle θc. The critical angle is proportional to the wavelength of incident radiation, therefore the position of the mirror can be adjusted to a position where the required wavelength of radiation is reflected, and the higher wavelengths eliminated. The mirror, like the monochromator can be bent to provide a degree of focussing.

It is essential to reduce background noise on the film caused by scattering of the beam from the sample holder and disordered parts of the specimen. Using a focussed beam is beneficial in this respect, since the intensity of randomly scattered rays decays with the square of the distance from the specimen. The intensity of the focussed diffraction pattern decreases proportionally with distance.

Consequently, at a fixed distance the contribution of scattered background intensity is decreased relative to that of the diffraction maxima.

3.1.2.2 Reduction of parasitic scatter

Various types of unwanted X-ray scatter are prevalent in X-ray cameras.

The first is scatter from gas molecules in the path of the beam, this has an effect of reducing the beam flux, and of blackening the central portion of the film. This may be reduced by the removal of air from as much of the beam path as possible by evacuation of the camera or replacement of the air by an electron sparse gas such as helium.

Another important source of scatter is inherent defects in the optical elements of the camera. This scattering can be decreased by the introduction of vertical and horizontal slits on either side of the optical components. These slits are positioned so as to allow the main beam to pass whilst attenuating the parasitic rays.

Only a small proportion of the incident X-ray beam is absorbed or diffracted by the sample. The through beam, is therefore of higher intensity than the diffraction peaks. The scattering of the through beam on the film surface and on the back of the camera results in blackening of the film. This can be ameliorated by the use of a small lead beam stop. The beam stop is cup shaped and catches the undiffracted beam and scattering from the beam stop surface.

3.1.3 BEAMLINES

The synchrotron storage ring provides electromagnetic radiation which is released down beam lines. Beamlines are situated at ports around the circumference of the storage ring where radiation is channeled to individual experimental stations. The purpose of this is to provide the radiation of correct wavelength for varied types of research. In the case of X-rays a number of beamlines at the Daresbury Synchrotron are dedicated to selecting a narrow range of radiation in the region corresponding to the Cu Kα peak of conventional X-ray sources. This being 1.5406Å.

Amongst these beamlines are those designated 2.1, 8.2, and 7.2 N.C.D. (Non Crystalline Diffraction). These beamlines each serve their role and have features which make them suitable for certain fields of X-ray experimentation.

The nature of the axial periodicity of collagen is such that the first order of diffraction can only be obtained using a camera of sufficient length to successfully separate it from the main beam. The length of camera capable of recording the first order must be at least 1.5 metres. The high degree of axial order in tendon, results in the diffraction pattern containing up to 100 observable Fourier components.

It has been found to be impracticle to attempt to obtain all of the Fourier components in one dataset. The first 17 orders are collected on small angle diffraction stations 2.1 or 8.2, and the orders 7 to over 50 are collected at the N.C.D. station 7.2.

3.1.3.1 Beamline 8.2

The optical instrumentation of beamline 8.2 consists of the following:

(1) MONOCHROMATOR. This is a bent triangularly shaped, flat, Ge crystal asymmetrically cut at 10.5° to the (1,1,1) plane. The monochromator is situated 20m from the storage ring exit port. The effect of this is to provide monochromatic X-ray radiation of 1.418Å and some degree of horizontal focus. The horizontal beam size is 0.3mm before the monochromator and 0.1mm after the monochromator.

(2) MIRROR. The mirror is 750mm long. It is bent, flat and uncoated. It is 22.2m from the storage ring and deflects radiation at a glancing angle of 4mrad providing at least a ten fold decrease in vertical beam size to less than 0.25mm.

The beam position is monitored by a scanning ion chamber and a two wire photoelectron detector.

The intensity at the sample is about 10^{11} photons per second with the storage ring operating at 200mA and 2GeV. The beamline 8.2 is shown in Figure 3.2.

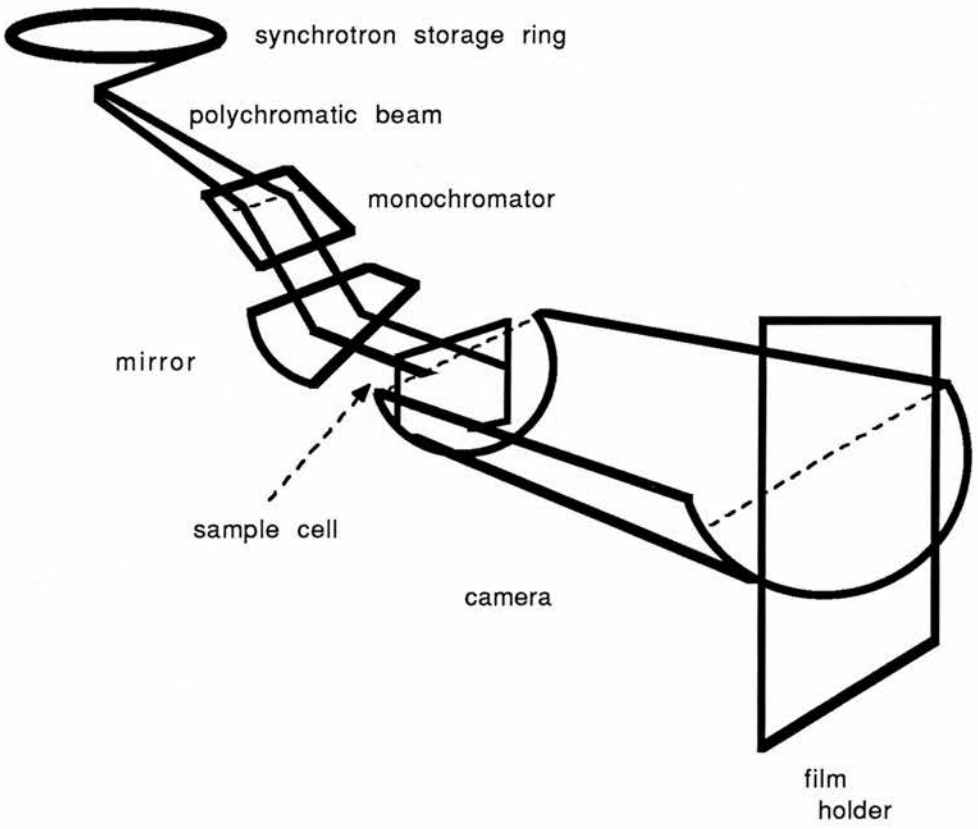


Figure 3.2, a diagrammatic representation of beamline 8.2.

3.1.4 CAMERAS

In this case the cameras are relatively simple. For small angle experiments, the sample is placed in front of a mica window. The body of the camera consists of an aluminium tube sealed at the other end by X-ray transparent mylar. Data is collected by use of a detector or a film pack to record the diffraction pattern. The main beam is stopped by a lead cup or beam stop, as described earlier, which is shaped so as to absorb the X-rays and stop them being scattered onto the detector or film. Air scatter, which produces a clouding on X-ray films around the main beam, is diminished by evacuation of the camera.

For higher angle diffraction experiments the camera on beam station 7.2 is used. This camera contains the collimator which reduces the size of the X-ray beam cross-sectional area. The camera length in this case is 240mm from collimator to film pack. The body of the camera is made of brass and has an aluminium extension, the whole system being flushed through with helium to reduce the X-ray scatter.

An evacuated camera is less practical in this case, since the sample is actually placed inside the camera and the sample cell is not vacuum proof.

3.1.5 SAMPLE CELL

One sample cell was designed for use in all of the X-ray experiments and its purpose was to:

(1) Allow the sample to be positioned correctly in the X-ray beam.

(2) Maintain the sample in a hydrated state.

(3) Remove the inherent crimp which exists in rat tail tendon.

The reason for the first of these requirements is self evident. The second, that a constant humidity should be maintained, is not so evident. The reason is that the intensity of the meridional and off meridional reflections varies with the humidity around the fibres. A constant relative humidity is attained by suspension of the sample over a salt solution of known concentration, and by keeping the external temperature around the cell constant. The reason for the third requirement, is that failure to remove the 200μm periodic crimp in rat tail tendon results in disorientation of X-ray diffraction patterns. This is due to the incident beam diffracting from fibrils in different orientations. The application of 4% stretch to the tendon removes the crimp and aligns the fibrils along the fibre and removes the disorientation effect. This makes data reduction easier, especially in the quantification of off meridional reflections. A sample cell was designed to meet the requirements outlined and a diagram of this is shown in Figure 3.3.

The features of this cell are:

(1) A large mylar window, which allows up to 20mm lengths of the tendon sample to be used in diffraction experiments.

(2) The "O" ring is present to allow the cell to be water tight and therefore permit the sample to be suspended above a saline solution maintaining the correct humidity of air around the sample.

(3) The internal tensioning system, which allows the sample to be clamped between the teflon washers, is present to allow tensioning of the sample by screwing the clamped washer until the crimp is removed.

(4) The tendon is suspended between rollers, the relatively large length of suspension allowing the twisting of the tendon to be minimised.

The sample cell, when loaded, is fitted to a goniometer head, the axis of vertical movement being colinear with the tendon length. The attachment of the sample cell to a goniometer allows precise alignment of the sample in the beam.

In X-ray diffraction the sample may be damaged due to incident radiation. This can impose a limit to the amount of X-ray flux that a sample can be exposed to, before beam damage decreases the quality of the diffraction pattern. The ionising nature of X-rays causes free radicals to be produced in the sample.
These can cause crosslinking between molecules in the sample, and also can disrupt the polypeptide backbone. These effects both cause the structure of the sample to deviate from that in vivo.

To produce X-ray diffraction patterns of tendon as shown in Plates 4.1-4.8, the exposure time had to be over one hour using synchrotron radiation. This length of exposure time at a high flux of radiation, causes damage to the sample.

A stepper motor was connected to the vertical drive of the goniometer and a programme "SAMPSHIFT"* was written for a Commodore Pet computer to control the stepper motor. This allowed the sample to be oscillated in the vertical, colinear with the sample. Therefore the effect of beam damage could be spread over a larger length of the tendon, and minimise the effect of beam damage at any one point. The length of oscillation was 7mm, this allowed the length of exposure in the beam to be increased by a factor of approximately 5 before beam damage could be recognised, by the deterioration of the off meridional Bragg peak scattering compared to background.

*Programme "SAMPSHIFT" written by J.P. Bradshaw.

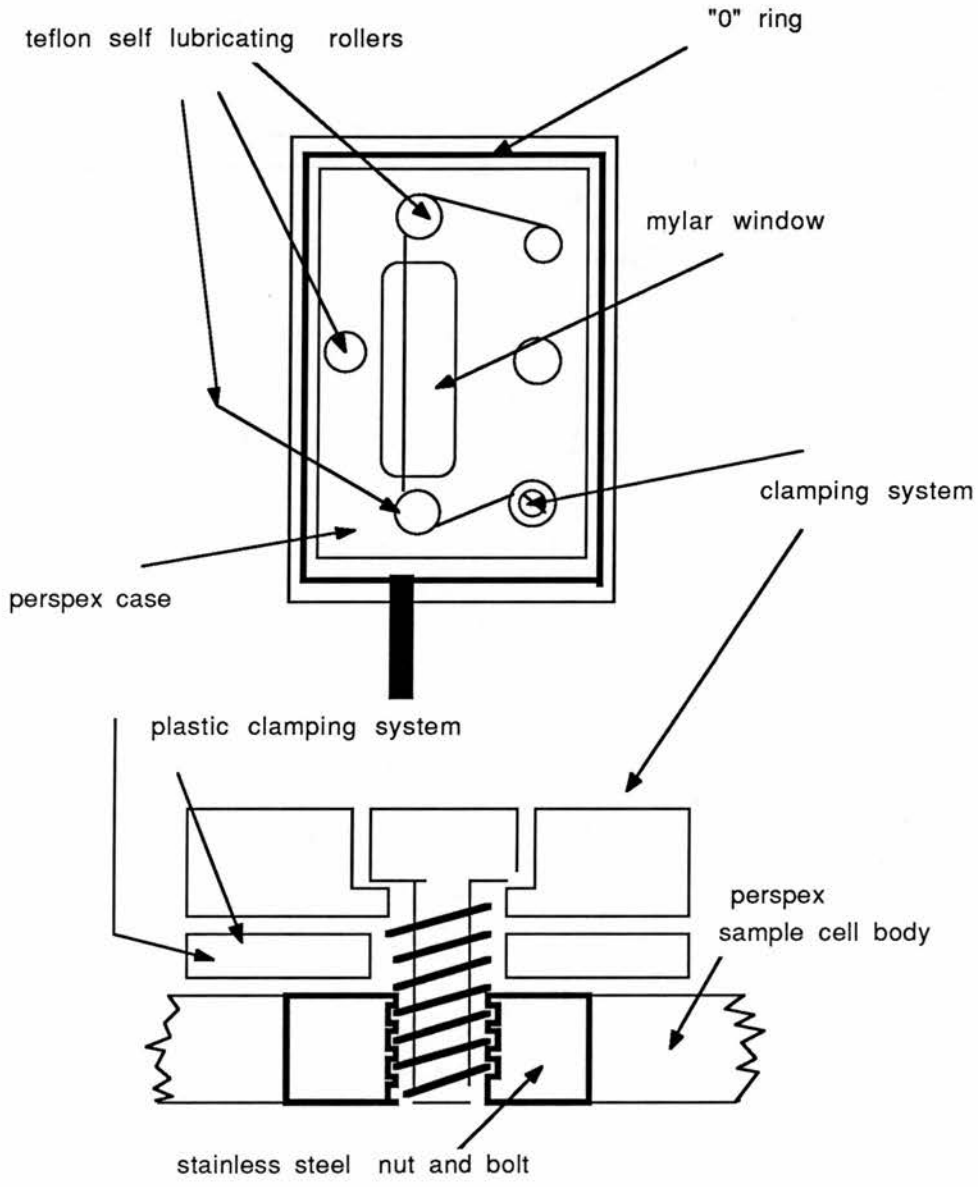


Figure 3.3, sample cell for X-ray diffraction.

3.2 DATA COLLECTION AND PROCESSING

The accuracy of a structure determined by X-ray or neutron diffraction depends on the accuracy with which the data are obtained. This encompasses a large range of topics from the quality and repeatability of sample preparation, to the method of data collection and subsequent interpretation and correction.

3.2.1 COLLECTION OF DATA

There are two main options available for the collection of X-ray data. One is the use of X-ray sensitive film, and the other is the use of electronic detectors. Both of these methods have advantages and disadvantages.

Electronic data collection is superior in terms of speed and ease of data processing, as the amount of data collected can be viewed as a function of time, and stopped when sufficient data has been judged to have been collected. The data is then readily available in digitised form and can be easily processed by suitable software.

A linear detector was used in data collection for this thesis, this can only collect a small part of the total diffraction pattern. For example, the detector collects a small section through each of the meridional reflections. The detector collects data in 1024 channels, this defines the amount of detail available.

Data collection on film has the disadvantage that the amount of data collected can only be judged by an initial process of trial and error. The film used here was Ceaverken reflex 25. The dynamic range of meridional intensities requires a pack of films (usually 5) to be used in each data collection. Each film pack gives a range of exposure values. The more intense diffraction peaks can be estimated from those films furthest away from the sample without saturation, whilst the less intense diffraction maxima can be observed with accuracy from those films on the top of the film pack.

The advantage of using film is that although the meridional diffraction peaks from collagen are essentially a one dimensional series, disorder and the shear in the triclinic unit cell spreads the intensity over an area perpendicular to the meridional series. Far more accurate results can be obtained if the intensities are integrated across the full width of the peaks.

The degree of isomorphism of the diffracting structure can also be judged more easily if a two dimensional diffraction image can be seen. Since in this case isomorphism has been judged by the maintenance of lattice parameters, this requires the off meridional peak position to be measured.

3.2.2 DATA PROCESSING

The conversion of diffraction data obtained on film to digital form, requires the films to be scanned on a Joyce Loebl Scandig 3 scanner. The raster size was selected as 50μm and therefore the optical density of each film could be recorded at pre-selected intervals on a raster system. The optical density range used on the system was 0-2 units and thus each point within the range was assigned a density grey value between 0 and 255. This information was then stored on magnetic tape and subsequently downloaded onto a VAX computer (DEC).

3.3 REACTOR NEUTRON SOURCES

Neutrons used in diffraction have wavelengths of the order 0.1-1nm. These are generated by the use of medium flux (10-20 megaWatts) or high flux (60-100 megaWatts) nuclear reactors. Neutrons are produced as the result of U^{235} fission in a reactor core. These are characteristically high energy reactions. The energy associated with the neutron is attenuated by collision as they pass from the core into a moderator such as D_2O or graphite at 300°C. Further moderation using a cold source such as liquid deuterium at 25°K enhances the longer wavelengths useful for diffraction.

Neutron guides then deliver neutrons from the moderator to externally located instruments, the first section of the guide being bent to eliminate gamma rays and fast neutrons.

A velocity selector, based on a rotating drum with a helical slit in it, monochromates the Maxwellian distribution of neutron wavelength (Perkins 1988). It minimises flux loss by permitting a narrow range of neutron wavelength with the correct velocity to travel down the helical path.

All neutron diffraction experiments described in this thesis were conducted at the Institute Laue Langevin in Grenoble.

3.3.1 NEUTRON CAMERAS

Neutron diffraction is similar to X-ray diffraction in that the collection of a diffraction pattern requires the use of a camera. The method of monochromation is described in the previous section (3.3).

The focussing of neutrons is impossible since the interaction of neutrons, at glancing angles, with mirror surfaces does not occur in the way found in X-ray optics. Beam definition therefore primarily results from collimation.

In contrast to the fine definition of a synchrotron X-ray beam, the neutron diffraction process uses a wide neutron beam of cross-sectional area > $1cm^2$. This requires a large sample surface area for diffraction and long camera lengths to resolve the peaks. The size of the beam is however necessary, due to the relatively low flux of neutrons compared to the flux of photons in a synchrotron.

Two instruments were used for the purpose of neutron diffraction and these were designated D11 and D16.

3.3.1.1 Instrument D11

Instrument D11 is a low angle neutron diffraction apparatus. The neutron beam passes down a neutron guide to the rotating monochromator, see Figure 3.4. The large beam size can lead to beam divergence errors and the collimation distance is varied by moveable segments of the neutron guides to optimise the desired collimation.

The beam area on the sample is adjusted using circular cadmium diaphrams which prevent neutrons scattering from the sample cell as opposed to the sample.

The detector assembly is positioned in vacuo in a large tube and ranges of sample to detector distances of 1 to 40 metres are possible. Neutrons are detected with BF_3 gas in an ionisation chamber (Bacon 1975).

The detector consists of a 64 x 64 array of $1cm^2$ individual neutron sensitive plates and this is able to record a two dimensional diffraction image.

3.3.1.2 Instrument D16

This instrument is a 4-circle diffractometer. The detector consists of a single channel, this is moved relative to the sample in an arc at a fixed distance. The information obtained is therefore a linear trace through the meridional reflections. The diffractometer is able to execute $\omega, 2\theta$ scans. This is performed by the

tilting of the sample so that the reciprocal lattice cuts the Ewald sphere in turn for each reflection.

The detector is positioned at the correct position to record diffraction from each reflection as it is stepped through the Ewald sphere.

3.3.2 SAMPLE CELLS FOR NEUTRON DIFFRACTION

The interaction of neutrons with biological samples, the flux of beamlines and the size of the beam dictate that the size of the sample must be larger than that used in X-ray diffraction, to produce satisfactory results. The sample cell for beamline D11 is therefore able to hold up to 1g of tendon.

It consists of a teflon body with large quartz windows, 15mm diameter, allowing the large collimated beam to interact with the tendon and not the sample cell body. The tendon is maintained in an aqueous medium to maintain the correct conformation.

In the case of D16, the sample cell consists of a frame of dimensions 30mm x 30mm to support the tendons (up to 1g) the tendon is maintained at the correct humidity by the frame being enclosed in an aluminium can, containing a trough for saline solution. The frame is attached to a base which has the ability to rotate for the reasons given in 3.3.1.2.

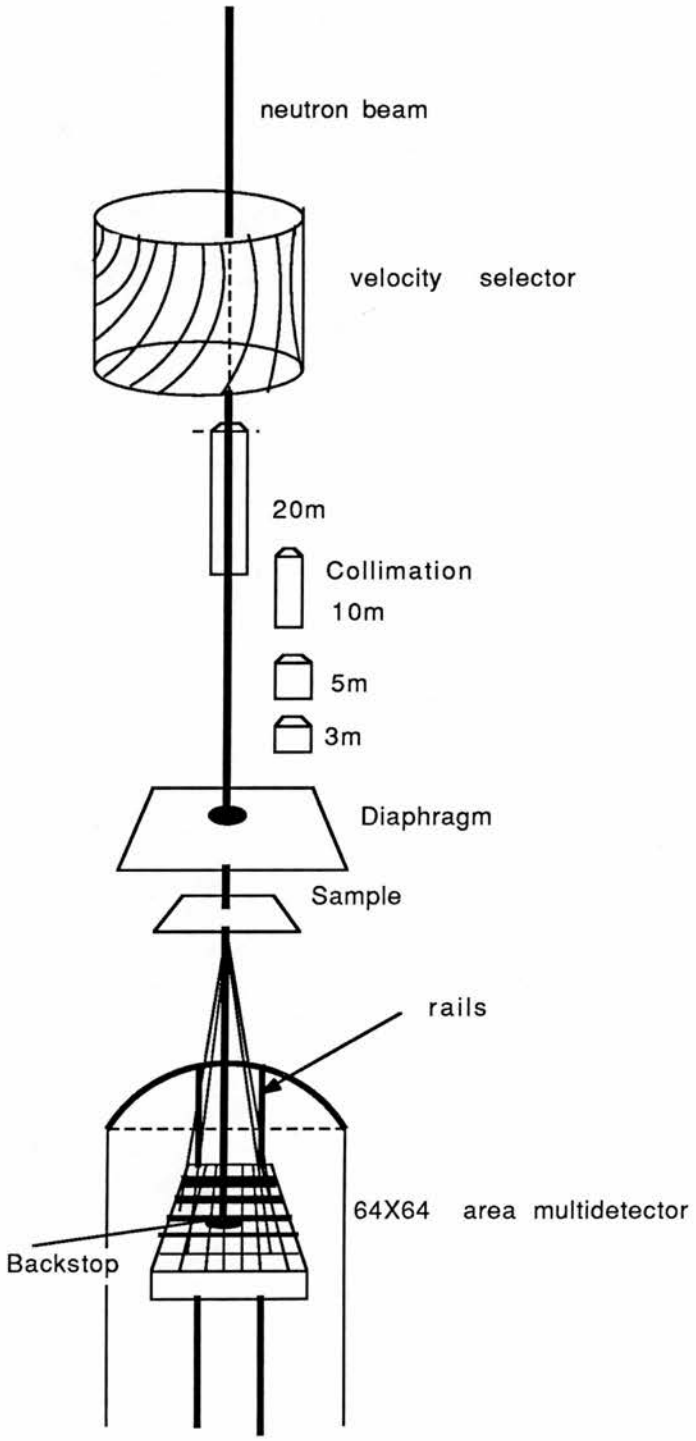


Figure 3.4, the low angle instrument D11

at the I.L.L. Grenoble.

CHAPTER 4

THE AXIAL ELECTRON DENSITY OF RAT TAIL TENDON

4.1 INTRODUCTION

The one dimensional structure, or axial projection, of wet collagen fibrils in rat tail tendon has been studied using a variety of physical techniques, principally electron microscopy, X-ray diffraction and neutron diffraction. Each has contributed to suggesting a structure in which the collagen molecules within the tendon fibrils are staggered, giving a fundamental periodicity of 670Å. This so-called D-repeat consisting of a "gap" and an "overlap" region, was first proposed by Hodge & Petruska (1963).

Each technique has inherent advantages and disadvantages. Electron microscopy allows the periodicities within the fibrils to be visualised by staining; the patterns produced have been correlated to type I collagen amino acid sequence information (Doyle et al, 1975). However, the resolution of the structural information provided by electron microscopy is only about 30Å, and the process of sample preparation may produce artifacts. In neutron diffraction the low scattering cross sections and the constraints of long exposure times and large samples, brought about by the low flux rates, can be a problem with biological material.

X-ray diffraction provides the potential for high resolution studies, and the advent of synchrotron radiation has made possible shorter exposure times and reduced radiation damage to the specimen. However, as in

the case of neutron diffraction, the data gathered constitutes only a part of the total information contained in the diffracted beams. The lost information consists of the phase angles of the diffraction maxima. The potential of X-ray diffraction to provide an accurate electron density distribution of the tendon fibrils makes the search for a solution to this "phase problem" worthwhile.

4.2 PREVIOUS X-RAY WORK

In X-ray diffraction the meridional reflections contain information about the one dimensional axial distribution of electron density within a single D-repeat of the collagen fibril. Up to one hundred meridional reflections have been observed. The amplitudes of these reflections (when each is combined with its corresponding phase angle) provide the set of structure factors which may be related by Fourier transformation to the electron density distribution. A number of attempts have been made to produce such an axial electron density map; these attempts may be discussed in two categories.

One strategy has been to calculate a model of the electron density distribution on the basis of the known amino acid sequence, amino acid axial spacing and the Hodge Petruska scheme. A variety of such models may be calculated in this manner before being Fourier inverted to determine which model fits the experimental data best.

It is clear that this approach, as used by Hulmes et al, (1977), is model dependent, and need not necessarily produce a unique and unambiguous solution.

The other main line of approach has been to determine the phase angles directly by using an experimental method such as isomorphous addition or replacement. This has been attempted by a number of groups with varying degrees of success (Ellis & McGavin, 1970, Kaesburg & Shurmann, 1953, Tomlin & Worthington, 1956, Stinson et al, 1979). However, assumptions of centrosymmetry, problems with scaling, lack of resolution and choice of stain have meant that the interpretation of the data, and the subsequent phase determination, were not conclusive.

4.3 AIMS

The approach of isomorphous addition was used with reagents selected using criteria of chemical reactivity. Stains were therefore used which labelled fewer sites in collagen than those used previously. Using this approach with data collected by synchrotron radiation has made possible the experimental determination of the phases of the first fifty meridional diffraction maxima of rat tail tendon.

4.4 THE SEARCH FOR SUITABLE DERIVATIVE STAINS

Previous approaches to the production of isomorphous derivatives of collagen frequently give the appearance of having been little more than a random "trial and error" search among the heavy atom reagents which have proved successful in crystallographic studies. An alternative approach has been to use electron microscopy stains, such as phosphotungstic acid and uranyl acetate. These stains are known to label at a large number of sites on the collagen molecule, principally the charged residues. This large number of potential binding sites introduces the potential for mixed and partial site occupation, complicates the data interpretation and can lead to a lack of isomorphism by disrupting the ionic interactions between charged side chains.

The chemical nature of the amino acid side chains of collagen lends itself to much more selective forms of labelling. Type I collagen contains a relatively small number of histidine, methionine and tyrosine residues which can be labelled by ligand exchange or covalent interaction. Suitable reagents were chosen after due consideration of the theory of heavy atom labelling and ligand chemistry, the success of crystallographic studies using these reagents, and their characteristic size and stability.

The search for suitable reagents was confined to those thought to react with the amino acids above, since

it was considered that they could be labelled more specifically, and to a higher degree of stability than the more abundant charged side chains of lysine, arginine, glutamate and aspartate.

The iodination of proteins has been extensively studied: it primarily seems to involve the formation of a covalent link with tyrosine residues, although the formation of iodohistidine has also been reported. Collagen has been iodinated by a number of standard techniques (Haworth & Chapman, 1977, Bernstein & Mechanic, 1980).

It has been reported that the $PtCl_4^{2-}$ ion and related compounds undergo a ligand exchange reaction with a number of sites in proteins, especially the -S-H group of cysteine, the free N-terminal, the $-S-CH_3$ group of methionine and the imidazole group of histidine (for reviews see Petsko et al, 1978 and Thomson et al, 1972). Of these four groups, only the last two are found in type I collagen. $K_2^+PtCl_4^{2-}$ has been used to produce isomorphous derivatives of a number of proteins, and collagen has been labelled for electron microscopy using chloroglycyl-L-methioninatoplatinum (II) (Pt-GLM) (Kühn et al, 1984).

Gold, in the form of $Na^+AuCl_4^-$, is thought to react with protein sidechains in a similar manner to the $PtCl_4^{2-}$ ion, but the nature of its reactivity, and its preferred sidechains are in doubt. NMR studies have shown that the interaction with free methionine is strong

(Bordignon et al, 1973), and the literature shows that in crystallographic studies the labelling sites have included histidine and possibly an ionic interaction with lysine (Sadler 1976).

A number of other heavy atom complexes have been used in crystallographic studies. The interactions between these complexes and proteins are principally through ligand exchange and ionic interactions. Of these available complexes, a number were selected that were reported to label at a limited number of sites (Blundell & Johnson 1976). The specific site of interaction with proteins in the case of many complexes such as $K^{+}AuCN_2^{-}$, $Na_3^{+}IrCl_6^{3-}$ and $K_2^{+}PdCl_4^{2-}$ is less well resolved than the heavy atoms discussed above.

The covalent attachment of chelators to proteins has been postulated as a method of producing isomorphous derivatives. Proteins can be made to chelate strongly various heavy metal ions very specifically. The binding of In^{3+} to proteins using this technique has been reported by Meares et al, (1976), Krejcarek & Tucker (1977), Sundberg et al, (1974).

4.5.EXPERIMENTAL

Suitable reaction conditions for each of the reagents were investigated, with the usual isomorphous derivative requirements of few binding sites and reproducibility together with true isomorphism.
The parameters of reagent concentration, reaction time, temperature and pH were varied for each of the reagents. Diffraction data were collected from treated collagen fibres.

4.5.1 CRITERIA FOR ISOMORPHISM IN COLLAGEN

The suitability of each set of reaction conditions was assessed in terms of the changes produced in the diffraction patterns. Reagents which altered the relative intensities of the meridional peaks, whilst retaining the spatial distribution of all the diffraction peaks, were chosen for further examination using synchrotron radiation.

Although the whole area of each diffraction pattern, out to c.7Å (meridional, equatorial and near equatorial peaks, 9.6Å layer line and 38.5Å and 26Å row lines), was examined to determine whether or not each set of structures were isomorphous, it was quickly realised that a good first indication was given by the equatorial and near-equatorial region which was rapidly disrupted by many of the reaction conditions tried. This region of the diffraction pattern describes the lateral packing of

collagen molecules into quasi-crystalline regions of the fibril, and has been reported to be particularly susceptible to pH and ionic strength (Jesior et al, 1980, Svendsen & Koch 1982).

4.5.2 DETERMINATION OF REACTION CONDITIONS

The optimum reaction conditions were investigated for each prospective derivative. In each case reaction conditions were investigated with 0.1g rat tail tendon (wet weight) in a final reaction volume of 10mls.

For the platinum derivative, samples of tendon were placed in reagent solution consisting of Ringer solution, ($9gl^{-1}$ NaCl, $0.45gl^{-1}$ KCl and $0.24gl^{-1}$ $CaCl_2$) borate buffer (5mMol) and $K_2{}^+PtCl_4{}^{2-}$ ranging from 1:1 to 50:1 mol reagent : mol methionine in collagen, all reactions took place at 20°C (room temperature).

A variety of reaction times (15 minutes to 48 hours, logarithmic scale) and pH values (pH 5 to 9, 0.1 step size) were attempted. After analysis of the resultant X-ray diffraction patterns, using the criteria described, it was determined that a suitable set of reaction conditions were, 2mgs $K_2{}^+PtCl_4{}^{2-}$ to 0.1g tendon at pH 7.5 for a minimum of two hours.

A similar experimental protocol was employed in the production of a gold derivative, this gave the optimum reaction conditions to be, 2mgs $Na^+AuCl_4{}^-$ at pH 7.5 for a minimum of two hours.

In the case of producing an iodinated derivative, a number of iodination techniques were investigated, these fall into three categories:

(1) Reaction with saturated iodine solution.

This method uses iodine in Ringer solution, where dissociation of the I_2 molecule to I_3^- and I^+ provides the reacting species. A saturated iodine solution was prepared and diluted to produce a series of solutions ranging from 10% to 100% strength of the original. The pH used was in the range 5 to 8 and a variety of labelling times were used to label 0.1g of rat tail tendon. Although the labelling technique did produce changes in the intensities, thus suggesting isomorphism, the shape of the meridional reflections was also altered, indicating a loss of the 4° shear in the unit cell. This was concurrent with alterations and weakening of the off meridional and equatorial diffraction pattern.

For true isomorphism, it was judged that this method was unsuitable for the high degree of structural resolution required.

(2) Iodination with iodine monochloride.

This method uses an oxidising agent to cause dissociation of I_2, and was first used for iodination of protein by Macfarlane (1958). Iodine monochloride solution was prepared and converted to hypoiodite by the addition of NaOH to raise the pH. Iodine monochloride

solution was prepared as a stock, see Bale (1966), this could be diluted to a suitable strength for a 1:1 to 1:50 mol tyrosine in collagen : hypoiodite.

The reaction with rat tail tendon was conducted at pH 8 using a variety of reaction times. This reaction technique did not provide a suitable isomorphous derivative, since it altered the position of reflections as for the method described in (1), without altering the intensities as significantly as the third method of iodination.

(3) Iodination with KI and Chloramine T.

This method had already been used for the iodination of collagen by Haworth & Chapman (1977), where paper chromatography of alkaline hydrolysates of iodinated collagen showed full iodination of the tyrosine residues with no trace of histidine iodination. In this case no I_2 was added since the chloramine T readily liberates I_2 as it reacts with KI in solution.

An initial "hanging drop" method of iodination was employed, since it was noted (Haworth & Chapman 1977) that prolonged exposure to chloramine T had a denaturing effect on collagen.

In this method, the rat tail tendon was placed in a glass vial and covered with a glass microscope slide, the underneath of the slide held a drop of chloramine T solution (10mgs ml^{-1}). The resulting oxidant, Cl_2, would then enter the solution by diffusion without the

chloramine T reacting with the collagen. Although a number of attempts were made using this method with different strengths of KI, the procedure was unsuccessful and a method using the addition of chloramine T to the reaction mixture was investigated. As for the case of Pt and Au labelling, a number of reagent concentrations, pH conditions and reaction times were attempted.
These produced a series of very successful methods, and resulted in optimum reaction conditions covering a broad range of concentrations at pH 7.5.

The reaction conditions employed in further studies were 2mgs KI, 3mgs chloramine T in buffered Ringer solution with 0.1g rat tail tendon for a minimum of one hour. The effect of chloramine T was investigated by the omission of KI to the reagent. The result of this being that it made no difference at all to the intensity or quality of meridional, off equatorial or equatorial reflections.

Of the other attempts to produce derivatives, many compounds such as $K^{+}AuCN_2^{-}$, $K_2^{+}Pd(NH_3)_2Cl_4^{2-}$, cis-platinum and $Na_3^{+}IrCl_6^{3-}$ produced diffraction patterns which were isomorphous, but either resembled the change in intensity produced by $K_2^{+}PtCl_4^{2-}$ or produced intensity differences that were uninterpretable, since their labelling position could not easily be determined.

The production of a collagen able to chelate indium was investigated using a method of attaching diethylene-triaminepentaacetic acid (DTPA) to collagen, using the

commercially available cyclic anhydride of this compound. The method of Hnatowich et al, (1982) was used, this resulted in the fibres being significantly weakened, and made mounting the sample in the sample cell very difficult. This method was therefore abandoned since it was regarded that such macroscopic alterations to the properties of the fibril would not produce a successful derivative.

After each type of labelling procedure the tendon samples were washed thoroughly with buffered Ringer solution to remove any non-specific stain from the sample. It was found that the $Na^{+}AuCl^{4-}$ labelled sample required careful washing since if this was not conducted, a series of thin lines appeared on the diffraction pattern between the meridional and 38Å row lines. This was interpreted as excessive gold stain accumulating around fibrils.

In every case, the reagent solutions were made and used as soon as possible. The hydration chemistry of $K_2^{+}PtCl_4^{2-}$ is known to proceed by ligand substitution of H_2O for Cl (Thomson 1974), as does $K^{+}AuCl_4^{-}$, the effect of ligand exchange of $K_2^{+}PtCl_4^{2-}$ with NH_3 of $(NH_4^{+})_2SO_4^{2-}$ has been shown to affect labelling studies (Petsko 1978). Therefore to get reproducible results and minimise this parameter, the above procedure was used.

4.5.3 DATA COLLECTION

Low angle meridional diffraction patterns were collected on two focussed beam X-ray cameras using Synchrotron radiation at the S.E.R.C. Synchrotron Radiation Source, Daresbury, England. Of these, XR2.1, and 8.2 with a specimen to film length of 2000mm, were used to collect the lower orders of diffraction from the D-repeat (1-17). The other, XR7.2NCD, was a shorter camera of 200mm focal length and was used for the higher orders (6-50+). These, and other, small angle diffraction facilities at the S.R.S. have been described in chapter 3. Each exposure was recorded on a pack of five films (Ceaverken reflex 25) to ensure that the complete range of intensities could be measured.

After processing, the intensity data of the films were digitised by scanning on a Joyce Loebl Scandig 3 microdensitometer with a raster setting of 50μm for the long camera films and 100μm for those exposed in the shorter camera. The total intensity of each reflection was corrected for the following systematic errors: absorption, specimen to film distance, polarisation (Whittaker 1952), disorientation, intersection with the Ewald sphere and beam convergence (Wang & Worthington 1975). These procedures are described fully by White (1977). Much of this work was carried out using the programme "GUCKMAL" on the Daresbury SASVAX DEC computer.

GUCKMAL was written by Daresbury computing laboratories, and was employed in this case since it

allows the user to draw a box around the perimeter of each reflection, and integrate the intensity found within the box whilst subtracting a background for that area. The background is determined by using values at the perimeter or by drawing a box adjacent, in this case off the meridian, to determine a local background count. The data obtained could then be collected on the Daresbury laboratory VAX (DEC), DLVE.

In order to determine the effect of each label on the collagen, and to establish whether or not a derivative is truly isomorphous, it is essential to determine the relative intensities of the meridional diffraction peaks from different samples: in other words it is necessary to scale data from different samples to each other. To do this, an additional set of these data was collected from the first 9 meridional diffraction maxima, using a position sensitive X-ray detector, before and after stain treatment in situ. Ion chambers, placed before and after the specimen, enabled this data to be normalised to allow the scaling of diffraction patterns from each stain treatment to each other and to the native pattern. The diffraction patterns of the native and derivative samples for high and low angle diffraction respectively are shown in Plates 4.1 to 4.8.

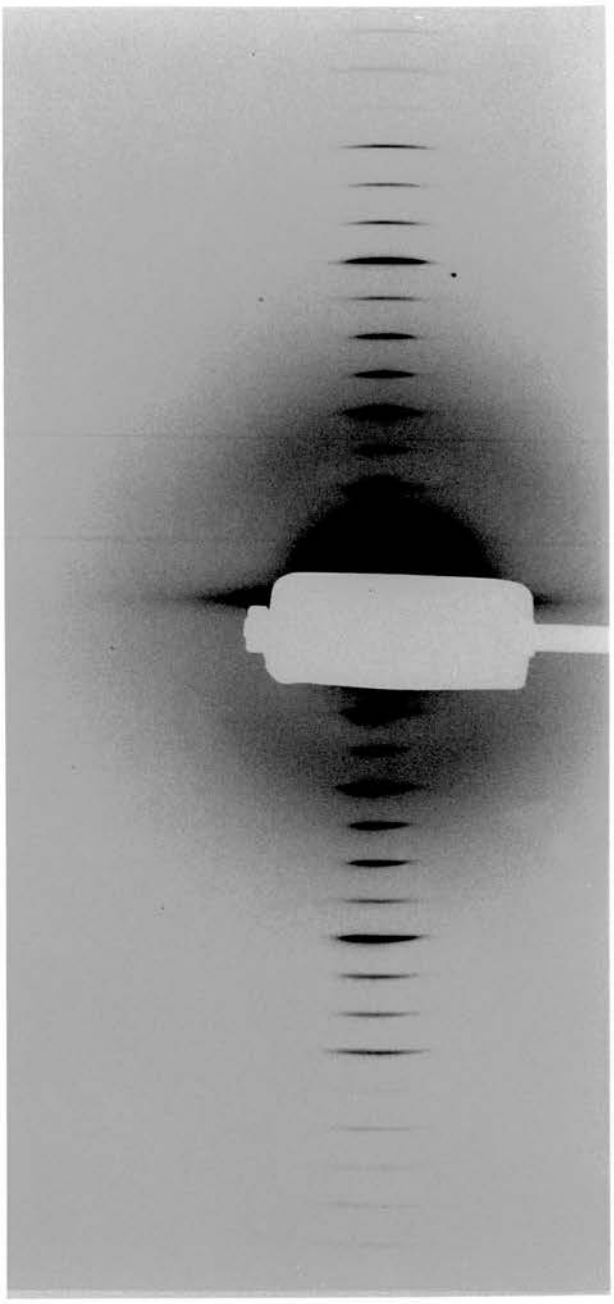


Plate 4.1 small angle diffraction pattern of native rat tail tendon. Taken on beamline 8.2 at the S.R.S., Daresbury.

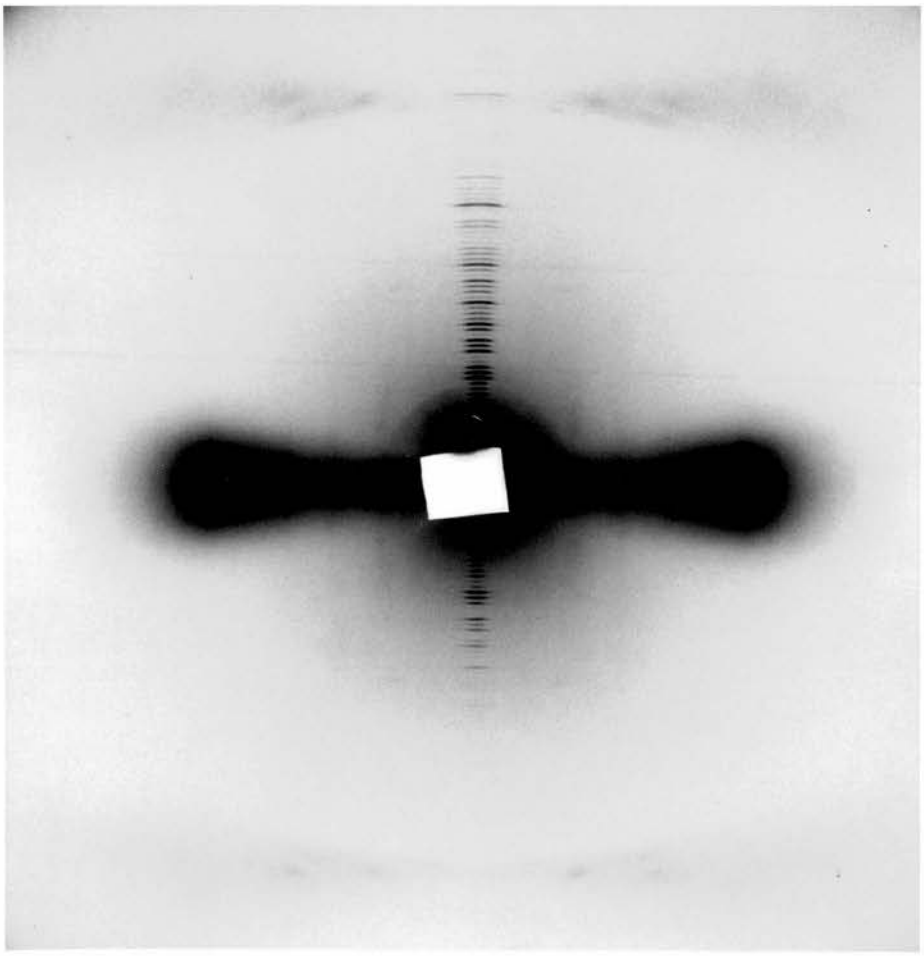


Plate 4.2 higher angle diffraction pattern of native rat tail tendon. Taken on beamline 7.2 at the S.R.S., Daresbury.

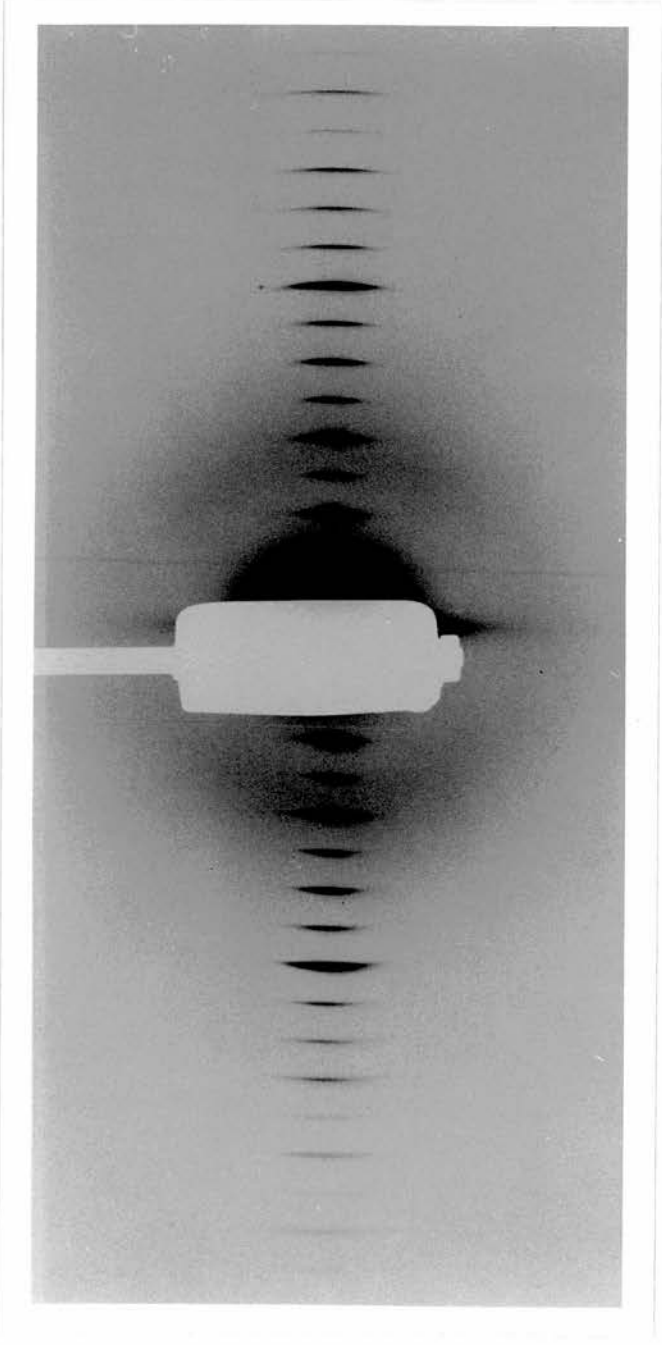


Plate 4.3 small angle diffraction pattern of rat tail tendon iodinated using chloramine T. Taken on beamline 8.2 at the S.R.S., Daresbury.

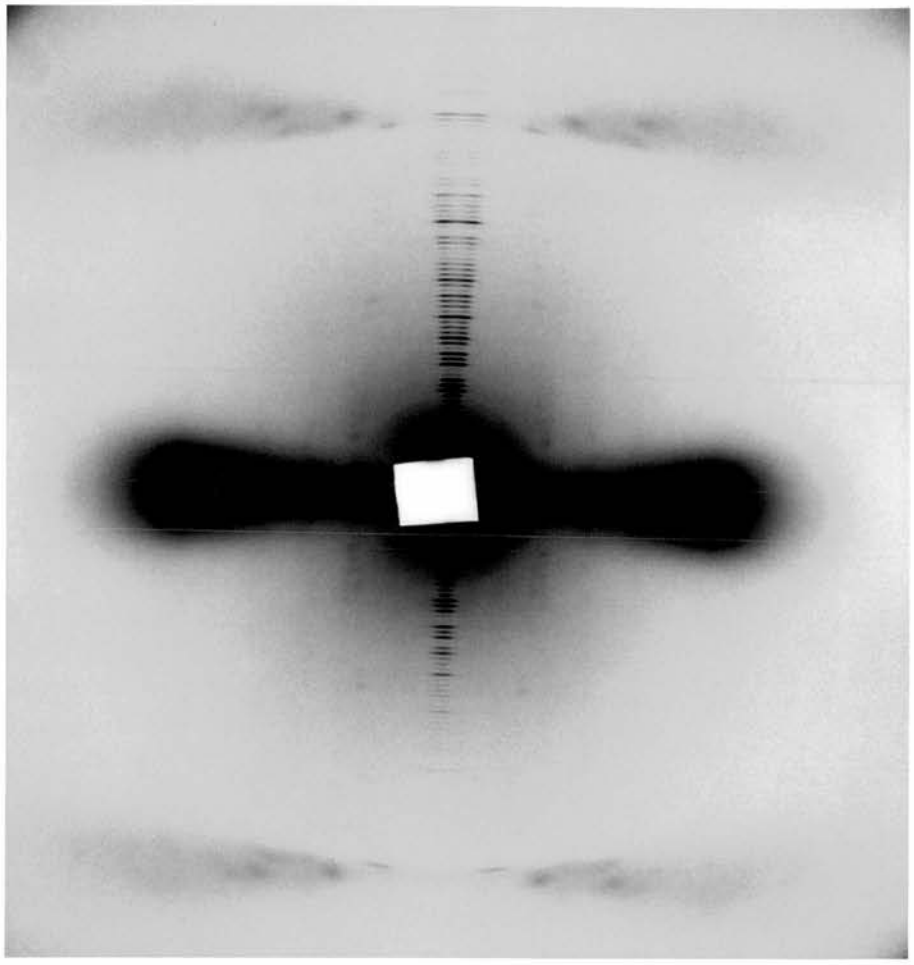


Plate 4.4 higher angle diffraction pattern of rat tail tendon iodinated using chloramine T. Taken on beamline 7.2 at the S.R.S., Daresbury.

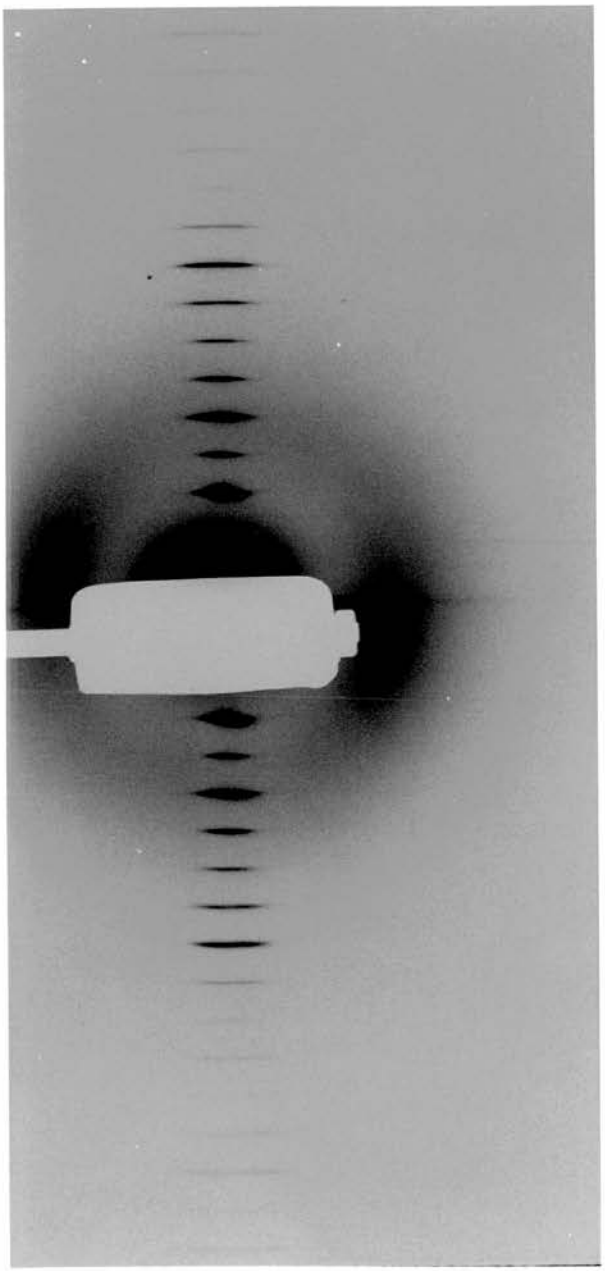


Plate 4.5 small angle diffraction pattern of rat tail tendon stained with $K_2^+PtCl_4^{2-}$.
Taken on beamline 8.2 at the S.R.S., Daresbury.

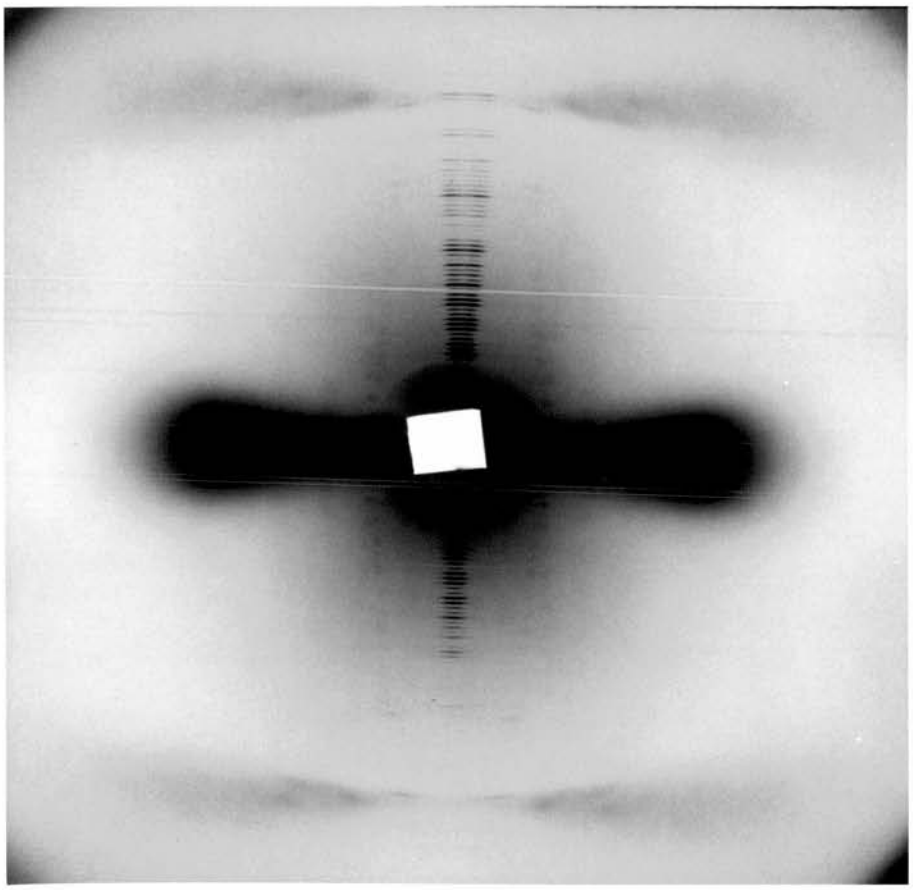


Plate 4.6 higher angle diffraction pattern of rat tail tendon stained with $K_2^+PtCl_4^{2-}$. Taken on beamline 7.2 at the S.R.S., Daresbury.

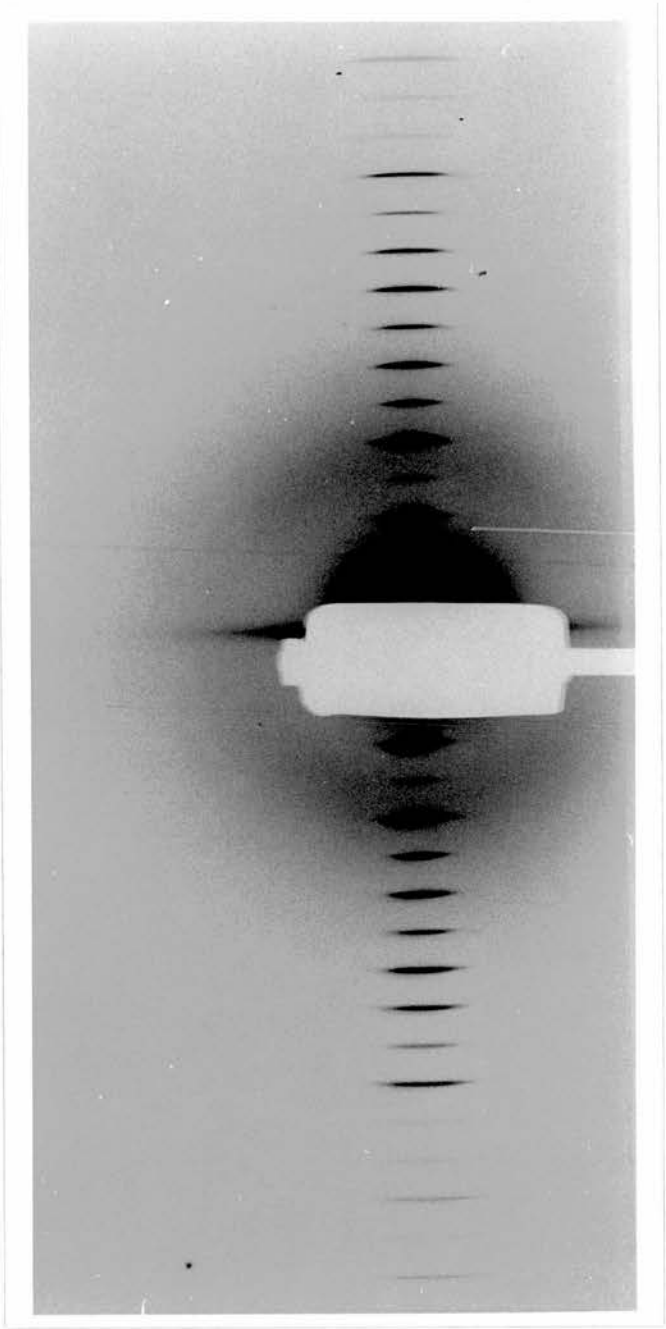


Plate 4.7 small angle diffraction pattern of rat tail tendon stained with $Na^{+}AuCl_4^{-}$. Taken on beamline 8.2 at the S.R.S., Daresbury.

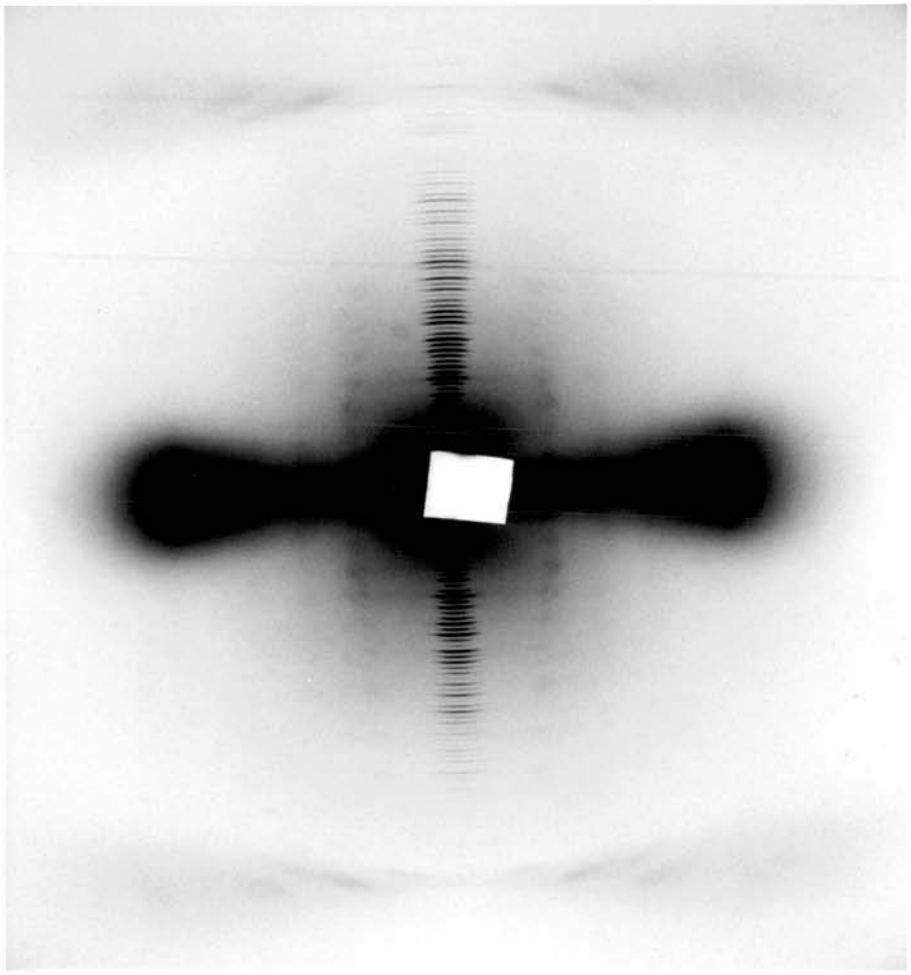


Plate 4.8 higher angle diffraction pattern of rat tail tendon stained with $Na^+AuCl_4^-$. Taken on beamline 7.2 at the S.R.S., Daresbury.

4.6 DATA ANALYSIS

Phase determination by the method of multiple isomorphous derivatives requires a knowledge of the location of the label within the unit cell. With this in mind, the first step in the analysis of the data was to calculate a difference Patterson function for each derivative type and the untreated control. This gave two levels of information.

Firstly, the presence of a few discrete peaks in the difference function showed a potentially usable derivative: too many peaks, or no peaks at all, meant that the particular set of reagents and conditions had not produced a suitable derivative.

The second level of information is contained in the positions and relative strengths of the peaks, which enabled us to determine which amino acid residues were being labelled in each preparation. This is the advantage of using reagents which label selectively over more indiscriminate electron microscope-type stains.
The latter were used in some earlier attemps to phase the collagen meridional diffraction pattern (Stinson et al, 1979). It is possible to interpret the much simpler difference Patterson functions in the light of the known reactivity of the reagent and the amino acid sequence of the protein.

When the best reaction conditions for each of the three reagents had been determined, and high and low angle

diffraction data from native and derivative tendon had been collected and treated, autocorrelation functions of the predicted label distributions were produced for direct comparison with the difference Patterson functions. The distribution of label was then adjusted and autocorrelated until a good fit was produced. Clearly, the two functions could not be expected to agree perfectly, since the autocorrelation function for each derivative only contains stain-stain vectors, whilst the difference Patterson function also contains protein-stain vectors.

Disregarding the background of the Patterson function, fitting the stain peaks, in terms of position and relative height, enabled the stain distribution to be determined. In each case the predicted distribution of labelling atoms needed only slight adjustment in order to fit the difference Patterson function, with the exception of any binding site which lay in the extrahelical telopeptide region of the molecule. The comparison of autocorrelation of labelling positions to difference Patterson maps is shown in Figures 4.1 and 4.2.

Iodine was initially assumed to label only tyrosine residues. The first model, therefore, contained peaks of electron density corresponding to the positions of tyrosine in the unit cell. An autocorrelation of this electron distribution function was then compared to the difference Patterson calculated using the iodine treated and control data. As a result of this comparison, electron density peaks were added at some of the histidine

residues before another autocorrelation function was calculated. In this way the label distribution map was modified and fitted to the observed data until the best fit was observed.

Although the starting point for this procedure was always the amino acid sequence of type I collagen and reported reaction characteristics of the label, the modifications made to the label distribution map in order to obtain a good fit to the observed data were not restricted to either of these factors. In other words, the final label distribution does not depend on any preconceptions of either sequence or reactivity.

Moreover, the refinement cycle described below further modified this label distribution. The effect of this can be seen in the case of iodine where the position of the iodine peaks in the telopeptide regions of the final map have moved closer together than in the original model, indicating a more contracted telopeptide conformation.

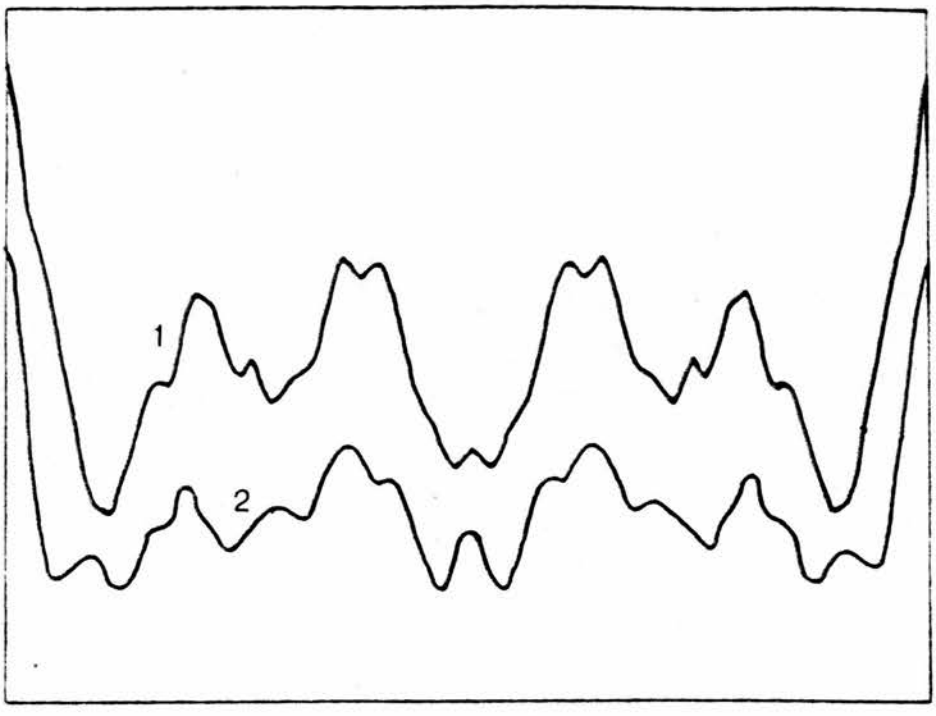


Figure 4.1 (for legend see overleaf)

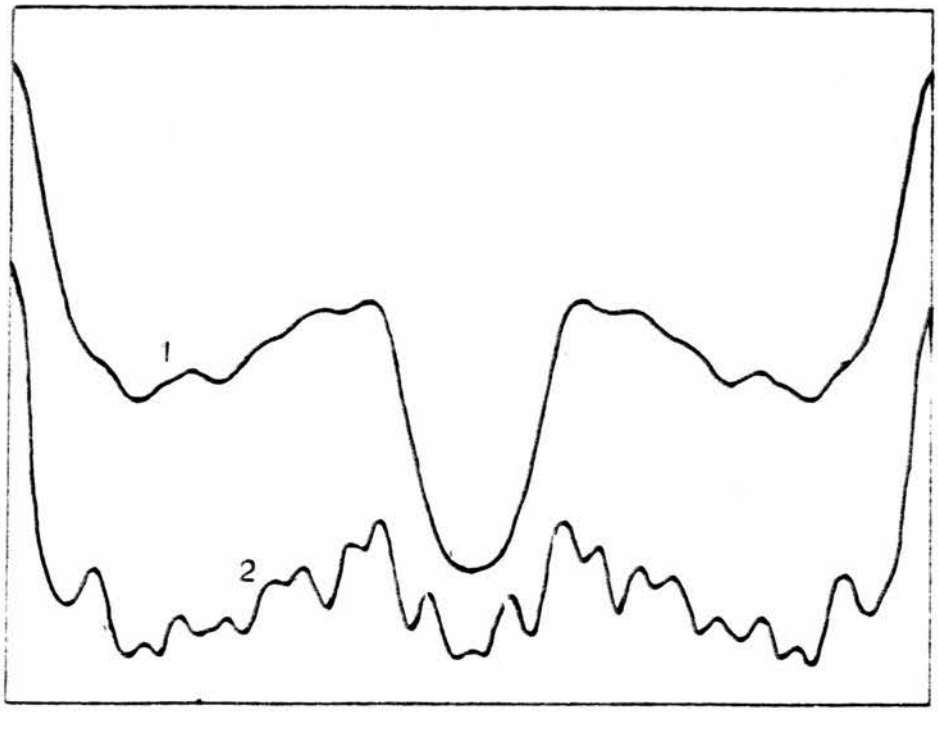


Figure 4.2 (for legend see overleaf)

Figure 4.1 (overleaf) shows the difference Patterson function of the derivative stained with $K_2{}^+PtCl_4{}^{2-}$ as profile 1, and the autocorrelation of proposed $K_2{}^+PtCl_4{}^{2-}$ staining sites as profile 2.

Figure 4.2(overleaf) shows the difference Patterson function of the iodinated derivative as profile 1, and the autocorrelation of proposed iodination sites as profile 2.

The stain distribution used to generate the autocorrelation functions shown overleaf, are those given in figure 4.6. Initial stain distribution of platinum for the autocorrelation, assumed labelling at methionine residues only. Initial stain distribution of iodine for the autocorrelation assumed labelling at tyrosine residues only.

The difference Patterson function is produced by applying the Patterson function to the differences between the intensity of the scaled derivative and native rat tail tendon meridional diffraction maxima (for the first fifty orders). The general equation for this is given by:-

$$\mathbf{P}(U)=a\Sigma[|F_{ph}|^2-|F_p|^2]\cos(2\pi U)$$

Where U is the fractional position of the heavy atom - heavy atom vectors.

4.6.1 REFINEMENT OF THE STRUCTURE

The positions of the labelling atoms within the unit cell of the collagen fibril were refined using the refinement cycle described in Figure 4.3. This system was designed to allow, with each cycle, the combination of the pairs of phases determined from each derivative, and the subsequent selection of one from each pair, and both passive (by means of the mathematical constraints of the cycle) and active (by means of comparisons with difference Patterson functions and background flattening) refinement of the stain peaks' height, shape and size.

The procedure starts with the proposed distribution of label electrons (a). Fourier inversion of this function is then combined with the three sets of observed intensities using the Harker construction to calculate the first set of native phases (b), and derivative phases (c). The phases determined by this method enable the calculation of the electron density distribution of not only the native fibril, but also of the three derivative fibrils (d). Simple subtraction of the native electron density from each derivative supplies a refined label distribution (e) which may be modified (f) (background flattening, adjustment to improve agreement with difference Patterson functions etc.) or left as it is before repeating the refinement cycle.

Since the phases and subsequent electron density distribution function calculated by each cycle are severely constrained by the four sets of observed data.

The process of cycling provides its own refining influence; for this reason the option of adding additional information was only used occasionally in the early stages to speed up the refinement. This process was continued until the label electron density showed good agreement with the difference Patterson functions, and remained unchanged after further cycles of the refinement procedure without input of further information at (f).

4.7 RESULTS

Table 4.2 gives the final sets of phases for the native and the three derivatives. The accuracy of these values varies from ± 0.02rads for the strong reflections to ± 0.15rads for the weakest ones. When these phases are combined in Fourier syntheses with the observed intensities (Table 4.1), they produce the four axial electron density distributions shown in Figures 4.4 and 4.5.

In order to eliminate termination error, the amplitudes were treated with a factor $\exp(-n^2B)$ where n is the order of diffraction and B is a constant chosen so as to reduce the amplitude of the 50th order to 1% of its original value (Sherwood 1976). The peaks in the maps produced by the Fourier of the amplitudes and phases from the native fibres correspond to regions of above average electron density in the axially projected collagen fibril structure. The maps obtained using the derivative data

contain essentially the same information, with the heavy atom label positions superimposed upon it. A much clearer indication of this heavy atom distribution may be obtained by subtracting the native from the derivative electron density maps (Figure 4.6).

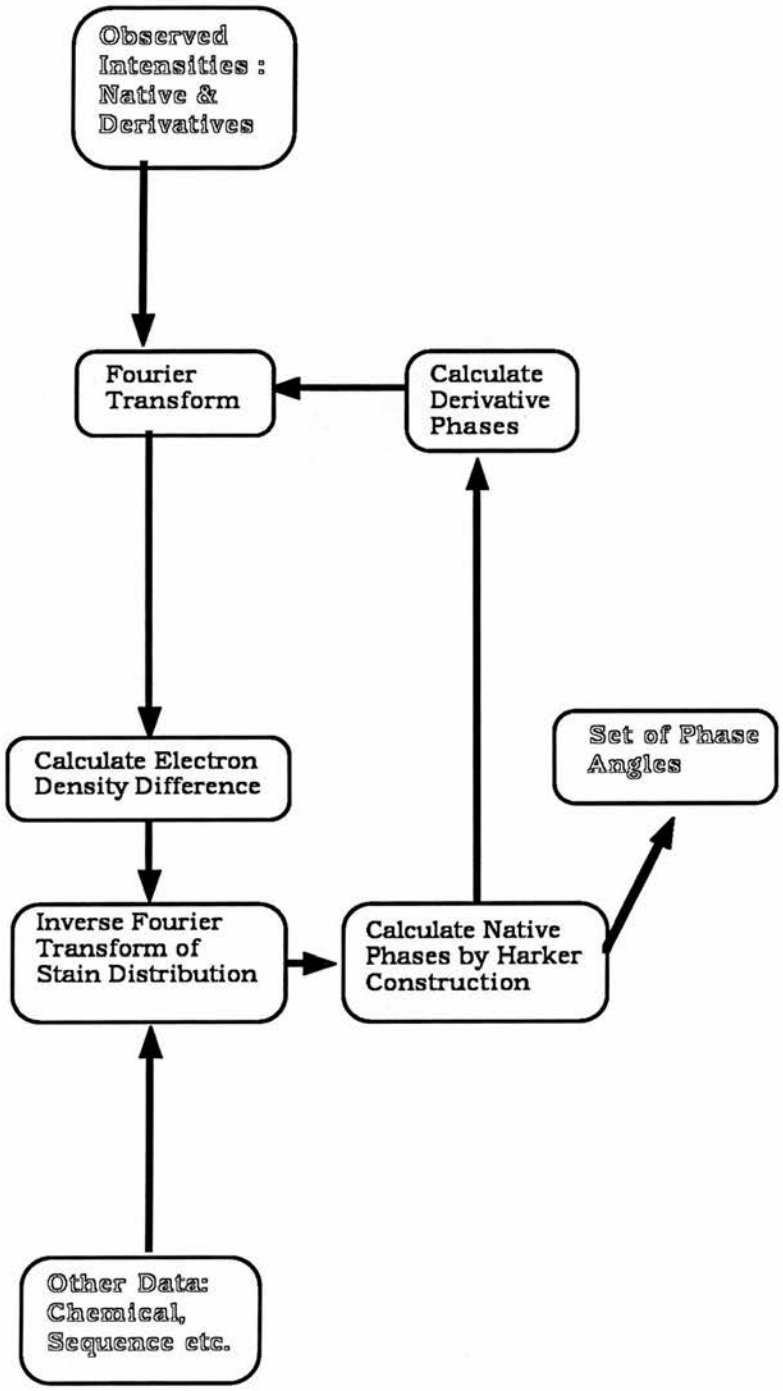


Figure 4.3 cycle of refinement used in the calculation of native and derivative phases

ORDER	CORRECTED INTENSITIES			
	NATIVE	IODINE	PLATINUM	GOLD
1	10000	10000	10000	10000
2	154	15	192	67
3	1417	3559	2586	2230
4	147	29	459	59
5	597	1612	2245	2037
6	228	451	700	268
7	209	557	259	506
8	110	358	350	264
9	586	619	572	189
10	138	460	200	364
11	131	177	79	19
12	220	353	36	269
13	26	40	21	30
14	57	103	257	5
15	40	82	213	267
16	37	40	47	42
17	63	62	92	120
18	14	39	2	76
19	21	56	77	26
20	188	347	352	663
21	176	156	122	144
22	72	100	147	236
23	5	3	22	43
24	3	13	14	25
25	111	100	151	268
26	30	50	165	50
27	76	92	122	174
28	3	16	21	36
29	35	12	10	9
30	82	72	109	69
31	19	16	90	26
32	19	18	93	38
33	12	3	23	47
34	86	54	98	24
35	14	4	21	25
36	22	17	60	11
37	24	25	30	17
38	41	30	76	42
39	9	4	17	20
40	25	8	21	34
41	63	34	60	65
42	29	14	35	2
43	16	13	15	36
44	27	12	1	28
45	3	5	1	12
46	12	3	5	0
47	37	4	1	4
48	20	12	15	24
49	13	11	10	22
50	1	0	10	1
51	5	3	6	0
52	46	28	43	58

Table 4.1 the corrected X-ray diffraction intensities of native and derivative tendon.

ORDER	PHASES (RADIANS)			
	NATIVE	IODINE	PLATINUM	GOLD
1	3.14	3.15	3.21	3.23
2	0.12	2.91	3.54	3.63
3	0.44	0.22	0.17	0.49
4	2.10	3.71	2.16	1.05
5	1.71	2.42	2.12	2.22
6	4.65	5.72	6.00	1.63
7	6.08	5.40	0.36	0.05
8	2.93	2.47	3.19	3.68
9	0.87	1.38	6.11	3.25
10	0.12	5.74	5.74	0.15
11	5.93	1.49	2.76	4.84
12	0.93	2.48	5.53	2.73
13	3.33	5.73	1.48	0.38
14	2.43	1.73	0.07	6.15
15	0.71	2.72	5.03	4.00
16	5.78	5.76	4.32	4.84
17	3.34	6.08	0.85	0.89
18	0.10	2.96	5.02	1.19
19	5.37	2.37	1.79	2.51
20	1.71	6.14	0.91	0.63
21	2.53	2.35	0.34	1.09
22	5.59	2.69	2.55	3.83
23	5.58	0.01	1.09	2.59
24	3.01	5.22	5.16	1.55
25	5.49	2.35	1.90	2.95
26	4.04	5.07	3.22	4.11
27	1.70	5.63	5.84	0.96
28	0.11	2.31	2.02	5.23
29	0.33	1.66	2.57	4.28
30	2.13	5.26	4.34	2.53
31	5.25	2.56	0.30	1.23
32	5.48	1.56	3.59	5.23
33	1.62	5.47	4.85	6.17
34	3.66	4.09	4.93	5.19
35	6.13	2.57	0.82	5.39
36	0.88	1.04	3.39	4.61
37	2.74	5.20	0.58	2.15
38	5.56	1.78	0.09	2.63
39	6.20	2.24	3.95	2.00
40	2.36	5.36	1.21	0.50
41	1.10	4.69	0.78	0.17
42	6.27	2.81	2.81	3.68
43	3.32	3.29	3.30	3.65
44	0.85	0.96	1.24	1.85
45	4.51	3.14	3.78	2.73
46	2.66	2.74	3.00	3.56
47	1.57	1.68	0.78	1.20
48	1.24	1.39	1.64	1.83
49	4.40	3.14	5.79	5.67
50	5.52	6.28	5.00	6.28

Table 4.2 the phases of native and derivative tendon reflections as determined by X-ray diffraction.

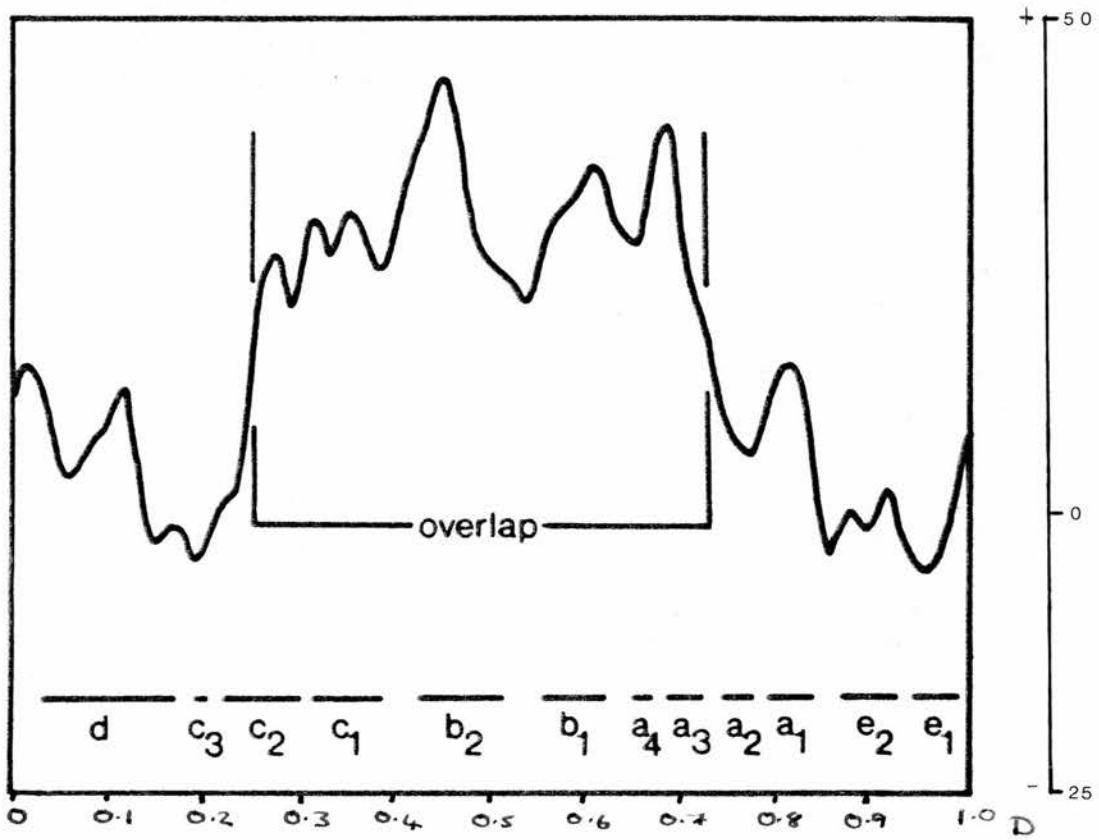


Figure 4.4 Fourier synthesis of the axially projected electron density map of a collagen fibril in wet rat tail tendon. The observed amplitudes were combined with the phases calculated from the observed data as described in the text. The positions of the electron microscope positive staining bands are indicated. The X-axis represents one D-repeat of 670Å; the Y-axis is the net number of electrons/molecule per unit axial translation.

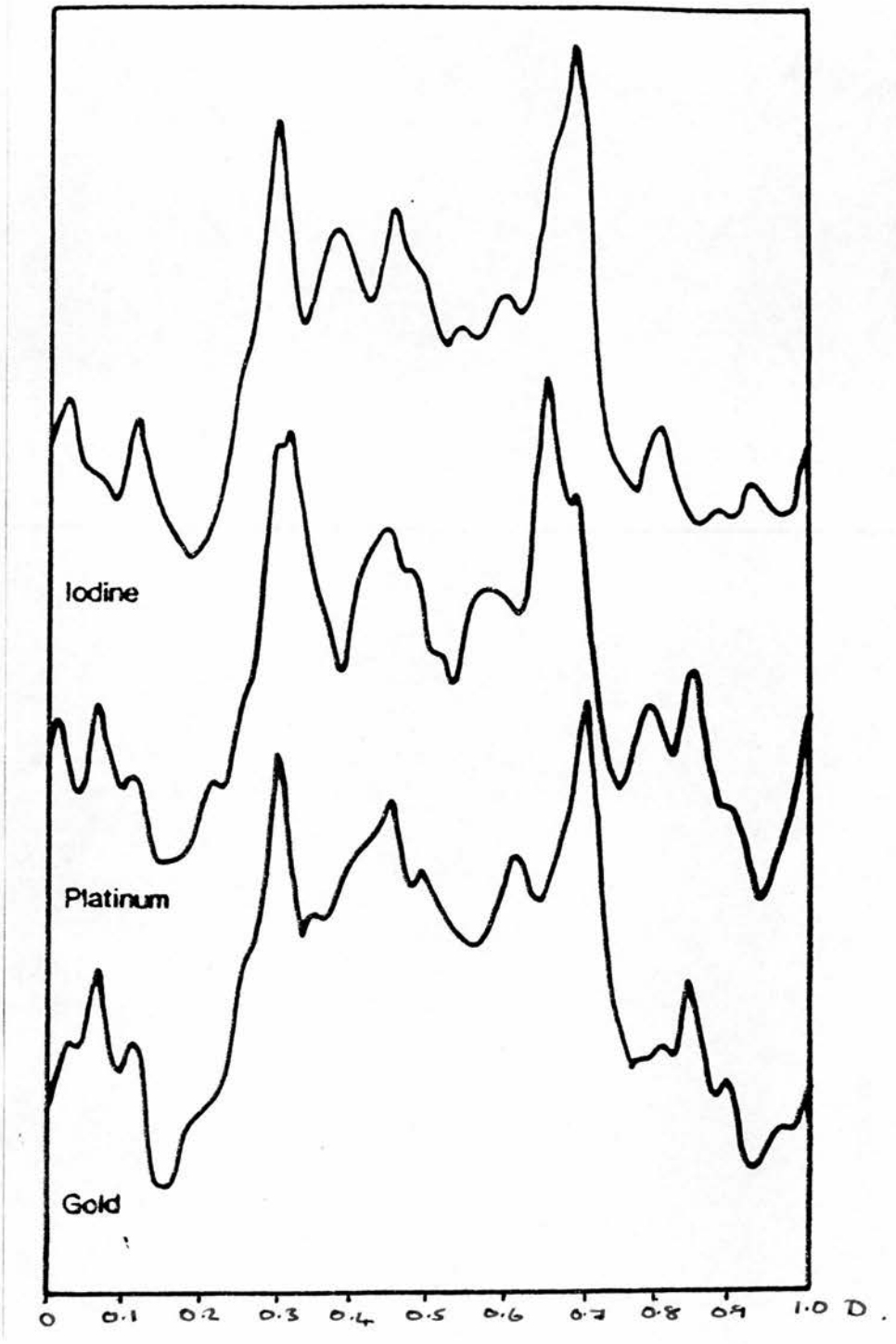
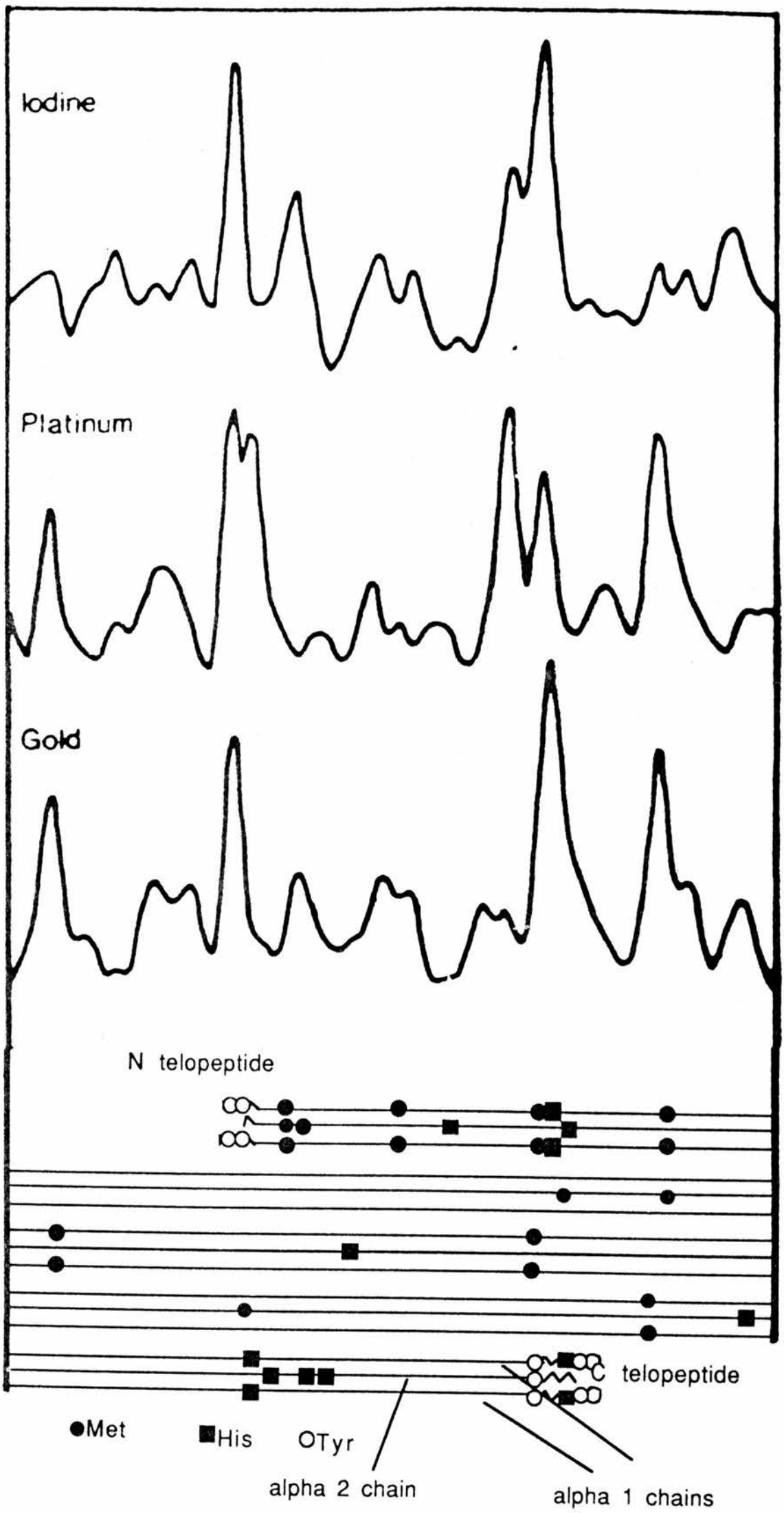


Figure 4.5 Fourier syntheses of the axially projected electron density maps of labelled collagen fibrils in wet rat tail tendon. The axes and scale are the same as in figure 4.4.


Iodine
Platinum
Gold
N telopeptide
C telopeptide
●Met
■His
OTyr
alpha 2 chain
alpha 1 chains

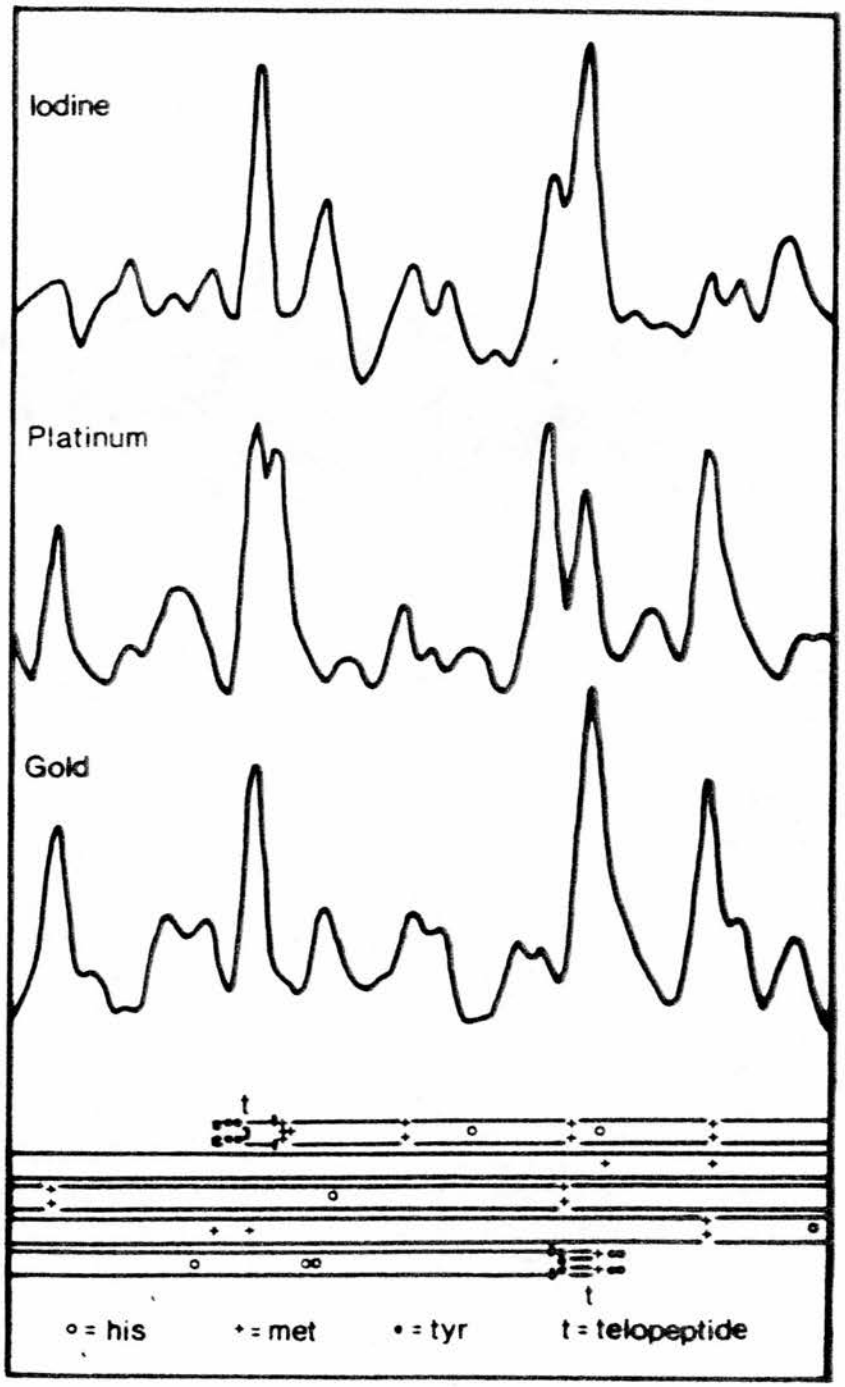


Figure 4.6 Difference Fouriers (calculated by subtracting native electron density from derivative electron density) representing the distribution of labelling atoms in the three types of derivative. The axes and scale are the same as in figure 4.4.

4.8 DISCUSSION

4.8.1 NATIVE ELECTRON DENSITY DISTRIBUTION

Each map (Figures 4.4 and 4.5) shows the characteristic gap/overlap function of electron density within a D-repeat axial unit cell. The length of the overlap was measured to be 0.48D. The limits of the overlap region were estimated to be at the two points where the electron density profile crosses the horizontal line of points equidistant between the highest overlap peak and the lowest trough in the gap region. These points are indicated in Figure 4.4. Since this measurement was based on data collected from 50 orders of the meridional diffraction maxima, an accuracy of $\pm$ 0.02D may be anticipated.

Previous structural analyses of collagen fibrils have not produced an unambiguous phase set, so have not supplied a conclusive value for the gap:overlap ratio. However, values of 0.4:0.6 (Spadaro 1970), 0.6:0.4 (Hodge & Petruska, 1963), 0.53:0.47 (Tomlin & Worthington, 1956) and 0.52:0.48 (Hulmes *et al*, 1980) have been cited. The relative lengths of gap and overlap presented here were determined using diffraction data which were both phased by a non model dependent technique, and of a higher resolution than previous studies, so these values may be regarded as the most significant to date.

The extra-helical telopeptide regions are known not to share the triple helical conformation of the rest of

the collagen molecule, since the regularly repeating sequence of amino acids, in the collagen helix, is not found in these regions. The net axial translation of the telopeptide residues is thought to be less than that of residues in the triple helix. Values of 2.82Å per residue in the helical region, and 2.41Å and 2.00Å in the N and C terminal telopeptides respectively (Hulmes et al, 1980). Capaldi and Chapman (1982) and Bender et al, (1983) also reported the C telopeptides to be condensed.

In the axial electron density map presented here the electron density in the region of the C terminal telopeptide is noticably higher than the N terminal region. This may be explained if the C termini are more contracted than the N termini. Other sections of collagen at the same axial position in the fibril unit cell will contribute electron density to this region. The position of a C terminal telopeptide with a smaller axial translation per amino acid (compared to the triple helix region) corresponds to the location of the A_3 staining band, and the C_2 band is sited in the region where an N-terminal telopeptide with reduced axial translation per amino acid would lie. There is agreement between the staining bands and the peaks in the electron density map presented here. This is expected since these regions are probably more rigid and make a large contribution to apparent electron density. However, the narrow bands A_2, A_4 and C_3 do not correspond to areas of high electron density in the map.

This could be due to the relatively small size of these bands, or possibly the axial mobility of these positions within the unit cell. Bands A_2 and C_3 are both in the gap region of the fibril, a region which is thought to be more mobile (Fraser et al, 1987).

One noticable difference between the map presented here and previous models of the axial electron density is the size of the peak located in the centre of the overlap region. The larger peak in our map corresponds well to the B_2 staining band of charged residues which forms a large, heavily charged area.

4.8.2 IODINE ELECTRON DENSITY DISTRIBUTION

The iodination procedure employed was thought to label tyrosine and histidine residues. Analysis of the label electron density distribution map (Figure 4.6) shows that the axial location of the iodination is mainly in the region of the C and N terminal telopeptides, in agreement with the findings of Ellis & McGavin (1970). However, iodine is also seen in two much smaller peaks in the overlap region. The height and width of these peaks would seem to indicate that labelling has occured at a small number of very specific sites, presumably the three histidine residues at the C_1 position on the $\alpha 1$ chain.

Analysis of the sequence of the N terminal telopeptide and laterally adjacent sections of the unit cell show that four tyrosine residues are located at the end of each $\alpha 1$ chain telopeptide (Fietzek & Kühn 1975),

and one at the end of the α2 chain (Bornstein & Traub 1979). Three histidine residues are found in the laterally adjacent sections. The axial separation of these groups appears to be insufficient, at the resolution of this study, to identify the individual labelling sites. This indicates that the groups must be less than 13Å apart. Moreover, the position of the peak in the electron density difference map, corresponding to the combined effect of these labelled residues is closer to the centre of the overlap region than would be predicted by a fully extended N terminal telopeptide. Taken together, these two observations support the view that this telopeptide has a contracted conformation.

At the C terminal telopeptide the electron density difference map shows two closely spaced peaks. Analysis of the amino acid sequences of the sections of molecule contributing to this region reveals two groups of tyrosine residues in the telopeptide, situated at the C terminus of the chain, and at the boundary between triple helix and non-helical telopeptide (Bernard et al, 1983, Chapman & Hulmes 1984). Despite the apparent contracted conformation of the telopeptide, these two groups can still be resolved on the electron density difference map, the distance between the two groups being approximately 0.04D.

The contracted nature of the N telopeptide is well supported by X-ray and neutron data (Hulmes et al, 1977, 1980), but the exact conformation is not known.

The apparent contraction may be caused by a Beta-turn in the amino acid chain (Helseth et al, 1979), or the reduced axial translation of each residue could result from a random coil conformation. Structure prediction routines have been run on the α1 N terminal telopeptide (Helseth et al, 1979) and on both the α1 N and C termini (Jones & Miller A. 1987). Although the axial structure presented here is unable to reveal the complete 3-dimensional conformation, it may prove to be a valuable method for distinguishing between the different conformations predicted by the different methods.

4.8.3 PLATINUM ELECTRON DENSITY DISTRIBUTION

The map (Figure 4.6) contains a series of sharp peaks which correspond well to the methionine residues which in axial projection form clusters in the unit cell (Chapman & Hulmes 1984)(Piez 1976). A number of histidine residues also seem to have been labelled.

Five major peaks can be recognised, which occur at the positions of methionine clusters. The relative heights of these peaks indicate that the occupancy of the various sites varies. This is presumably related to steric hindrance prevailing at each site and variations in the ability of the different methionines to undergo an SN_2 type ligand substitution. One peak at the axial location of two methionine residues in the gap region is comparable in size to another peak at the site of five methionine residues, implying that a similar number of platinum ligands are found at both locations.

Comprehending what the electron density map reveals about the labelling of histidine residues with platinum is more complicated, since the major groups of histidines occur near the clusters of methionines, except for a band of six histidine residues close to the N terminal end of the overlap region. These residues are labelled by iodine, but not by platinum. This may be caused by a local environment which presents unfavourable conditions for the labelling. Our trials with different reaction conditions showed that optimal labelling occured with the reagents buffered to physiological pH. At this pH only

approximately 10% of the histidine residues will be protonated. Petsko (1978) has stated that, at around this pH, the S-CH_3 of methionine will react preferentially with platinum derivatives.

The results of this platinum labelling technique may be compared to the work of Kühn et al, (1984) who used chloroglycyl-L-methionatoplatinum (II) (Pt-GLM) as a heavy atom label for Scanning Transmission Electron Microscopy (S.T.E.M.) Labelling was carried out at pH5, at which all the histidine residues are expected to be fully protonated and therefore difficult to label. The distribution of label in reconstituted rat tail tendon treated in this way shows a series of peaks with a high degree of correlation to that reported here, but to a lower resolution.

4.8.4 GOLD ELECTRON DENSITY DISTRIBUTION

Like platinum, gold is believed to have a strong interaction with methionine, and, also as platinum, an analysis of gold chemistry as used in conventional crystallography reveals a process of ligand exchange with the solvent, thus complicating any understanding of the underlying reactions.

Comparison of the gold and platinum electron density difference maps (Figure 4.6) shows a good correlation between the positions of the peaks. Moreover, the location of the residual labelling seems to be similar in the two maps. There are, however, a number of peaks which

are present in the platinum map but absent in the gold. There is little evidence that these differences can be accounted for by the presence of histidine residues at the same axial position, since the correlation of these peaks is poor.

The differences may result from the preferential labelling of methionine residues by platinum to gold in the overlap region since it is most interesting that both of the peaks absent from the gold map are found in the overlap region, just behind the telopeptides. This may imply that the telopeptides play a role in controlling the size of ions present in the overlap region, but the physicochemical implications of this are beyond the scope of this study.

4.9. CONCLUSION

The electron density maps produced by this study were found to be relatively easy to analyse in terms of the known chemical reactivity of the reagents used, and the available amino acid sequence data. The native electron density map contains peaks which correlate well with the electron microscopy positive staining bands, and since this information was obtained from unfixed, non-embedded material the findings should be directly applicable to the *in vivo* state. This constitutes the first unambiguous determination of the meridional phases of type I collagen, and supplies the methodology for future structural studies of, not only normal, but also pathological collagenous tissues.

CHAPTER 5

THE NEUTRON SCATTERING PROFILE OF RAT TAIL TENDON

5.1 INTRODUCTION

The use of neutron diffraction in structure determination has usually been to complement X-ray diffraction studies and provide information that cannot be directly obtained from X-ray studies. The advantages of neutron diffraction to other methods of structural analysis result from two properties of the scattering phenomenon.

1) The radiation is non-ionising, as the neutrons interact with the nucleus and therefore do not produce free radicals. A result of this is that damage to the specimen is vastly reduced compared to X-rays.

2) Scattering at the nucleus confers properties associated with the difference in scattering length of atoms. See chapter 2.

The technique has therefore been employed to determine the position of hydrogen atoms in protein crystal structures. Deuterated samples have also been used to determine the distances between subunits in proteins for example the ribosome (Moore et al, 1984). In the case of tendon collagen, a number of reducible sites exist primarily as crosslinks that can be labelled with deuterium and used in structural determination.

5.2 AIM

The ability to reduce crosslinks and other groups in rat tail tendon, indicates that the structure can be determined by a process of isomorphous addition, similar to that used in studies with X-ray diffraction in chapter 4. The reduction of the crosslinks by insertion of deuterium at these sites provides the derivative structures and analysis of the native and resultant derivative axial structures indicates the axial position of the reducible sites within the D period projection. This would aid both structural investigation of collagen and the chemical analysis of crosslink position in tendon.

5.3 PREVIOUS WORK

The previous investigations relevant to this study can be divided into two categories.

1)Neutron diffraction analysis of collagen.
A number of investigations of collagenous structures have been conducted, these include Hulmes et al, (1980), Miller A. et al, (1975), White et al, (1977). The first of these studies made a contribution to structure determination by a process of model building, in this, data from X-ray model building procedures were tested with neutron diffraction data from rat tail tendons at various D_2O/H_2O ratios.

A set of phases was obtained using the first twelve orders of diffraction. The investigation confirmed both the Hodge Petruska model and the contracted conformation of the telopeptides. White et al, (1977) used neutron diffraction to determine the position of the mineral phase in turkey leg tendon.

2) Chemical analysis of the crosslinks.

Amino acid analysis has been combined with amino acid sequence information to determine the nature and position of crosslinks, various techniques have been reviewed by Robins (1982). A number of crosslink types have been isolated, see Fig 1.2, however the existence of various types such as histidino-hydroxymerodesmosine has been questioned (Robins and Bailey 1973), since the process of hydrolysis and separation during analysis could produce artifacts.

The main terminology applied to crosslink types has been based on their ability to be reduced by agents such as $NaBH_4$ and its related isotopic compounds. This has been related to the acid solubility of collagens, since collagen from sources containing high levels of reducible crosslinks and low levels of non reducible crosslinks can be solubilised to a greater extent (expressed as percentage of total mass) than those with high levels of non-reducible crosslinks.

The degree of solubility has also been related to the age of the tissue in question. Tendon samples from young animals tend to be more soluble than those from older animals of the same species (Bailey et al, 1974). From this a process of crosslink maturation has been inferred. The reducible crosslinks of collagen are regarded as intermediates in the process of maturation (Robins & Bailey 1975) (Bailey & Skimokomaki 1971). Studies by Robins & Bailey (1975) using periodite degradation techniques demonstrated that the mechanism of crosslink stabilisation occurs through an Amadori rearrangement of various immature crosslinks, especially dihydroxylysinonorleucine.

Rat tail tendon is reported to contain allysine crosslinks (Bailey & Peach 1968). More recent structural and chemical investigations have postulated hydroxylysine mediated crosslinks (Nakamura 1987). There is a general concensus of agreement that rat tail tendon contains crosslinking between telopeptides of adjacent N and C termini when in axial register. Another crosslink has been found to link the telopeptide with a residue in the triple helix. However the existence of other crosslinks is debatable.

Crosslink analysis reveals the presence of other reducible groups in tissues such as tendon. This includes residues such as hexitol lysines (Rosenberg et al, 1979). Here the covalent Schiff base linkage of a hexose sugar to the εNH_2 of a lysine

residue readily lends it to reduction by borodeuteride. Reducible compounds and their deuterated derivatives can be seen in Figure 5.3.

5.4 EXPERIMENTAL

5.4.1 PRODUCTION OF DEUTERATED COLLAGEN

A method of specific reduction using $NaBD_4$ was developed using an adapted method of Bailey et al, (1968). Reaction conditions were selected that would allow a series of deuterated derivatives to be produced with two requirements in mind.

1) The amount of deuteration must be sufficient to produce different diffraction pattern intensities.

2) The derivative must be isomorphous with the native samples.

To comply with the first requirement, experimental conditions had to be varied with respect to the time of labelling and the concentration of $NaBD_4$.

Specific reduction of collagen was investigated by Svendsen and Koch (1982), where they found that the X-ray diffraction pattern of non-reduced collagen deviated from the normal, if the sample was exposed to a low pH and then returned to pH 7.5. This did not occur in the deuterium reduced samples where the diffraction pattern was identical to that of the native state.

This could be used as an indication of the degree of labelling.

To comply with the requirements of isomorphism, derivative samples were examined by X-ray diffraction. The resulting diffraction patterns showed no difference in the position and intensity of reflections.

The production of deuterated collagen with respect to the above criteria resulted in the following experimental conditions, using 1g of wet rat tail tendon obtained from 4-6 month old Wistar rats.

20 mgs $NaBD_4$ in buffered Ringer solution
at pH 7.5 for 6 mins. reaction volume 25 mls,
at room temperature.

20 mgs $NaBD_4$ in buffered Ringer solution
at pH 7.5 for 30 mins. reaction volume 25 mls,
at room temperature.

The number of derivatives had to be limited since the amount of beam time available on instruments D11 and D16 was insufficient to test more derivatives.

5.4.2 NEUTRON DIFFRACTION OF SAMPLES

Tendons were dissected from rat tails and immersed in buffered Ringer solution at pH 7.5. Derivative samples were produced by methods given in 5.4.1.

Approximately 150 tendon fibres of average individual diameter approximately 200μm were clamped close together on a frame so as to produce a parallel array of collagen fibrils. These could then be stretched by 5% to remove the crimp. For experimentation on D11 the frame was sealed in a teflon cell with quartz windows as described in 3.3.2 and for experimentation on D16 the frame was mounted and sealed in an aluminium can, see 3.3.2.

Diffraction experiments were carried out on D11 for lower axial orders of diffraction (1-7) and on the diffractometer D16 for the higher orders (5-23).
Experiments on D11 were carried out with the collagen fibres perpendicular to the beam, with the axis of the fibres vertical. Two sample to detector distances were used in order to collect a low angle data set. To record the first 3 orders and the transmitted intensity, on the 64cm x 64cm area detector, neutrons of wavelength 9Å (deviation of wavelength spread $\delta\lambda\backslash\lambda$=9%) were selected and the sample to detector distance set at 5.6m.

The distance from the effective source of the neutrons was set at 5m by moveable collimation to give a beam cross section profile of 50mm high x 30mm wide.

Further collimation was used in order to prevent neutrons hitting the sample cell. This used a circular cadmium aperture, at the sample, of diameter 10mm.

Transmitted beam intensities were measured for a fixed time interval after attenuation of the beam at source with a graphite moderator. The beam stop was removed to record the flux of the through beam.

In order to record orders 2 to 7 the same wavelength neutrons were used and the sample to detector distance shortened to 2.53m by moving the detector.

The D11 data were initially corrected for detector response efficiency using the isotropic incoherent scattering of H_2O. The resulting diffraction profiles were stored on tape and all subsequent data processing conducted on VAX (DEC) and microcomputers at Edinburgh.

Higher orders of axial reflections were measured on the four circle diffractometer D16 using $\omega/2\theta$ scans with the beam axis horizontal and the collagen fibres in the horizontal plane. The sample arrangement is described in chapter 3. The beam was rendered monochromatic of wavelength 4.43Å by use of a pyrolytic graphite crystal. Collimation was provided by gadolinium coated steel Soller slits. The total beam cross sectional area was $100mm^2$. Measurements were made by $\omega/2\theta$ step scanning (0.08° steps in ω for ~8 minutes per step). This allowed the intensities of orders 5 to 23 to be recorded for each sample. Data were collected and stored on tape to allow processing at Edinburgh.

5.4.3 DATA PROCESSING

Integrated intensities for each reflection obtained using instrument D11 were determined by using a suite of programmes written in Fortran 77 for the BIOVAX at Edinburgh University. The programme NUTSPEC allowed sections of the 64 x 64 data array from the detector to be viewed. The width of each reflection was then estimated. From this, values of intensity for each order were obtained and corrected against a local background. It was therefore possible to merge the data from the two different camera lengths by using standard programmes as described in chapter 4 for the X-ray studies of meridional reflections. The intensity of the second and third orders in each data set were used to determine a scale factor for merging the data together.

In order to relate these intensities to the true structure factor for each reflection, it was necessary to correct for specimen disorientation, beam divergence, wavelength spread and the intersection of each reflection with the Ewald sphere (Lorenz factor). The magnitude of these corrections was determined experimentally by Hulmes et al, (1980). Here a rocking curve analysis was used on the instrument D17, this was used to determine a correction term for data obtained using instrument D11. This correction was found to be small. For example, with the ratio of the fully corrected first order intensity of D17 to the integrated intensity of the first order from D11 (both samples at 100% H_2O), the correction was a

Fortran 77 programmes written by J.P. Bradshaw.

monotonically increasing function of the order number. Therefore for the sixth order the correction ratio was determined to be 2.5. The corrections were applied to each D11 data set as conducted by Hulmes et al, 1980.

D16 data were analysed using the programmes D16ADD and NUTSPEC. This allowed the intensity of each reflection to be estimated by the relative height of each peak. The intensities could then be corrected for local background scattering and a Lorenz factor applied to each reflection. These data were then scaled to the corrected intensities of D11 data by using the fifth, sixth and seventh orders common to both data sets. The resulting intensities were all scaled to a first order value of 10000 and are shown in Table 5.1.

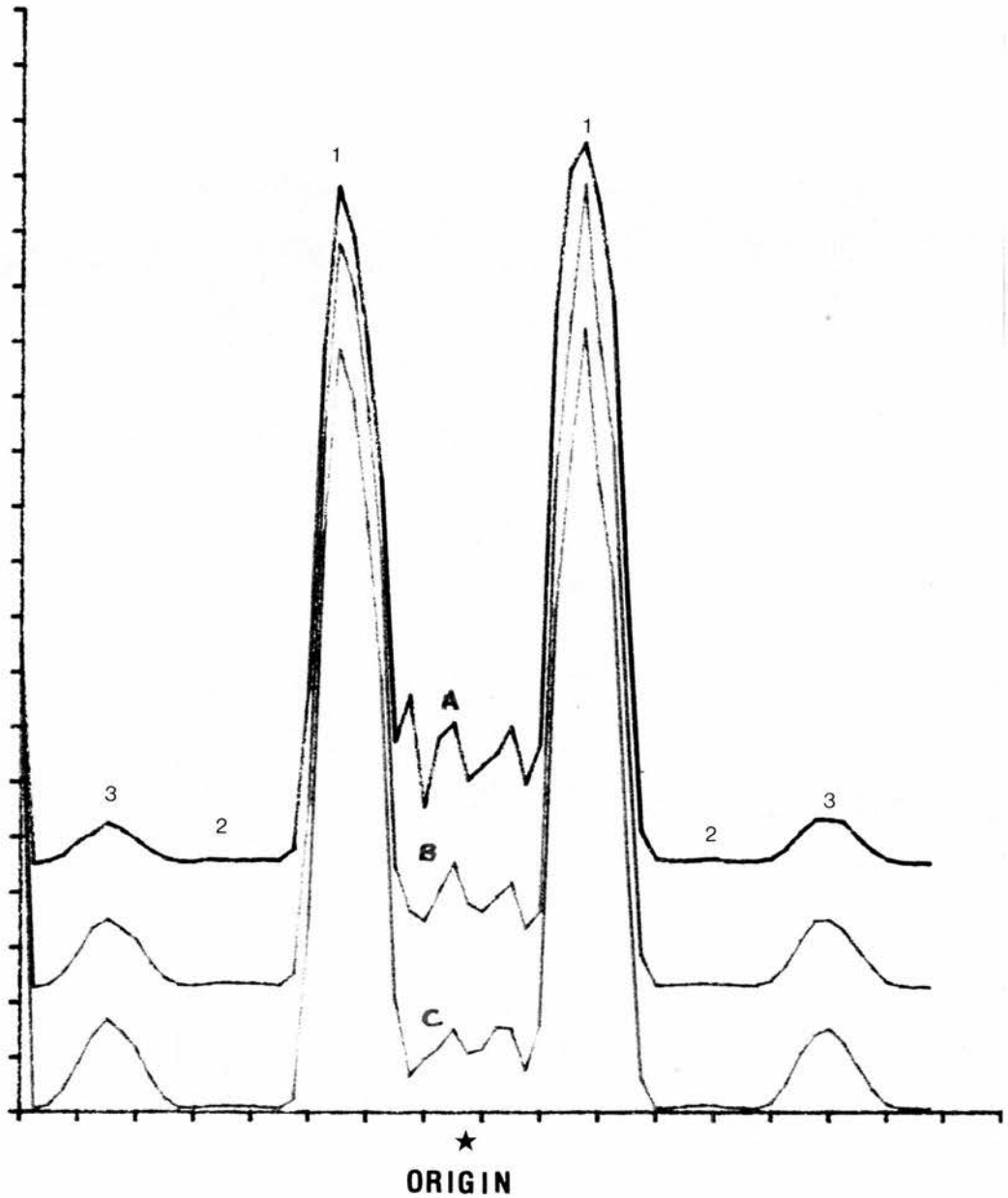


Figure 5.1 plots of neutron diffraction intensities for orders 1-3 for native and derivative samples, at 0% D_2O.

Plot A wet rat tail tendon labelled with $NaBD_4$ for 30mins

Plot B wet rat tail tendon native

Plot C wet rat tail tendon labelled with $NaBD_4$ for 6mins

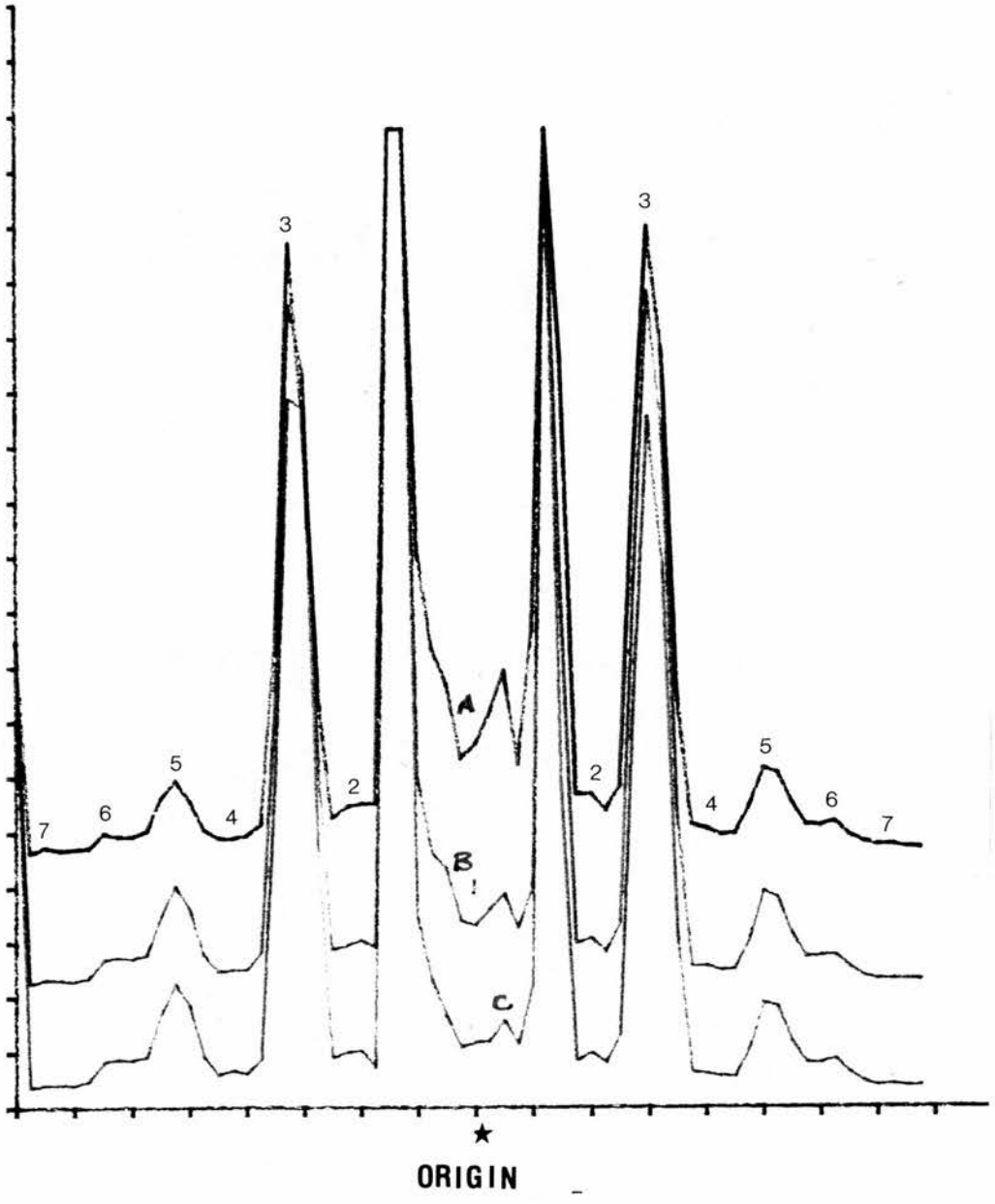
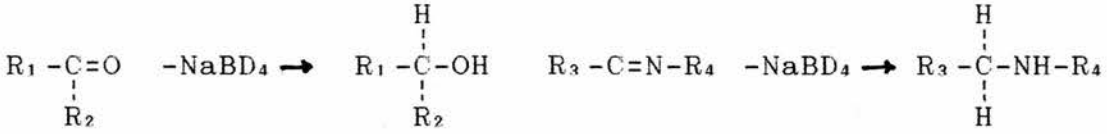


Figure 5.2 plots of neutron diffraction intensities for orders 2-7 for native and derivative samples, at 0% D_2O.

Plot A wet rat tail tendon labelled with $NaBD_4$ for 30mins

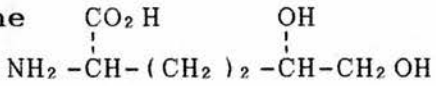
Plot B wet rat tail tendon native

Plot C wet rat tail tendon labelled with $NaBD_4$ for 6mins

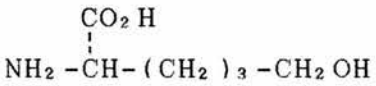


The reduction of carbonyl and aldimine compounds by $NaBD_4$

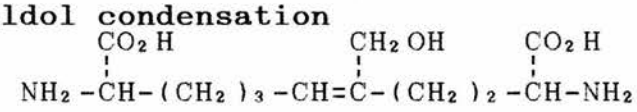
Dihydroxynorleucine



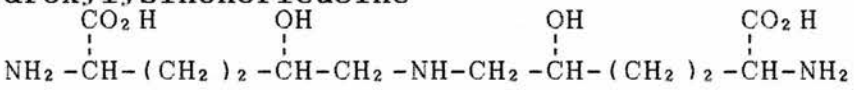
Hydroxynorleucine



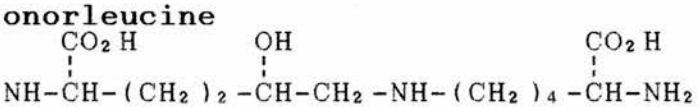
Reduced aldol condensation product



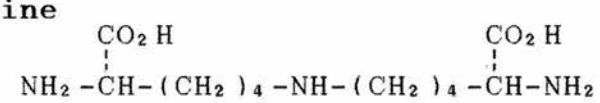
Dihydroxylysinonorleucine



Hydroxylysionorleucine



Lysinonorleucine



Hexitol-lysine

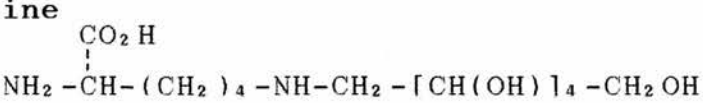


Figure 5.3 components of type I collagen reducible by $NaBD_4$, in reduced state.

ORDER	CORRECTED INTENSITIES		
	Native	6minNaBD$_4$	30minNaBD$_4$
1	10000	10000	10000
2	86	80	116
3	1945	2215	1900
4	106	114	118
5	600	620	455
6	430	330	400
7	100	130	134
8	270	225	330
9	190	190	220
10	135	110	100
11	130	140	160
12	195	255	190
13	47	90	20
14	69	146	90
15	101	126	101
16	38	157	30
17	25	190	55
18	112	113	112
19	92	100	120
20	129	208	170
21	129	240	129
22	175	210	210
23	120	100	70

The error in each intensity is approximately the square root of the measured value (White 1977).

Table 5.1 corrected intensities of native and deuterated samples.

5.4.4 DATA ANALYSIS

As for phase determination of a molecule using X-ray diffraction, the method of multiple isomorphous addition requires information relating to the position of label in the unit cell. This information was obtained from two sources, the difference Patterson function resulting from derivative and native diffraction intensities and known amino acid sequence information. These were compared by producing autocorrelation functions of the proposed labelling sites with the difference Patterson maps.

The difference Patterson functions show discrete peaks indicating deuteration at specific sites within the unit cell. The number of peaks was judged to be sufficiently small as to allow phase determination. The position and relative strengths of these peaks were used to construct and modify models of the labelling sites within the molecule. This was conducted until an autocorrelation for each derivative had a similar peak distribution to the corresponding difference Patterson function.

Initial autocorrelations only used the crosslinking sites at the telopeptides for positions of deuteration. The proposal of labelling at two sites cannot explain the number of discrete peaks produced in the difference Patterson maps. The possibility of other sites being deuterated was investigated.

Many molecular species have been isolated from collagen that can be reduced by $NaBH_4$. These include the crosslinking groups in various forms, hexitol lysine residues and the processed precursors of crosslinks, the aldehyde derivatives of lysine and hydroxylysine (Bailey et al, 1974) (Robins 1982). It is possible that these components can be deuterated in rat tail tendon and therefore contribute to the difference in scattering between native and derivative samples.

A number of trial distributions of deuteration in the axial unit cell were tested. These distributions were dictated by the position of lysine and hydroxylysine residues in a D repeat. The autocorrelation functions tested, began to correspond well with the difference Patterson function when hydroxylysine residues were used as sites of deuteration. A predominant deuteration site at a group of lysine residues in the gap was also postulated.

The combined contribution of deuteration at the telopeptide regions, hydroxylysine rich regions and a lysine rich region in the gap, produced the autocorrelation functions that matched the difference Patterson maps of each derivative sufficiently well. The significant differences in the 6 minute and 30 minute $NaBD_4$ labelled derivatives could be produced by altering the relative occupancy at each labelling site. This gave a starting point that allowed phase determination.

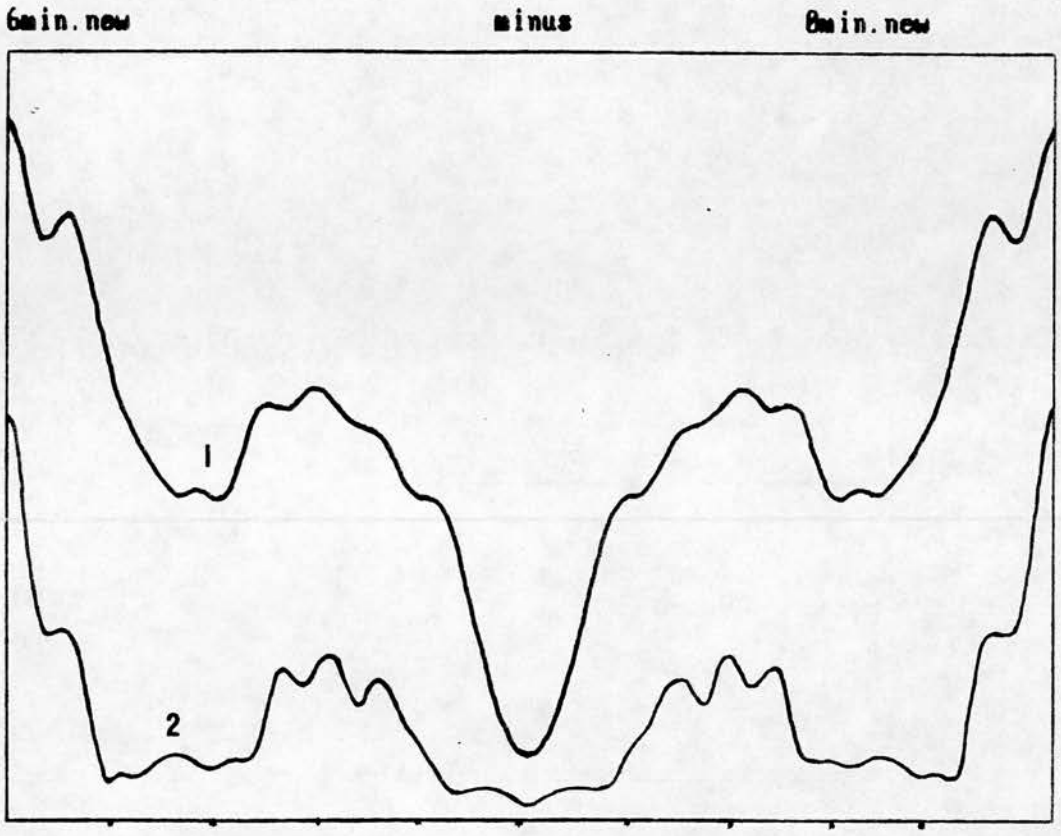


Figure 5.4 the comparison of (1) the difference Patterson of 6minute $NaBD_4$ labelled collagen and (2) an autocorrelation calculated from proposed labelling positions. The model used to generate the autocorrelation function was the final deuterium labelling position as shown in figure 5.6. The position of peaks of the difference Patterson map represent the vectorial distance between sites of deuteration. The height of the peaks relates to the amount of deuterium at the ends of a vector this length.

5.5 FOURIER SYNTHESIS AND PHASE DETERMINATION

A map of proposed sites for deuterium labelling in the unit cell was used to obtain the vector F_h by a process of Fourier inversion. The resulting vectors could then be used in the Harker construction. (see Figure 2.5). The combination of Fh with |Fph| and |Fp| for each meridional reflection recorded, allowed the phase associated with each native tendon reflection to be obtained. The use of two derivatives where the labelling of each in the unit cell is different, means that the phase determination by this process is unambiguous.

The production of an initial set of phases for the native sample allows the phases associated with the derivative structures to be obtained, by a process analogous to that used in 4.6.1. Therefore neutron scattering profiles for the native and derivative samples can be produced. The subtraction of the derivative from the native profiles, results in neutron scattering density difference maps that can be used on entering a cycle of refinement.

5.5.1 REFINEMENT

The refinement cycle used in the X-ray diffraction analysis of tendon structure was used here to refine the structure determined by neutron diffraction.

The progress of refinement was judged by the comparison of autocorrelations resulting from each neutron scattering difference profile. The process of refinement was continued until an optimum correlation between the autocorrelation and original difference Patterson maps was achieved, taking into account the absence of cross vector terms corresponding to protein deuterium vectors in the autocorrelation function, with little change in the phases on subsequent cycles.

5.6 RESULTS

The corrected intensities for the native and derivative tendon samples in 0% D_2O are given in Table 5.1. The final set of phases for the native and derivative samples is given in Table 5.2.

The corresponding neutron scattering profile of native tendon derived from these values is shown in Figure 5.5. The difference profiles that are obtained by the subtraction of the derivative scattering profile from the native provide an effective way of observing the alterations made to the scattering profile and are shown in Figure 5.6. Since the neutron scattering of each sample was conducted in a medium of 100% H_2O, each scattering profile exhibits a step function, characteristic of the gap/overlap function found in tendon. The contribution of the water in the gap to the coherent scattering being less than that of the overlap region that contains a higher density of protein.

The resolution of this study is to $\pm$ 0.043D and constitutes a presentation of the neutron scattering profile of tendon to the highest resolution so far.

ORDER	PHASES (RADIANS)		
	NATIVE	6minutes $NaBD_4$ derivative	30minutes $NaBD_4$ derivative
1	3.37	3.22	3.13
2	0.84	0.68	2.23
3	0.64	0.82	0.66
4	5.72	1.61	5.32
5	3.74	3.69	3.75
6	0.66	1.34	0.97
7	2.44	2.86	2.56
8	2.69	3.35	3.09
9	4.96	4.61	4.62
10	5.14	4.82	4.95
11	4.12	4.38	4.64
12	3.58	3.58	2.86
13	4.12	3.64	3.52
14	0.61	0.98	5.46
15	2.98	3.02	2.05
16	4.51	3.11	3.01
17	1.79	2.43	2.99
18	4.62	5.13	5.08
19	5.27	1.27	1.73
20	1.61	2.39	2.82
21	0.51	0.68	0.63
22	0.39	0.96	1.18
23	2.33	2.85	2.97

Table 5.2 the phases of derivative and native tendon samples as determined by neutron diffraction.

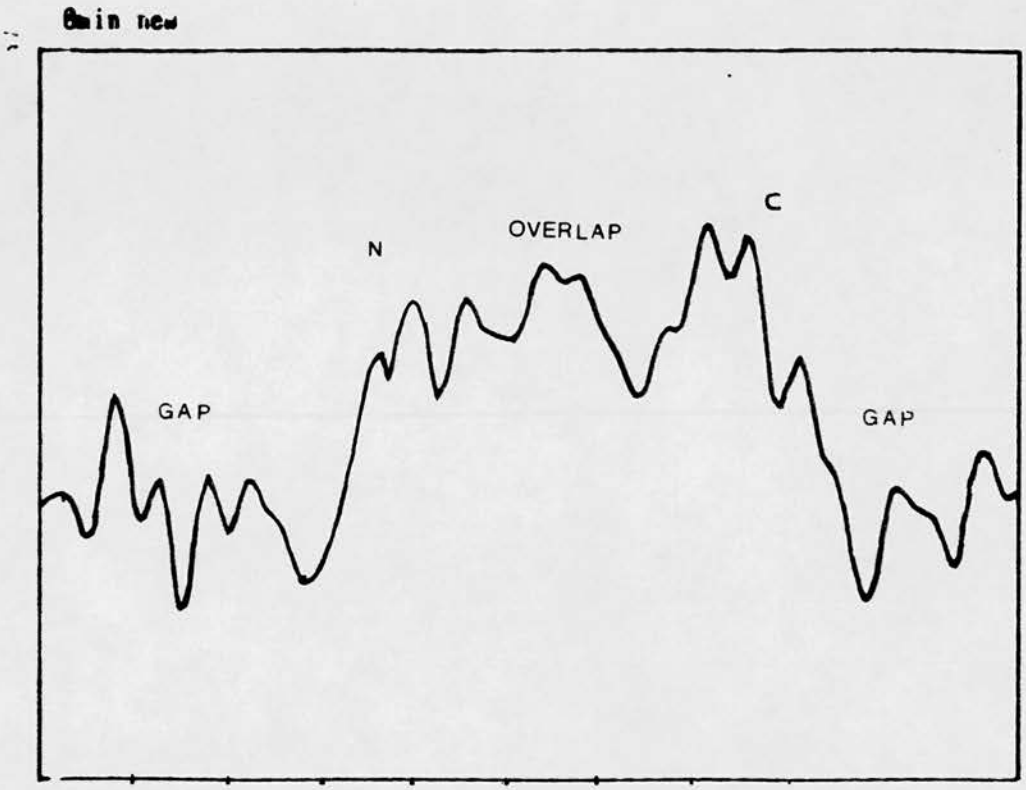
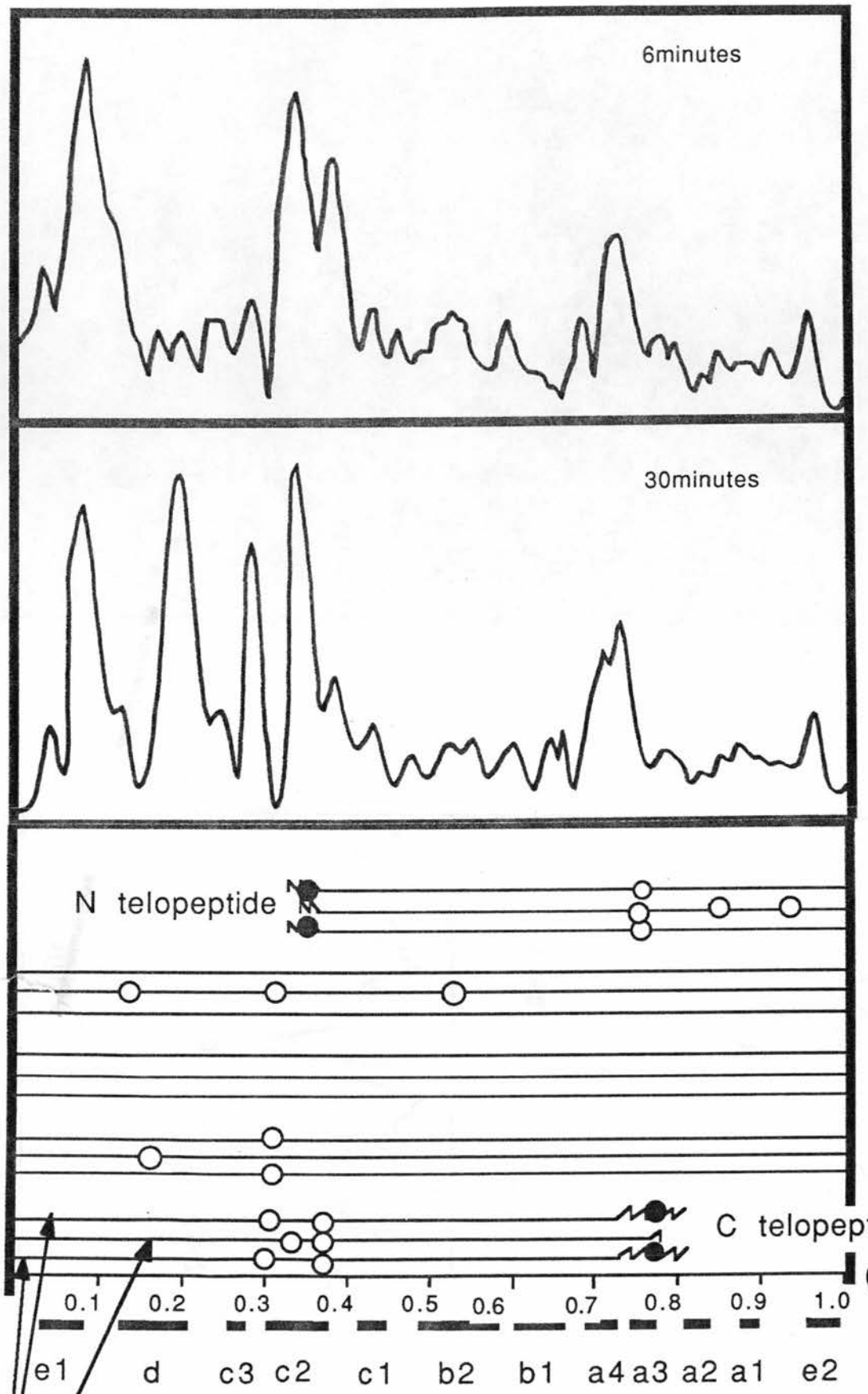


Figure 5.5 Fourier synthesis of the axially projected neutron scattering profile of native wet rat tail tendon. The X-axis represents one D-repeat of 670Å: the Y-axis is the net neutron scattering factor/molecule per unit axial translocation. The proposed position of the N and C telopeptides are shown by N and C, these correlate with the deuterium labelling positions at the N and C termini in Figure 5.6.



Positive staining bands found in collagen d repeat by electron microscopy, these correspond to parts of the structure containing charged amino acids and therefore lysine residues.

Figure 5.6 difference Fouriers calculated by subtracting native neutron scattering profiles from derivative scattering profiles. The peaks represent the distribution of deuterium labelling in the axial unit cell.

5.7 DISCUSSION

5.7.1 THE NATIVE SCATTERING PROFILE

The native profile of rat tail tendon as determined by neutron diffraction (Figure 5.5) shows an overlap:gap ratio of 0.49:0.51. This is in agreement with the value produced by the structural determination using X-rays (Chapter 4), within the limits of experimental error.

The profile shows a number of peaks and troughs superimposed on the gap overlap function. There is some degree of correlation with the banding pattern of positively stained electron micrographs, however the negative scattering factor associated with the hydrogen atom implies that the profile will not be directly related to that determined by X-ray diffraction or electron microscopy.

The amino acid residues of collagen with the highest scattering factor are those that contain an aromatic group. Of these, tyrosine is of special relevance to collagen since it is confined to the telopeptide regions at the interface of the gap and overlap regions.
The profile shown here has telopeptide regions that appear as peaks at the end of each gap region.

As in the case of X-rays it is difficult to compare the size of both telopeptide peaks. The scattering contribution of the amino acids from adjacent positions of the other segments in the axial structure will result in a different background. It is difficult therefore to judge

the contribution of the telopeptide region to this part of the profile.

The N terminal telopeptide does however appear as a smaller peak than that found at the C terminal. This can be related to the difference in the axial translation of residues in the N and C termini respectively.
The difference in the amount of tyrosine in the N and C termini (as judged from a composite rat, bovine and chick sequence), can also be related to the relative size of the peaks, since the C terminal telopeptides contain 9 tyrosine residues in total whilst the N-terminal telopeptides only contain 5 (Galloway 1982).

The native axial structure of collagen has been deduced by a process of model building, (Hulmes et al, 1980) and this structure shows a general similarity to the structure shown here, but to a lower resolution.

5.7.2 THE 6MINUTE $NaBD_4$ LABELLED DIFFERENCE MAP.

This difference map shows a series of discrete peaks. These can be related to the reducible components in collagen and their position in an axial projection of tendon structure. The difference map contains two peaks situated approximately 0.4D apart. The position of these peaks corresponds to the telopeptide position on the native neutron scattering profile, and since chemical analysis of tendon has revealed that crosslinking occurs in these regions, it can be assumed that it is the

crosslinks or the precursors of crosslinks that are being reduced here.

The map also reveals a series of peaks that cannot readily be related to the crosslinks commonly postulated by methods involving chemical degradation and analysis of the collagen. The possibility that hydroxylysine is modified by lysyl oxidase and subsequently deuterated in the procedures given in 5.4 is supported here.

The majority of the peaks in this difference map relate to the axial position of hydroxylysine residues in the D repeat. This also has the disadvantage of masking the labelling of crosslinks at the N terminal, since this part of the axial structure contains the majority of hydroxylysine residues.

The resolution of the map is insufficient to allow the separation of the N terminal crosslink(s) and the 8 surrounding hydroxylysine residues. The large amount of reduced hydroxyallysine, or crosslinks resulting from these, can account for the large peaks in the difference map at the N termini. The peak found in the region of the C termini probably corresponding to the crosslinks occuring at this site with some contribution being made by the hydroxyallysine derivatives of 3 hydroxylysine residues in this region. The degree of contraction in both types of telopeptide may also have some bearing on the amount of deuterium label at the telopeptide sites.

The more contracted nature of the C termini may have a blocking effect on the labelling of residues in this region and therefore make the diffusion of $NaBD_4$ slower in this region of the molecule.

The map also shows other peaks corresponding to the position of hydroxylysine residues in the gap region of the D repeat. The size of these peaks are large, and may be related to the ability of these residues to be more readily labelled than those in the telopeptide or overlap regions. The method of labelling here cannot distinguish between the reduced crosslinks and the reduced precursors of crosslink formation. Therefore it is impossible from these data to categorically assume that the reduced residues are crosslinks or precursors of crosslink formation.

There are a number of lysine residues in the gap region, some of which correspond to the peaks in the difference maps for example lysine 855 in the $\alpha 1$ chain. It is possible that lysine residues in these regions have been processed by lysyl oxidase, however there is little evidence for the existence of an intermolecular crosslink in the gap region.

The process of non enzymatic glycosylation may play a role here. The glycosylation of lysine or hydroxylysine residues in collagen allows these residues to be reduced by $NaBD_4$. Since this large peak in the 6 minute labelling difference map corresponds to residues in the gap region, it is possible that these residues have reacted with

interstitial sugars, and subsequently have been reduced. The discrete nature of the peak in the gap region indicates that if these peaks are the result of deuteration of a sugar linkage, the process may be more specific than expected.

5.7.3 THE 30 MINUTE $NaBD_4$ LABELLED DIFFERENCE MAP

This map contains a number of additional peaks to those already described in the 6 minute difference Fourier map.

The principal differences to this being the extra peaks adjacent to the N terminal telopeptide. In the axial amino acid sequence these correspond to a small group of lysine residues at position 434 in the $\alpha 1$ chain, and two hydroxylysine residues at positions 174 and 648 in the $\alpha 2$ amino acid sequence. In the 6 minute difference Fourier map, these peaks only correspond to small peaks. As discussed in the previous section, the way in which these residues have been altered to allow deuteration is a subject for conjecture.

With regard to the peak corresponding to the lysine residues at position 434, the existence of an intermolecular crosslink can be ruled out since there are no other residues in axial register that could support crosslink formation. The existence of an intramolecular crosslink at this point could be the reducible group, however the reactivity of this type of crosslink would

probably be the same as an intermolecular crosslink and therefore the relative size of this peak in the 30 minute difference Fourier map and the 6 minute difference Fourier map is difficult to explain.

The conclusion that can be drawn for both of the extra peaks, is that the reduction of the species corresponding to these peaks requires a significantly longer time. In the case of the peak corresponding to the hydroxylysine residues 174 and 648 of the $\alpha 2$ chain, the local environment of these residues may preclude their deuteration with short exposure to $NaBD_4$. Both these residues lie within the D band (electron microscopy notation), this region of the axial structure is thought to bind proteoglycan (Scott et al, 1988), this could cause the diffusion of $NaBD_4$ into this region to be slowed down and the deuteration would therefore only take place after longer periods of exposure. In the case of the peak corresponding to lys 434, a final possibility is that the main chain polypeptide backbone may have been broken at this site by the action of the reducing agent. This effect is only thought to occur after lengthy exposure of the sample to $NaBD_4$ or through the use of stronger reducing agents such as $LiAlH_4$ (Ruttenberg et al, 1964, Crestfeld et al, 1963).

5.8 CONCLUSION

The data presented here aid the structural investigation of collagen in tendon. The neutron scattering profile of tendon is described at the highest resolution so far obtained. The method of phase determination is unambiguous and therefore is not dependent on the constraints imposed by model building procedures. The profiles described show agreement with the generally accepted parameters of collagen's axial structure in tendon.

The data made available from the difference Fourier maps indicates that the labelling of collagen by $NaBD_4$ is specific, however the contribution of this study to the determination of crosslink position and the nature of other labelling sites can only be tentative.

CHAPTER 6

CONCLUSION

6.1 CONCLUSION

The experimental methods, data analysis and results shown here, give a detailed analysis of the axial structure of collagen in rat tail tendon. The techniques employed, X-ray and neutron diffraction and the method of phase determination, allow the structure of tendon to be determined to a high resolution and also give an insight into the *in situ* chemical reactivity of side groups in the constituent $\alpha 1$ and $\alpha 2$ chains of type I collagen.

The use of these two different diffraction techniques allows one to compare and contrast their uses on a specific molecule. The nature of neutron diffraction allows the position of the crosslinks and other reducible groups in tendon to be located.

The reduction of these residues with $NaBD_4$ provides a basis for the use of multiple isomorphous addition by deuteration. These reducible components cannot be easily identified by X-ray analysis, since these groups are not chemically reactive with heavy atom stains; as used in X-ray diffraction. It is essentially the chemical reactivity of these residues that is of importance, and the technique of neutron diffraction is ideal to study this, especially when complemented with evidence from chemical analysis.

The use of X-ray diffraction in this study provides a high level of resolution. This is one of the advantages of this technique with respect to neutron diffraction.

Neutron diffraction provides a resolution of D/23 as opposed to the resolution of D/50 provided here by X-ray diffraction. The technique also has the potential for an even more detailed study of the axial structure of tendon, since the first hundred meridional intensities can be observed.

The nature of labelling in the X-ray diffraction studies also provides information about the ability of various side groups especially tyrosine, histidine and methionine to react with specific heavy atoms.

The results given here can be compared and contrasted with other methods of investigation, to produce a fuller, less ambiguous understanding of the nature of tendon.

The methods of analysis shown here can be used where suitable in a number of related fields.
Phase determination by isomorphous replacement/addition can be used in other investigations into type I collagen, other collagens and other biological fibres.

The association of Na^{+}, K^{+}, and Ca^{2+}, ions with collagen in tendon can be studied by the introduction of heavy atom analogs of these, such as ions from the lanthanide series. This could be used to produce suitable derivatives for phase determination. Type II collagen as found in the lamprey notochord, provides an ideal subject for the techniques used here, since it has a regular axial structure that could be investigated by both X-ray and neutron diffraction. Type III collagen contains cysteine

residues, the labelling of these by mercurial reagents and $K_2^+PtCl_4^{2-}$ could provide a firm basis for the production of derivative samples.

The nature of diseased connective tissue can also be studied. The alterations in the amino acid sequence of collagen in various types of osteogenesis imperfecta allows the binding of heavy atoms to healthy and diseased tissue to be studied. X-ray diffraction has already been used to study tendon from patients with osteogenesis imperfecta (Bradshaw and Miller 1986).

The reaction of collagen with sugars when the concentration of interstitial sugars is raised, as in diabetes could be studied by neutron diffraction. The adducts formed by the reaction of glucose with lysine and hydroxylysine could be located since these can be reduced by $NaBD_4$. The in vitro simulation of this disorder can also be studied for example by the reaction of ribose and related sugars with collagenous tissues as conducted by Tanaka et al, (1988). This would provide a new series of isomorphous derivatives for phase determination and also allow the specificity of attachment of sugars to lysine and hydroxylysine to be studied. X-ray diffraction also provides the potential for the structural determination of collagen packing in three dimensions. The analysis of off meridional Bragg peaks in native and a series of derivative samples can be used to determine the structure of collagen in a triclinic unit cell.

The nature of heavy atom labelling aids this study in 3 ways.

1) The method of study provides unambiguous phase determination for each **hkl** reflection.

2) The information obtained from the axial packing of collagen in tendon and the reactivity of specific heavy atom stains with residues in collagen, gives a starting point for the study and proposed structures in three dimensions can be analysed in the light of these results.

3) The method of heavy atom labelling enhances the signal to noise ratio of the off meridional rowlines and therefore makes the measurement of these intensities easier and more reliable.

It is hoped that the methods of analysis developed in research for this thesis can be used in the future structural analysis of connective tissue.

BIBLIOGRAPHY

ASTBURY, W. T. (1933) Trans. Faraday Soc. 29, 193-211.

BACON, G. E. (1975), Neutron Diffraction. Clarendon Press Oxford.

BAILEY, A. J. (1968) Biochim. Biophys. Acta. 160, 447-453.

BAILEY, A. J. and PEACH, C. M. (1968) Biochem. Biophys. Res. Comm. 33, 812-819.

BAILEY, A. J., ROBINS, S. P. and BALIAN, G. (1974) Nature 251, 105-109.

BAILEY, A. J. and SHIMOKOMAKI, M. (1971) FEBS Lett. 16, 86-88.

BALE, W. F., HELMKAMP, R. W., DAVIS, T. P., IZZO, M. J., GOODLAND, R. L., CONTRERAS, M. A. and SPAR, I. L. (1966) Proc. Soc. Exp. Med. 122, 407-414.

BARNARD, K., LIGHT, N. D., SIMS, T. J. and BAILEY, A. J. (1987) Biochem. J. 244, 303-309.

BEAR, R. S. (1942) J. Am. Chem. Soc. 64, 727.

BEAR, R. S. (1944) J. Am. Chem. Soc. 66, 1297-1305.

BEAR, R. S. (1952) Advances in Protein Chemistry 7,69-160

BENDER, E., SILVER, F., HAYASHI, K.(1983) Coll. Rel. Res. 3, 407-418.

BERG, R. A. and PROCKUP, D. J. (1973) Biochem. Biophys. Res. Comm. 52, 115-120.

BERNARD, M., MYERS, J., CHU, M., RAMIREZ, F., EIKENBERRY, E. F. and PROCKOP, D.(1983) Biochemistry 22, 1139-1145.

BERNSTEIN, P. H. and MECHANIC, G. L. (1980)
J. Biol. Chem. 255, 10414-10422.
BIGI, A., RIPAMONTI, A., ROVERI, N., GARBISA, S., GOTTE, L. and VOLPIN, D. (1977) Conn. Tiss. Res. 5, 37-39.
BIRK, D. E., FITCH, J. M., BABIARZ, J. P. and LINSENMAYER, T. F. (1988) J. Cell Biol. 106, 999-1008.
BLUNDELL, T. L. and JOHNSON, L. N. (1976) Protein crystallography, Academic press, New York.
BOEDTKER, H. and DOTY, P. (1956) J. Am. Chem. Soc. 78, 4267-4280.
BORDIGNO, E., CATALIN, L., NATALIE, G. and SCATTURI, A. (1973) Chem. Comm. 22, 878-879.
BORNSTEIN, P., (1969) Biochemistry 8, 63-71.
BORNSTEIN, P.and TRAUB, W. (1979) in: The Proteins., NEURATH, H. and HILL, R., Eds., London, A. P.,411-632.
BOWES, J. and KENTEN, R.,(1948) Biochem. J. 43, 358-365
BRADSHAW, J. P. AND MILLER, A.(1985) Annals of the Rheumatic Diseases. 45,750-756.
BRODSKY, B.and EIKENBERRY, E. F. (1982) Methods in Enzymology 82, 127-174.
CAPALDI, M. J. and CHAPMAN, J. C. (1983) Biopolymers 21 2291-2313.
CHAPMAN, J. C. and HULMES, D. J. S. (1984) in: Ultrastructure of the connective tissue matrix., RUGGERI, A. and MOTTA, P. M., Eds., Boston, Martinus Nijhoff 1-33.
COCHRAN,W., CRICK, F.,VAND, V.(1952) Act. Cryst. 5,581-6.
COHEN,C. and BEAR, R.S.(1953) J.Am.Chem. Soc. 75, 2783-4.

COWAN, P. M., McGAVIN, S. and NORTH, A. C. T. (1955) Nature 176, 1062-1064.

COWAN, P. M., NORTH, A. C. T. and RANDALL, J. T. (1953) in: Nature and Structure of Collagen, RANDALL, J. T., Ed., London, Butterworth, 241.

CRAIG, A. S. and PARRY, D. A. D. (1981) J. Ultrastruct. Res. 74, 232-239.

CRESTFELD, A. M., MOORE, S. and STEIN, W. H. (1963) J. Biol. Chem. 238, 622-627.

DAVIDSON, P. F. and BRENNAN, M. (1982) Conn. Tiss. Res. 11, 135-151.

DION, A. S. and MYERS, J. C. (1987) J. Mol. Biol. 193, 127-143.

DOYLE, B. B., HUKINS, D. W. L., HULMES, D. J. S., MILLER, A. and WOODHEAD-GALLOWAY, J. (1975) J. Mol. Biol. 91, 79-99.

DOYLE, B. B., HULMES, D. J. S., MILLER, A., PARRY, D. A. D., PIEZ, K. A. and WOODHEAD-GALLOWAY, J. (1974) Proc. R. Soc. Lond. B. 186, 67-74.

EIKENBERRY, E. F., CHILDS, B., SHEREN, S. B., PARRY, D. A. D., CRAIG, A. S. and BRODSKY, B. (1984) J. Mol. Biol. 176, 261-277.

ELLIS, D. O. and McGAVIN, S. (1970) J. Ultrastruct. Res. 32, 191-211.

ERICSON, L. G. and TOMLIN, S. G. (1959) Proc. R. Soc. Lond. A. Math. Phys. Sci. 252, 197-216.

EWALD, P. P. (1921) Z. Kristollagr. Miner. 56, 129-135.

EYRE, D. (1987) Methods in Enzymology 144, 115-140.

HART, R. W. and FARRELL, R. A. (1969) J. Optical Society of America 59, 766-774.

FIETZEK, P. P. and KUHN, K. (1975) Molecular & cellular Biochemistry 8, 141-157.

FOSTER, J. A., BRUENGER, E., GRAY, W. R. and SANDBERG, L. B. (1973) J. Biol. Chem. 248, 2876-2879.

FRASER, R. D. B., MACRAE, T. P., CHEW, M. W. K. and SQUIRE, J. M. (1987) in: Fibrous Protein Structure, SQUIRE, J. M. and VIBERT, P. J., Eds., London, 173-191.

FRASER, R. D. B., MACRAE, T. P. and MILLER, A. (1987) J. Mol. Biol. 193, 115-125.

FRASER, R. D. B., MACRAE, T. P., MILLER, A. and SUZUKI, E. (1983) J. Mol. Biol. 167, 497-521.

FRASER, R. D. B., MACRAE, T. P. and SUZUKI, E. (1979) J. Mol. Biol. 129, 463-481.

GALLOWAY, D., in: Collagen in Health and Disease., WEISS, J. B. and JAYSON, M. I. V. Eds,. Edinburgh, Churchill Livingstone.

GELMAN, R. A. and PIEZ, K. A. (1980) J. Biol. Chem. 255, 8098-8102.

GREEN, D. W., INGRAM, V. M. and PERUTZ, M. F. (1954) Proc R. Soc. Ser. A. 225 287-307.

HALL, C. E. and DOTY, P. (1958) J. Am. Chem. Soc 80, 1269-1274.

HALPERN, O. and JOHNSON, M. H. (1939) Phys. Rev. 55, 898-923.

HARKER, D. (1956) Acta Cryst 9, 1-11.

HARKNESS, M. L. R., HARKNESS, R. D. and JAMES, D. W. (1958) J. Physiol. 144, 307-313.

HASCALL, V. C. and KIMURA, J. H. (1982) Methods in Enzymology 82, 769-800.

HAWORTH, R. A. and CHAPMAN, J. A. (1977) Biopolymers 16, 1895-1906.

HELSETH, D. L., LECHNER, J. H. and VEIS, A. (1979) Biopolymers 18, 3005-3014.

HNATOWICH, D. J., LAYNE, W. W. and CHILDS, R. L. (1982) Int. J. Appl. Radiat. Isot. 33, 327-332.

HODGE, A. J. and PETRUSKA, J. A. (1963) in: Aspects of Protein Structure, RAMACHANDRAN, G. N., Ed., London and New York, Academic Press, 289-300.

HODGE, A. J. and SCHMITT, F. O. (1960) Proc. Nat. Acad. Sci. USA 46, 186-197.

HOFMANN, H., FIETZEK, P. P. and KUHN, K. (1980) J. Mol. Biol. 141, 293-314.

HOFMANN, H., VOSS, T., KUHN, K. and ENGEL, J. (1984) J. Mol. Biol. 172, 325-343.

HULMES, D. J. S., JESIOR, J. C., MILLER, A., BERTHET-COLOMINAS, C. and WOLFF, C. (1981) Proc. Nat. Acad. Sci. USA. 78, 3567-3571.

HULMES, D. J. S. and MILLER, A. (1979) Nature 282, 878-880.

HULMES, D. J. S., MILLER, A., PARRY, D. A. D., PIEZ, K. A. and WOODHEAD-GALLOWAY, J. (1973) J. Mol. Biol. 79, 137-148.

HULMES, D. J. S., MILLER, A., WHITE, S. W. and BRODSKY DOYLE, B. (1977) J. Mol. Biol. 110, 643-666.

HULMES, D. J. S., MILLER, A., WHITE, S. W., TIMMINS, P. A. and BERTHET-COLOMINAS, C. (1980) Int. J. Biol. Macromol. 2, 338-345.

HUKINS, D. W. L. (1984) in: Collagen in Health and Disease., WEISS, J. B. and JAYSON, M. I. V., Eds., Edinburgh, Churchill Livingstone.

HUXLEY, H. E. and BROWN, W. (1967) J. Mol. Biol. 30, 383-434.

JESIOR, J. C., MILLER, A. and BERTHET-COLOMINAS, C. (1980) FEBS Lett. 113, 238-240.

JONES, E. Y. and MILLER, A. (1987) Biopolymers 26, 463-480.

JOSSE, J. and HARRINGTON, W. F. (1964) J. Mol. Biol. 9, 269-287.

KAESBERG, P. and SHURMAN, M. (1953) Biochim. Biophys. Acta 11, 1-6.

KASTELIC, J., GALELSKI, A. and BAER, E. (1978) Conn. Tiss. Res. 6, 11-23.

KREJCAREK, G. E. and TUCKER, K. L. (1977) Biochem. Biophys. Res. Comm. 77,581-585.

KUHN, E. M., MARENUS, K. D. and BEER, M. (1984) J. Ultrastruct. Res. 87, 172-179.

LIEBOVICH, S. J. and WEISS, J. B. (1970) Biochim. Biophys. Acta. 214, 445-454.

MANDELKOW, E., MANDELKOW, E.M., BORDAS, J. Trends in Biochemical Sciences. 8, 374-377.

MCFARLANE, A. S. (1958) Nature 182, 53.

MEARES, C. F., GOODWIN, D. A., LEUNG, C. S-H., GIRGIS, A. Y., SILVESTER, D. J., NUNN, A. D. and LAVENDER, P.J. (1976) Proc. Nat. Acad. Sci. (USA) 73, 3803-3806.

MEEK, K. M., CHAPMAN, J. A. and HARDCASTLE, R. A. (1979) J. Biol. Chem. 254, 10710-10714.

MILLER, A., BRODSKY DOYLE, B., HULMES, D. J. S., JENKINS, G. T., WHITE, J. W., HAAS, J., IBEL, K. and TIMMINS, P. (1975) Proceedings of the 1975 Brookhaven Symposium 27, 86-100.

MILLER, A. and WRAY, J. S. (1971) Nature 230, 437-439.

MILLER, E. J. (1971) Biochemistry 10, 1652-1659.

MILLER, E. J. (1976) Molecular & Cellular Biochemistry 13, 165-192.

MILLER, E. J. and GAY, S. (1987) Methods in Enzymology 144, 3-41.

MILLER, E. J. and MATUKAS, V. J. (1969) Proc. Nat. Acad. Sci. USA. 64, 1264-1268.

MOORE, P. B., ENGLEMAN, D. M., LANGER, J. A., RAMAKRISHNAN, V. R., SCHINDLER, D. G., SCHOENBORN, B. P., SILLERS, I. Y. and YABUKI, S. (1984) in: Neutrons in Biology, SCHOENBORN, B. P., Ed., New york, Plenum Press, 25-46.

MORGAN, P. H., JACOBS, H. G., SEGREST, J. P. and CUNNINGHAM, L. W. (1970) J. Biol. Chem. 245, 5042-5048.

NAKAMURA, Y. (1987) Int. J. Biol. Macromol. 9, 281-290.

NAVE, C., HELLIWELL, J. R., MOORE, P. R., THOMPSON, A. W., WORGAN, J. S., GREENALL, R. J., MILLER, A., BURLEY, S. K., BRADSHAW, J. P., PIGRAM, W. J., FULLER, W., SIDDONS, D. P., DEUTSCH, M. and TREGEAR, R. (1985)
J. Appl. Chrystallog. 18, 396-403.

NEMETSCHEK, T., GRASSMAN, W. and HOFFMAN, V. (1955)
Z. Naturforsch. B. 10, 61-68.

NOLD, J. G., KANG, A. H. and GROSS, J. (1970) Science 170, 1096-1098.

NORTH, A. C. T., COWAN, P. M. and RANDALL, J. T. (1954)
Nature 174, 1142-1143.

OGSTON, A. G. (1970) in: Chemistry and Molecular Biology of the Intracellular Matrix Vol 3, BALAZS, E. A., Ed., London, Academic Press, 1231-1240.

OTTER, A., SCOTT, P. G. and KOTOVYCH, G. (1988)
Biochemistry 27, 3560-3567.

(a)PARRY, D. A. D., BARNES, G. R. and CRAIG, A. S. (1978)
Proc. R. Soc. Lond. B. 203, 305-321.

(b)PARRY, D. A. D., CRAIG, A. S. and BARNES, G. R. (1978)
Proc. R. Soc. Lond. B. 203, 293-304.

PERKINS, S. J. (1988) New Comprehensive Biochemistry 11b, 143-265.

PETSKO, G. A., PHILLIPS, D. C., WILLIAMS, R. J. P. and WILSON, I. A. (1978) J. Mol. Biol 120, 345-359.

PIEZ, K. A. (1976) In: Biochemistry of collagen., RAMACHANDRAN, G. N. and REDDI, A. H., Eds., New York, Plenum Press, 1-44.

PIEZ, K. A., EIGNER, E. A. and LEWIS, M. S. (1963) Biochemistry 2, 58-65.

PIEZ, K. A., LEWIS, M. S. and MARTIN, G. R. (1961) Biochim. Biophys. Acta. 53, 596-599.

PIEZ, K. A. and TRUS, B. L. (1981) Bioscience Reports 1, 801-810.

POPE, F. M. and NICHOLLS, A. C. (1984) in: Molecular Medicine, MALCOLM, A. D. E., Ed., IRL Press, 117-175.

PRESTON, B. N., SNOWDEN, J. M. and HOUGHTON, K. T. (1972) Biopolymers 11, 1645-1659.

RAMACHANDRAN,G.N.and KARTHA,K.(1954) Nature 174, 269-270.

RICH, A. and CRICK, F. W. C. (1955) Nature 176, 915-916.

ROBINS, S.P. (1982) Meth. Biochem. Analysis 28, 329-379.

ROBINS, S.P. BAILEY, A.J. (1973) Biochem. J. 135 656-665

ROBINS, S. P. and BAILEY, A. J. (1975) Biochem. J. 149, 381-385.

ROSENBERG, H., MODRAK, J. B., HASSING, J. M., AL-TURK, W. A. and STOHS, S. J. Bioc. Biophys. Res. Comm. 91,498-501.

ROSENBLOOM, J. (1987) Methods in Enzymology 144, 172-196.

ROWE, R. W. D. (1985) Conn. Tiss. Res. 14, 9-20.

RUTTENBERG, M. A., KING, T. P. and CRAIG, L. C. (1964) Biochemistry 3, 758-764.

SCHMITT, F. O., GROSS, J. and HIGHBERGER, J.H. (1955) Exp. Cell. Res. Suppl. 3, 326-330.

SCHMITT, F. O., HALL, C. E. and JAKUS, M. A. (1942) J. Cell. Comp. Physiol. 20, 11-33.

SCOTT, J.E. and HAIGH, M.(1985) Bioscience Rep. 5, 71-81.

SCOTT, J. E. and HAIGH, M. (1988) Biochem. J. 253,

607-610.

SHERWOOD, D. (1976) Crystals, X-Rays and Proteins. Longmans. London.

SMITH, J. W. (1968) Nature 219, 157-158.

SPADARO, J. A. (1970) Nature 228, 78-79.

SPIRO, R. G. (1969) J. Biol. Chem 244, 602-612.

STINSON, R. H., BARTLETT, M. W., KURG, T., SWEENY, P. R. and HENDRICKS, R. W. (1979) Biophys.J. 26, 209-222.

SUNDBERG, M. W., MEARES, C.F., GOODWIN, D. A. and DIAMANTI, C. I. (1974) J. Med. Chem. 17 1304-1307.

SVENDSEN, K.H. and KOCH, M. H. (1982) EMBO J. 1, 669-674.

SVENDSEN, K. H., THOMSON, G. and BARTELS, K. (1983) Int. J. Biol. Macromol. 5, 204-208.

TANAKA, S., AVIGAD, G., BRODSKY, B., EIKENBERRY, E. (1988) J. Mol. Biol. 203, 495-501.

THOMSON, A. J. (1974) Rec. Result. Cancer Res. 48, 38-62.

THOMSON, A. J., WILLIAMS, R. J. P. and RESLOVA, S. (1972) Structure Bonding 11, 1-46.

TKOCZ, C. and KUHN, K. (1969) Eu. J. Biochem. 7, 454-462.

TOMLIN, S. G. and WORTHINGTON, C. R. (1956) Proc. R. Soc. Ser. A 235, 189-201.

TRAUB, W. and PIEZ, K. A. (1971) Advances in Protein Chemistry 25, 243-352.

TRUS, B. L. and PIEZ, K. A. (1976) J. Mol. Biol. 108, 705-732.

VEIS, A., MILLER, A., LEIBOVICH, S. J. and TRAUB, W. (1979) Biochimica et Biophysica Acta 576, 88-98.

VON DER MARK, K., WENDT, P., REXRODT, F. and KUHN, K. (1970) FEBS Lett. 11, 105-108.

WANG, S. K. and WORTHINGTON, C. R. (1975) Biophys. J. 15, 322a.

WEISS, J. B. and JAYSON, M. I. V., Eds. (1982) Collagen in Health and Disease, Churchill Livingstone, Edinburgh.

WHITE, S. W. (1977) Oxford, Thesis, D.phil.

WHITE, S. W., HULMES, D. J. S., MILLER, A. and TIMMINS, P. A. (1977) Nature 266, 421-425.

WHITTAKER, E. J. W. (1952) Acta Cryst. 6, 222-223.

WOOLFSON, M. M. (1970) X-ray Crystallography, Cambridge University Press, Cambridge.

YONATH, A. and TRAUB, W. (1969) J. Mol. Biol. 43, 461-477.