HYDROXYLYSINE GLYCOSIDES OF CORNEAL COLLAGEN

HYDROXYLYSINE GLYCOSIDES OF CORNEAL COLLAGEN

BY

JAMAL IBRAHIM

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE UNIVERSITY OF OXFORD

1986

CONTENTS

	Page
Acknowledgements	(i)
Abstract	(ii)
Chapter 1 : Introduction	1
I. Cornea	1
II. Collagen	6
III. Introduction to the present work	33
Chapter 2 : Experimental materials and methods	36
I. Materials	36
II. Methods	37
Chapter 3 : Isolation and characterisation of some cyanogen bromide peptides of bovine corneal collagen	51
I. Purification and characterisation of corneal collagen	51
II. Cyanogen bromide peptides of bovine corneal collagen	53
III. Relative electrophoretic mobilities of peptides α1-CB7 and α1-CB8 on SDS gels	59
Chapter 4 : Pinpointing the sites of hydroxylysine glycosides in peptide α1-CB7 of bovine corneal collagen	65
I. Tryptic peptides of citraconylated peptide α1-CB7	66
II. V8-protease peptides of peptide α1-CB7	76
III. Sites of hydroxylysine and hydroxylysine glycosides in peptide α1-CB7 of bovine corneal collagen: A summary	96
Chapter 5 : Discussion	102
References	
Appendix	

Acknowledgements

I would like to thank Dr. John Harding for his supervision and friendship. I also wish to thank A.J. Bron for approving this work to be carried out at the Nuffield Laboratory of Ophthalmology.

For his attention to the amino acid analyser (and companionship) I would like to express many thanks to Ken Rixon, head technician.

I thank (Dr.) Howard Beswick for discussing some of the techniques as well as Anne Colwin for doing some electron-microscopy. I am grateful to Tony Gascoyne at 'Immunochemistry' for carrying out some amino acid analysis for me when our analyser was being serviced.

With regard to the typing of this thesis, my appreciation goes to Ian and John at 'Daily Information' for their advice and making their equipment available for hire.

For grants and financial support, my gratitude goes to the Royal National Institute for the Blind and the government of Malaysia.

I dedicate this thesis to my parents, sister and brothers, as well as Kathy and all my friends in England, Jamaica and Malaysia.

Title: Hydroxylysine glycosides of corneal collagen.

Author: Jamal Ibrahim

College: New College, University of Oxford.

Submitted for the degree of Doctor of Philosophy, Trinity term, 1986.

ABSTRACT

The universally found thin and uniform features of collagen fibrils constituting the corneal stroma (Craig and Parry, 1981), are essential for corneal transparency (Goldman et al., 1968). Corneal collagen contains more hydroxylysine−glycoside than most collagens. In relation to the 'D−stagger' arrangement of collagen molecules in the fibril, it was postulated that an excess of hydroxylysine−glycoside occurring in the 'overlap' region may limit fibril growth in the lateral direction by steric−hindrance (Morgan et al., 1970).

The ultimate aim of this work was to pinpoint the sites of hydroxylysine−glycosides in the corneal collagen molecule, bearing in mind their location with respect to the fibril gap/overlap structure. Determination of hydroxylysine−glycosides were carried out on an automated amino acid analyser with the much reduced total analysis time of 1¼ hours per sample, using a two−buffer elution system developed during this work. Cyanogen bromide peptides α1−CB3, α1−CB7 and α1−CB8 were isolated and shown to contain 1.0, 0.97 and 1.21 residues of hydroxylysine−glycoside respectively per peptide.

It was also proven beyond doubt that peptide α1−CB8 has a higher electrophoretic mobility on SDS−gels than peptide α1−CB7, pointing to mistakes found in the literature.

In the 271−residue peptide α1−CB7, analysis of tryptic and V8−protease peptides led to the discovery of seven glycoside sites, the major site being at lysyl position 684 (numbers from start of α1[I] helical sequence). This lysyl position was found to be 46% hydroxylysine−glycoside, being 34% glucosylgalactosyl−hydroxylysine and 12% galactosylhydroxylysine. Of position 564, 17% was hydroxylysine−glycoside while of the other positions 573, 603, 648, 657 and 756, between 5% and 12% was hydroxylysine−glycoside. In relation to the fibril gap/overlap structure, position 756 is in an overlap zone and position 564 is just inside one. Positions 573 and 684 are just inside a gap zone which also contains the other glycoside sites. Another possible site at position 729 was not found to contain hydroxylysine nor therefore, any glycoside.

These findings are discussed with respect to the possible role of hydroxylysine−glycosides in limiting collagen fibril diameter. Comparisons of the amino acid sequences around the seven glycoside sites however gave no clues as to what makes some lysyl residues more susceptible to modification than others. The possible reasons for the high extent of lysyl modification in the cornea are also discussed.

CHAPTER 1

INTRODUCTION

I. CORNEA

I A. GROSS ANATOMY AND FUNCTIONS

The cornea is the tough transparent window at the front of the eye (fig. 1.1). It refracts light, contributing nearly three quarters of the dioptric power of the eye and transmits this light with minimal scatter or absorption. In addition to this optical function, together with the sclera it protects the eye from mechanical injury and helps to maintain the shape of the eye by withstanding the intraocular pressure.

The adult human cornea is almost circular, with a diameter of 11 to 12 mm. It is about 0.52 mm thick at the centre (Maurice and Giardini, 1951a; Donaldson, 1966), thickening to about 0·65 mm in the limbal region (Martola and Baum, 1968), with insignificant variations between sexes or with increasing age. The human cornea consists of five principal layers, namely the epithelium, Bowman's layer, stroma, Descemet's membrane, and endothelium (fig. 1.2). The Bowman's layer is not usually present in the cornea of non−primates, including the 0.8 mm thick bovine cornea.

I B. CORNEAL STROMA

The stroma represents 90% of the thickness of the cornea. Research into the nature of corneal transparency is thus directly related to the origins and maintenance of stromal transparency.

1. Structure and biochemistry of the stroma

The stroma consists of sheets or lamellae of collagenous material lying parallel to the corneal surface. The collagen fibrils in each lamella make large angles to those in adjacent lamellae above and below. Electron micrographs of corneal sections (Jakus, 1964) reveal layers of lines which are longitudinal collagen fibrils, alternating with layers of dots which are transverse collagen fibrils (fig. 1.3). There are about 300 and

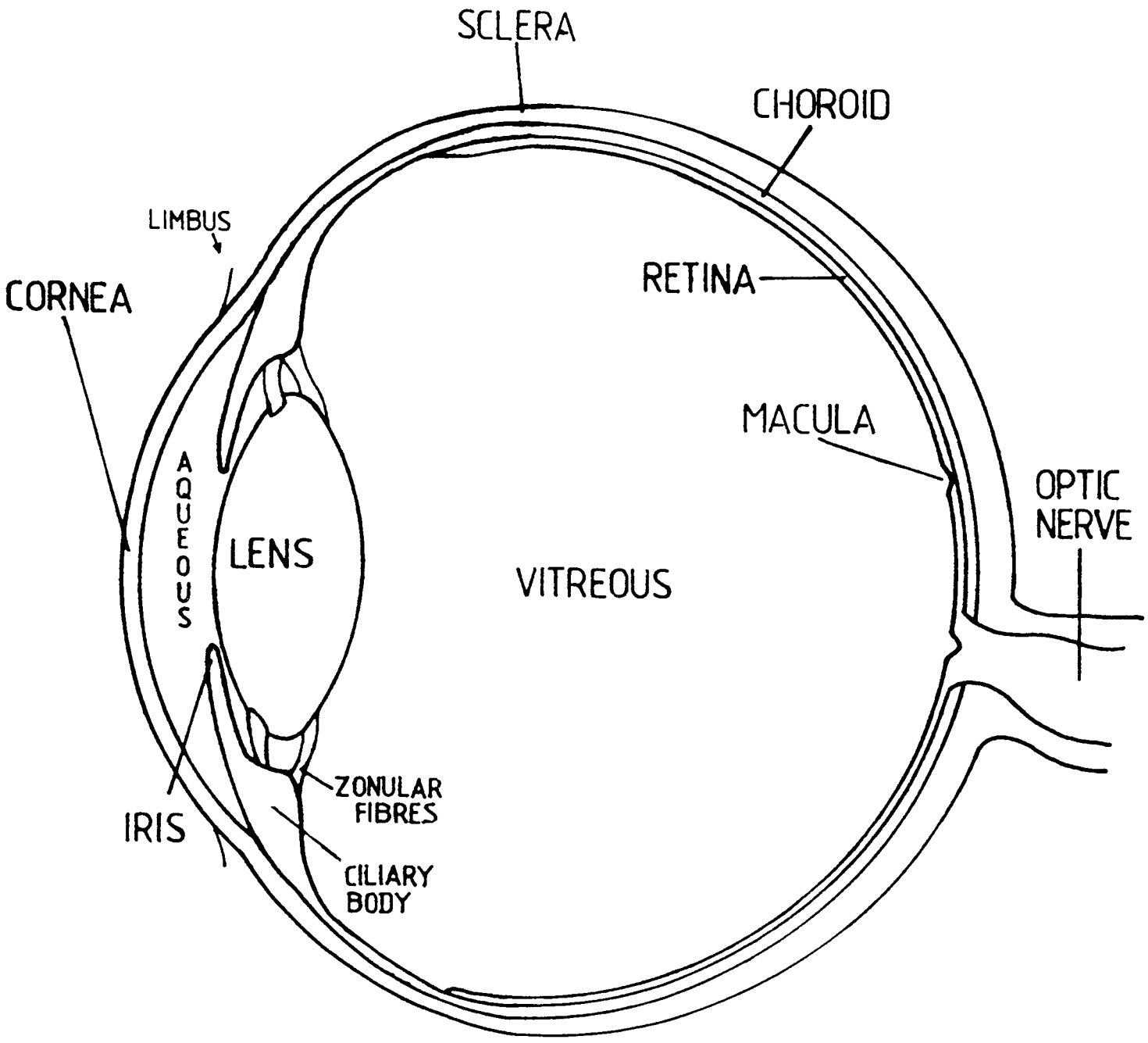


Figure 1.1 Diagrammatic section of an eye. Modified from *Krause (1934)*.

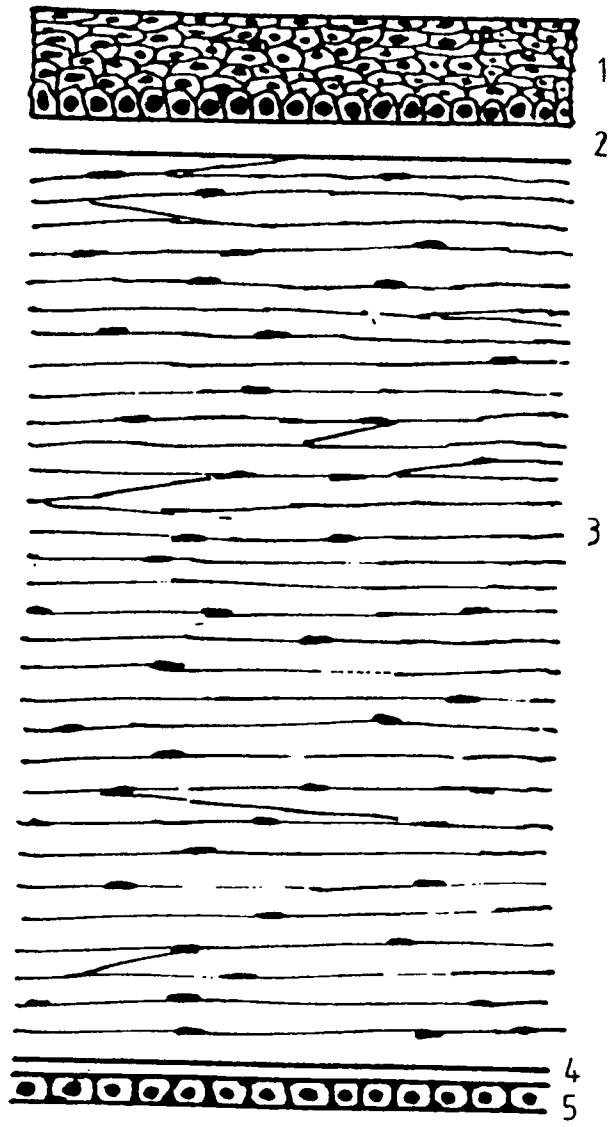


Figure 1.2 Diagrammatic section through the depth of the cornea: *1*, epithelium; *2*, Bowman's layer; *3*, stroma; *4*, Descemet's membrane; *5*, endothelium. (From *Krause, 1934*).

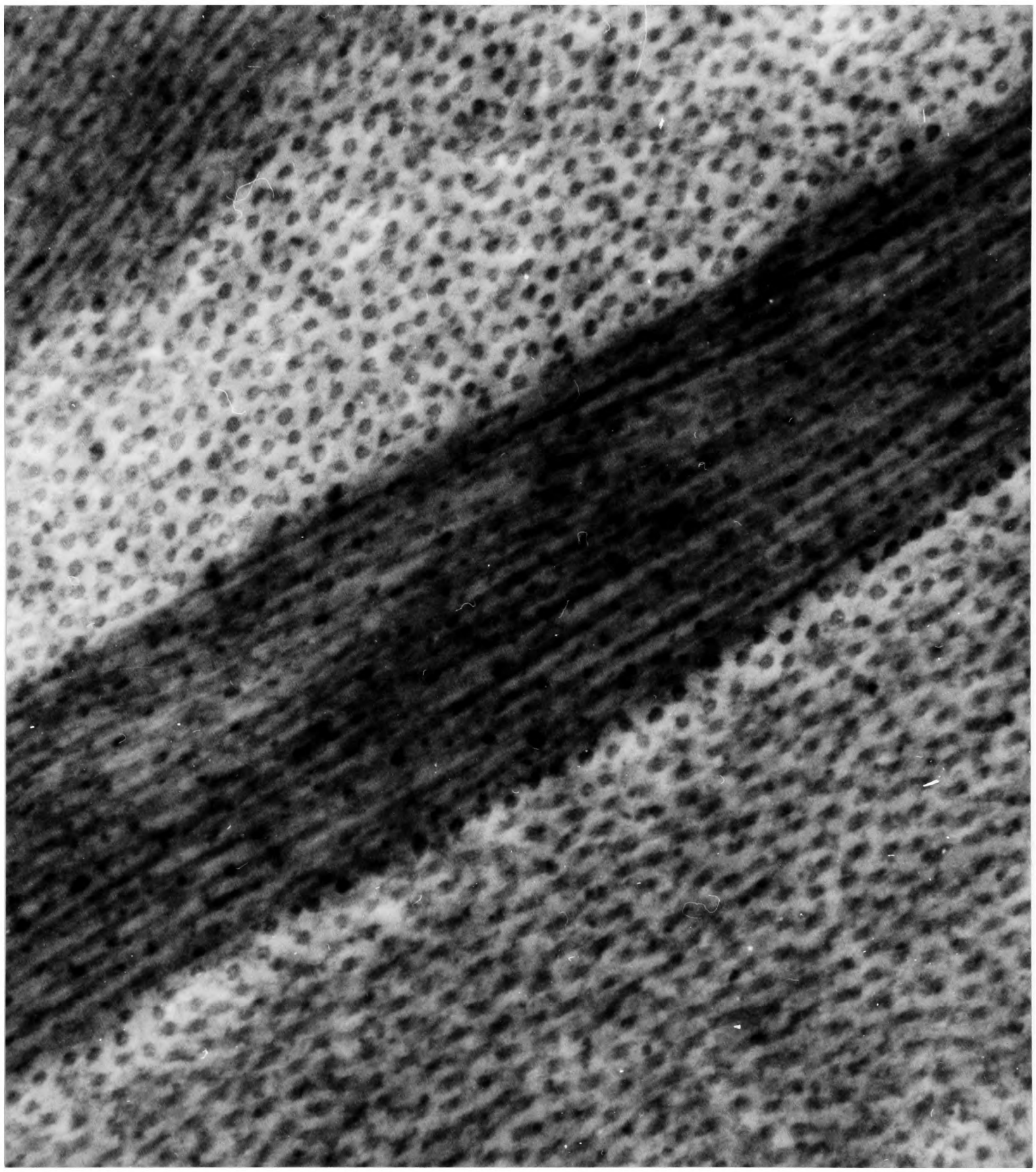


Figure 1.3 Electronmicrograph of a section through rat corneal stroma. (Prepared by *Anne Colwin*).

500 lamellae through the thickness of the central and peripheral human cornea respectively. This striking architecture may originate from the cell−surface assembly of collagen fibrils by specifically oriented corneal stromal keratocytes (Birk and Trelstad, 1984). Keratocytes, analogous to the fibroblasts of skin, are dispersed throughout the corneal stroma, constituting 2 to 3% of the stromal volume.

Of the corneal dry weight, 70% is the protein collagen, while most of the remainder consists of hydrophilic proteoglycans (Maurice, 1984). Water makes up about 77% of the corneal wet weight. The corneal stroma can thus be seen as a lamellar organisation of collagen fibrils in a hydrated ground substance of hydrophilic proteoglycans.

2. Theories of corneal transparency

The cornea transmits light of the visible range, wavelength 400−700 nm with minimal absorption and scatter. The transmissivity of the rabbit cornea for example, is 98% at 700 nm falling to 80% at 400 nm (Farrell et al., 1973).

The light refractive index of dry collagen is significantly higher than that of the ground substance, and that of the stroma as a whole. Maurice (1957), calculating on the basis that the collagen fibrils scatter light individually of one another, stated that the human corneal stroma should be opaque as it would scatter more than 90% of a beam of light, wavelength 500 nm. He thus put forward the lattice theory of corneal transparency (Maurice, 1957). According to this theory, collagen fibrils of the corneal stroma are arranged in a crystalline lattice within each lamella. By analogy with a diffraction grating, scattering of light by individual fibrils is suppressed by destructive interference. Thus, only light travelling in the direction of the original incident beam would be transmitted. However, the required crystalline order has not been observed in electron micrographs (Schwarz and Graf Keyserlingk, 1969). Furthermore the random architecture of the thick Bowman's layer of the shark cornea does not result in an opaque cornea (Goldman and Benedek, 1967).

A more accepted theory of corneal transparency is based on a quasi−regular

structure with short−range order (Goldman et al., 1968; Benedek, 1983; Sayers et al., 1982). The theory states that, where fluctuations in refractive index occur over distances less than half the wavelength of visible light i.e. about half of 500 nm (green light), the principles of diffraction precludes any scattering of light in the forward direction.

3. Short−range order of the collagen fibrils in the corneal stroma

The uniform interfibrillar spacing of thin collagen fibrils of uniform diameter constitute the short−range order of the corneal stroma.

a) Thin, uniform collagen fibrils

Collagen fibrils of the corneal stroma are consistently about 25 nm thick in vertebrate species such as fish, amphibians, reptiles, birds and mammals including humans, with very little variation across the corneal thickness (Craig and Parry, 1981). In contrast, collagen fibrils of the opaque corneal scar are, on average, thicker and of a wider range of diameters (Schwarz and Graf Keyserlingk, 1969). Similarly, the naturally opaque sclera contains fibrils which are of varying diameters of the range 30 to 300 nm (Goldman and Benedek, 1967; Maurice, 1984). The regulation of collagen fibril diameters is the central issue of the current project and will be discussed later (Chapter 1.IIE).

b) Uniform interfibrillar spacing

Collagen fibrils of the corneal stroma are uniformly interspaced at centre−to−centre distances of 55 to 60 nm (Jakus, 1961; Goldman et al., 1968). The hydrophilic proteoglycans of the stromal ground matrix are thought to maintain the uniform spacing.

Proteoglycans are the major ground matrix components of collagenous connective tissues such as cartilage, tendon, bone, skin, sclera and cornea (see reviews by Hascall and Kimura, 1982; Poole, 1986). They constitute a heterogeneous family of macromolecules with different sizes, determined by the size of their protein cores, the

number and lengths of covalently-attached negatively charged glycosaminoglycan side-chains and their state of aggregation. The corneal stroma contains two types of proteoglycans which, in contrast to those of cartilage, are small and non-aggregating. The main type has keratan sulphate side-chains while the remainder has chondroitin- or dermatan-sulphate side-chains attached to relatively small protein cores (Axelsson and Heinegard, 1975). In the normally hydrated cornea, mutual repulsion between the negatively charge glycosaminoglycan chains of these proteoglycans maintains the uniform interfibrillar spacing of stromal fibrils (Bocherding et al., 1975).

The induction of opaque rabbit corneal scars have been shown to be accompanied by a change in types and an increase in size of proteoglycans to at least three times the normal sizes, resulting in the disruption of uniform interfibrillar spacing (Hassell et al., 1983). After a year, the corneas regained transparency with a corresponding return to the occurrence of normal corneal proteoglycans. Similarly, Bocherding et al. (1975) noted that the transition from the ordered fibril organisation of cornea to the less organised sclera corresponds to a decrease of the small keratan sulphate proteoglycans and an increase in the amounts of larger proteoglycans.

Due to the hydrophilic nature of proteoglycans, the corneal stroma has a tendency to imbibe water from the hypotonic aqueous humour (Maurice, 1984). With swelling, the cornea becomes opaque, presumably due to disruption of the normally uniform interfibrillar spacing. In the functioning cornea, this is prevented from happening by the limiting layers, especially the endothelium.

I C. THE LIMITING LAYERS :

MAINTENANCE OF OPTIMUM CORNEAL HYDRATION

In maintaining the optimum level of corneal hydration, the limiting layers firstly serve a passive barrier function. In addition, the endothelium possesses a fluid-pumping mechanism which pumps fluid back into the hypotonic aqueous humour. Damage to the epithelium causes swelling of the cornea while endothelial damage

causes a more pronounced swelling (Maurice and Giardini, 1951b). Addition of metabolic inhibitors or deprivation of oxygen also cause swelling (Langham and Taylor, 1956). Similarly, cooling the cornea causes swelling (Mishima and Hedbys, 1967) which can be reversed by reincubating at 37°C (Mishima and Kudo, 1967). Swelling also occurs when ouabain is applied to the corneal endothelium (Trenberth and Mishima, 1968). An energy-requiring fluid pump associated with the the endothelium has thus been proposed (Hodson, 1977; Mayes and Hodson, 1979; Hodson and Wigham, 1983). According to the proposed mechanism, an endothelial ion-pump pumps bicarbonate ions, perhaps in conjunction with K+, Na+ and Cl- ions into the aqueous humour. Whether water follows as a result of localised osmotic effects or by other mechanisms is still unclear. The enzymes Na+K+ATPase and carbonic anhydrase are thought to play major roles in this 'pump-leak' mechanism (Maurice, 1984; Mclaughlin et al., 1985) in which, water leaking into the cornea from the aqueous humour is constantly removed by the endothelial fluid pump.

I D. CORNEAL NUTRITION, RESPIRATION AND METABOLISM

Metabolism of glucose in the cornea produces the energy required to sustain cellular turnover (Mishima and Hedbys, 1968) as well as the proposed 'endothelial pump'.

Glucose reaches all layers of the cornea from the aqueous humour with small amounts entering across the epithelium from the limbus (Hale and Maurice, 1969). Amino acids enter from the aqueous humour by diffusion but keratocytes and epithelial cells may be able to accumulate amino acids by a transport mechanism (Riley, 1977). Oxygen diffuses to all layers of the cornea from the atmosphere via the epithelium with a small but significant transendothelial diffusion from the aqueous humour (Riley, 1969; Weismann et al., 1981). When the eye is shut, the conjunctiva, and to a lesser extent, the aqueous humour replace the atmosphere as a source of oxygen (see Maurice, 1984).

Metabolism of carbohydrate in the cornea occurs via the glycolytic, citric acid and

pentose phosphate pathways (Kinoshita, 1962). It is thought that stromal glucose oxidation occurs entirely via the citric acid cycle. Of special note is the high proportion of glucose metabolised via the pentose phosphate pathway in the epithelium (Kuhlman and Resnik, 1959). This corresponds to the high level of protein (hence RNA) synthesis required to replace epithelial cells weekly. Protein synthesis in the stroma is rapid during development but is extremely slow in the adult cornea (Maurice and Riley, 1970). The half−life of collagen in most mature tissues is at least 100 days (Neuberger and Slack, 1953), the turnover of adult corneal collagen being extremely slow (Smelser et al., 1965).

II. COLLAGEN

II A. PERSPECTIVE

Collagen of the corneal stroma is in the form of thin fibrils of uniform diameter. This particular quality of corneal collagen is essential for corneal transparency. In order to ascertain the possible factors that may regulate collagen fibril diameter, it is necessary to understand the biosynthesis and structure−function relationships of collagen.

Collagen is the most abundant of animal proteins, and may constitute up to a quarter of the body protein (Harkness, 1961). As a major component of connective tissues such as tendon, bone, ligament, cartilage, basement membranes, cornea and sclera, it provides an extracellular framework for all multicellular animals by its ability to form supramolecular aggregates such as fibres and sheets. In the past ten years several reviews have appeared on various aspects of collagen research (Grant and Jackson, 1976; Fessler and Fessler, 1978; Bornstein and Traub, 1979; Prockop et al., 1979; Bornstein and Sage 1980; Eyre, 1980; Miller, 1982; Eyre et al., 1984; Prockop and Kivirikko, 1984; Sykes, 1985; Cheah, 1985; Martin et al., 1985). In addition, books and volumes reporting the proceedings of symposia are also available which are dedicated to aspects of collagen research (edited or authored by: Ramachandran and Reddi, 1976; Parry and Creamer, 1979; Woodhead−Galloway, 1980; Viidik and Vuust,

1980; Cunningham and Fredriksen, 1982; Weiss and Jayson, 1982; Fleischmajer et al., 1985) while the series entitled "*International Review of Connective Tissue Research*" provides periodic reviews of major topics in collagen research.

The collagens are a family of proteins that exhibit a wide diversity of structural and chemical features (see Miller, 1985). Collagen molecules are trimers, consisting of three co-aligned polypeptide chains called 'α' chains. The molecules contain significant or extensive domains where the three α-chains are in a triple helical conformation. In the triple helical regions, the α-chain amino acid sequences consist of repeating 'Gly-X-Y' triplets within each chain, an absolute requirement for the collagen triple-helix. The ability of collagen molecules to form higher-order structures depends on molecular length, shape and extent of triple helical domains in relation to the positions of intermolecular interacting sites as determined by the primary stucture of collagen.

According to the present classification system, the collagen family of proteins consists of at least ten genetic 'types', designated by the Roman numerals I to X, consisting of particular trimeric combinations of at least eighteen genetically distinct collagen α-chains (see Martin et al., 1985 and fig. 1.11). 'Type I' collagen accounts for about 90% of the collagen of the body, being almost the only collagen in bone and tendon, and is the major collagen type in skin, teeth, cornea and sclera (see Bornstein and Sage, 1980; von der Mark, 1981). This chapter will thus mainly discuss type I collagen. The type I collagen molecule is a heterotrimer of two α1(I) and one α2(I) chains. Unless stated otherwise, references to 'α1' and 'α2' throughout this thesis refer to the α-chains of type I collagen.

II B. COLLAGEN GENES

In humans, genes coding for the different collagen α-chains have so far been found to be dispersed throughout the genome and occur as single copies per haploid genome (see Cheah, 1985). The genes coding for the α1(I) and α2(I) chains are on chromosomes 17 and 7 respectively. The other collagen genes located are, the α1(II),

α1(III) and α1(IV) genes on chromosomes 12, 2 and 13 respectively.

Characterisation of isolated gene fragments have revealed the remarkable structure of collagen genes. Genes coding for collagen types I and III, both fibre−forming collagens, consist of about 50 coding exons separated by introns (Yamada et al., 1984; Chu et al., 1984). Of the 18 kilobase pair α1(I) and the 39 kilobasepair α2(I) genes, only 6 kb pairs and 5 kb pairs respectively consist of coding sequences. About forty of the exons are either 54 base pairs long or related to this basic unit. The first codon of these characteristic exons always code for glycine while the last codon always codes for the amino acid occurring in the 'Y' position of the 'Gly−X−Y' triplet sequence. It has thus been suggested that the present day collagen genes have resulted from successive amplifications of a 'primordial' 54 base−pair unit which could have coded for exactly six 'Gly−X−Y' triplets (Tate et al., 1982; Runnegar, 1985). However, exons related to this basic unit have not been found in genes coding for α−chains of type IV (Kurkinen et al., 1985), type IX (Lozano et al., 1985) and invertebrate collagens (see Tate et al., 1982).

II C. BIOSYNTHESIS OF COLLAGEN

Type I collagen α chains are synthesised in the cell as their larger precursor forms, the 'pro−α' chains (see Fessler and Fessler, 1978). Prior to the assembly of triple helical 'procollagen' molecules, particular amino acid side−chains undergo enzyme−catalysed post−translational modifications which affect the properties of pro−α chains and the assembled procollagen. Following secretion into the extracellular space, procollagen is processed to yield collagen which subsequently assembles into ordered fibrils (see figures 1.4 and 1.5).

1. Intracellular events

a) Transcription and translation

The multi−exon structure of the type I collagen genes means that about 50 non−coding sequences need to be spliced out of the 'primary' messenger RNA to yield

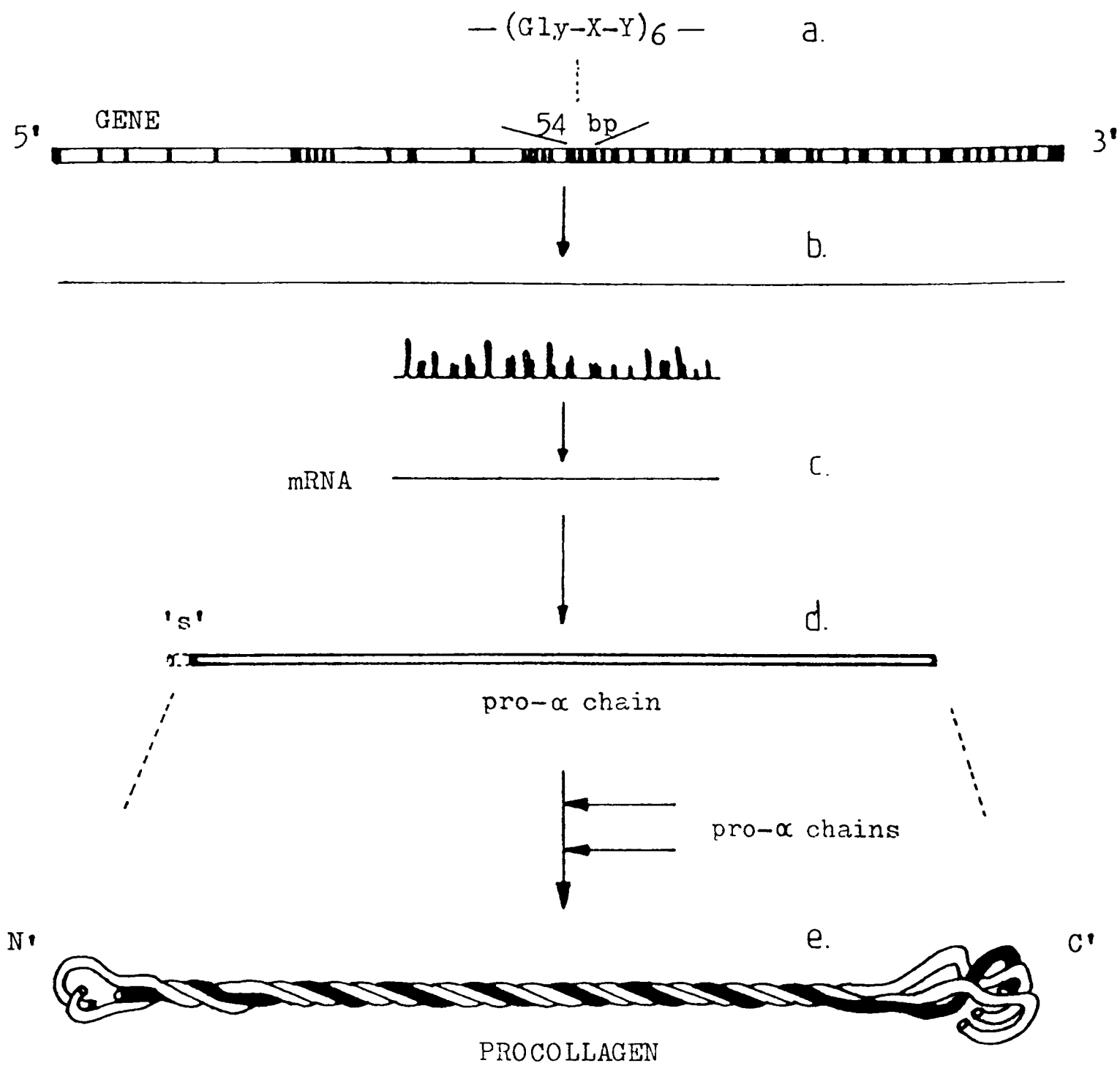


Figure 1.4 A scheme showing the progression from the multi−exon collagen gene to the triple−helical procollagen (Type I) molecule. **a)** 18 kilobase−pair pro−α1(I) gene, showing a typical 54 base−pair exon coding for six 'Gly−X−Y' amino acid triplets; **b)** pre−mRNA; **c)** mRNA (about 6 kb); **d)** pro−α1(I) chain including the signal peptide ('s') at the N−terminus; **e)** procollagen molecule consisting of two pro−α1(I) chains and one pro−α2(I) chain. Note: the signal peptide is cleaved off before assembly of the procollagen molecule. The globular 'propeptide' domains at the N'− and C'− terminal ends lack the repeating −Gly−X−Y− 'helical' sequence except for the occurrence of a short stretch in the 'N−propeptide'. (Data from *Prockop and Kivirikko, 1984; Fessler and Fessler, 1978).*

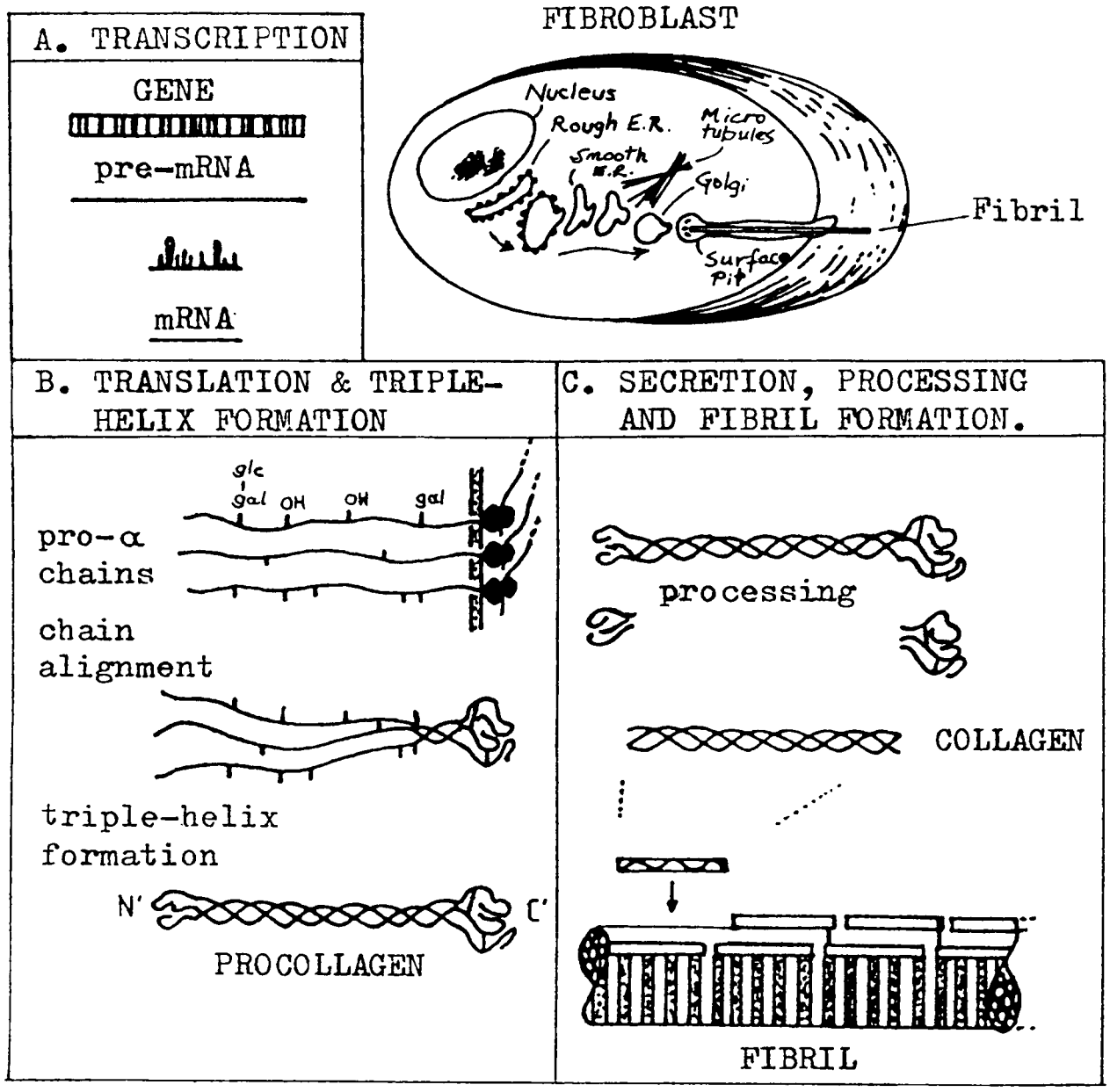


Figure 1.5 **Biosynthesis of type I collagen**

A: the procollagen genes are transcribed in the nucleus; B: after RNA splicing, the mRNAs are translated on ribosomes bound to rough endoplasmic reticulum (E.R.). Hydroxylases and glycosyltransferases modify some lysyl and prolyl residues on the nascent chains within the rough E.R. (*OH* = hydroxyl group; *gal* = galactose; *glc* = *glucose)*. Chain alignment and interchain disulphide bond formation between three pro−α chains are followed by triple helix formation to yield the procollagen molecule; C: procollagen is translocated via the smooth E.R. and Golgi−derived vesicles to the cell−surface and secreted. Outside the fibroblast, procollagen is 'trimmed' by proteinases to yield collagen. Collagen molecules then assemble into fibrils followed by cross−link formation.

the 'mature' mRNA for translation (see Prockop and Kivirikko, 1984). Transcriptional control or selective degradation of pro−α2 mRNA maintains pro−α1 and pro−α2 mRNAs in a 2:1 ratio (Vuust et al., 1983). Procollagen mRNA is then translated by ribosomes bound to endoplasmic−reticulum (Grant and Jackson, 1976, Veis et al., 1985) this being equally efficient for pro−α1 and pro−α2 mRNAs. A leader hydrophobic signal peptide sequence of about 22−26 amino acid residues (refer to fig. 1.10) directs the elongating pro−α chain into the lumen of the 'rough' endoplasmic reticulum as with other secreted proteins (see Harwood, 1980), chain elongation continuing after the signal peptide is cleaved off. Translation of a complete type I pro−α chain takes about 6 minutes (Vuust and Piez, 1972).

The completed pro−α1(I) chain consists of an extensive, uninterrupted 'helical' domain of 1014 amino acid residues (338 repeating 'Gly−X−Y' triplets), flanked by N−terminal and C−terminal 'propeptides' of 139 and 246 residues respectively. These propeptide domains are largely of 'globular' nature except for a 'helical' $(Gly-X-Y)_n$ sequence of about 40 residues within the N−propeptide. Between the propeptide domains and the major helical domain, are short non−triplet sequences of 16 residues and 25 residues, also known as N− and C−terminal 'telopeptides' respectively (see Miller, 1985 and refer to fig. 1.10). The N−propeptide and the telopeptides of the pro−α2(I) chain are shorter than their counterparts in the pro−α1(I) chain.

b) **Post−translational modifications of proline and lysine**

While the pro−α chain is being translated, some lysyl residues are hydroxylated to yield (5)−hydroxylysine. The hydroxylation is catalysed by the enzyme lysyl hydroxylase, which requires the presence of molecular oxygen, ferrous iron, α−oxoglutarate and a reducing agent such as ascorbate (see Kivirikko and Myllylä, 1980, 1982, 1985). Amino acid sequence analysis of collagens (refer to fig. 1.10) have, without exception, indicated that only those lysyl residues occurring in the 'Y' positions ('pre−Gly' positions) of the repeating 'Gly−X−Y' triplet sequences may be hydroxylated. Out of thirteen and sixty hydroxylated lysyl positions so far located

during amino acid sequence determinations of the bovine α1(II) (Francis et al., 1978) and human α1(IV) (Babel and Glanville, 1984; Glanville et al., 1985) chains respectively, all have been found to be in 'pre−Gly' positions, the minimum sequence requirement being '−Lys−Gly−X−Y−Gly−'. There may however, be a different form of lysyl hydroxylase that is specific towards lysyl residues in the short, non−helical 'telopeptide' regions of collagen chains (Royce and Barnes, 1985).

Hydroxylysyl residues, in turn, may be glycosylated to yield the hydroxylysine glycosides galactosylhydroxylysine (GH) and glucosylgalactosylhydroxylysine (GGH). These glycosylations are catalysed by the enzymes galactosyltransferase, which catalyses the O−glycosidic linkage of galactose via the hydroxyl group of hydroxylysine to form the monoglycoside GH, and glucosyltransferase which catalyses the linkage of glucose to the galactose of GH to form the diglycoside GGH (Kivirikko and Myllylä, 1979). The reactions require the presence of UDP−glycoside (the galactose or glucose donor), and Mn^{2+}. Galactosyltransferase occurs in two forms of molecular weights 450 kilodaltons and 200 kilodaltons while glucosyltransferase occurs as a single chain of molecular weight 70 kilodaltons. The enzymes are glycoproteins and probably contain sulphydryl groups at their catalytic sites.

Prolyl residues occupy about 100 of the 'X'− (post−Gly), and 100 of the 'Y'− (pre−Gly) positions in the repeating Gly−X−Y triplet sequences of the type I α chains (refer to fig. 1.10). Most of the pre−glycine prolyl residues are hydroxylated to 4−hydroxyproline by the catalytic action of the enzyme prolyl−4−hydroxylase (Kivirikko and Myllylä, 1980). Post−glycine prolyl residues are not modified by this enzyme (Tryggvason et al., 1977). Prolyl hydroxylation requires the same co−substrates as lysyl hydroxylation. About one post−glycine prolyl residue per α1(I) chain is hydroxylated to 3−hydroxyproline by the enzyme prolyl−3−hydroxylase (Fietzek and Kühn, 1975).

Although there are only small tissue variations in prolyl−4−hydroxylase and glucosyltransferase activities, there are significant tissue variations in the activities of the enzymes lysyl hydroxylase and galactosyltransferase. Thus, enzyme assays have

shown higher activities of lysyl hydroxylase and galactosyltransferase in cartilage cells as opposed to tendon and bone cells, as well as in foetal skin cells as opposed to adult skin cells (Spiro and Spiro, 1971; Ryhänen and Kivirikko, 1974; Antinnen et al., 1977; Risteli et al., 1979). These findings are consistent with the more extensive modification of lysyl residues of cartilage (type II) collagen (see Miller and Gay, 1982) and collagens of young tissues (Barnes et al., 1974; Royce and Barnes, 1977).

With respect to post−translational glycosylation it should also be noted that oligosaccharides are also found linked to procollagen but these are linked via asparagine residues located in the propeptide domains (see Kivirikko and Myllylä, 1982).

c) **Assembly of the procollagen molecule**

The completed pro−α chains are released from the ribosomes into the lumen of the endoplasmic reticulum, after which, modifications of lysyl and prolyl residues may still continue (Oikarinen et al., 1976). Although the mechanism of collagen triple−helix formation is still not fully understood, it probably begins with the folding of individual pro−α chain N− and C− propeptides with the formation of intra−chain disulphide bonds in these regions. Mutual chain selection between two pro−α1 chains and one pro−α2 chain is facilitated by interactions between their folded N−propeptides and between their folded C−propeptides culminating in the formation of interchain disulphide bonds (Fessler and Fessler, 1978; Bachinger et al., 1981; Fessler et al., 1985). Then, provided that enough prolyl residues have been hydroxylated to (4)−hydroxyproline, the three pro−α chains coil into a right−handed triple helix through all the 'Gly−X−Y' triplet sequences starting from the C−terminus. Cis→trans isomerisation of peptide bonds (especially X—Pro, X—Hyp) is thought to be the rate−limiting step (Bruckner and Eikenberry, 1984; Sarkar et al., 1984). The triple helical structure requires every third residue to be a glycine residue as glycine is the only amino acid small enough to fit into the space down the tight triple heix while the imino acids proline and hydroxyproline confer a degree of rigidity to the triple

helix due to constraints imposed by their characteristic ring structures (Ramachandran and Ramakrishnan, 1976). The triple helix is stabilised by interchain hydrogen bonds transverse to helix axis. Hydroxyprolyl residues contribute significantly to hydrogen−bonding, probably mediated by 'water−bridges'. Underhydroxylation of prolyl residues therefore leads to poor triple helix formation and explains the relatively low melting temperatures of under−hydroxylated procollagen and collagen molecules (Berg and Prockop, 1973).

The triple helical conformation inhibits further modification of prolyl and lysyl residues (see Grant and Jackson, 1976). Blumenkrantz et al. (1984) have found that prolyl hydroxylase resides mainly in the lumen of the endoplasmic reticulum while most of the lysyl hydroxylase and glycosyltransferase activities reside in the membrane of the reticulum. They propose a sequence of events whereby, modifications of lysyl and hydroxylysyl residues occur prior to the completion of prolyl hydroxylation. When sufficient prolyl residues have been hydroxylated (about 100 per chain), triple helix formation occurs rapidly, inhibiting further modification of lysyl and hydroxylysyl residues.

d) Translocation and secretion of procollagen

The assembled procollagen molecule is translocated from the rough endoplasmic reticulum, via the smooth endoplasmic reticulum and Golgi−derived vesicles to the cell surface. Procollagen secretion is inhibited by $\alpha\alpha'$−dipyridyl (an inhibitor of prolyl hydroxylase), antimycin A (a metabolic inhibitor), and colchicine (an inhibitor of tubulin assembly). Translocation and secretion of procollagen thus require the triple helical conformation and, are energy−dependent processes mediated by microtubules (see Grant and Jackson, 1976). Trelstad and Hayashi (1979) have reported electron microscopic observations, in tendon fibroblasts, of 'secretory vacuoles' containing aggregates of procollagen, en route to deep cell−surface pits. Birk and Trelstad (1984) reported similar findings in corneal keratocytes.

2. Extracellular events

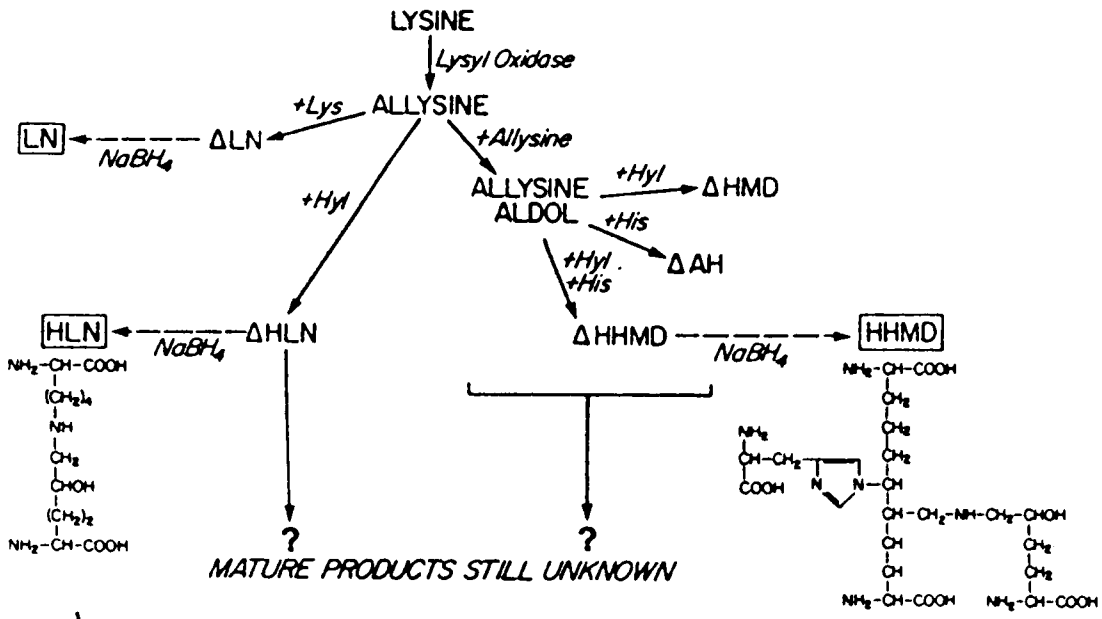
a) Processing of 'procollagen' to 'collagen'

In the extracellular space, perhaps in the deep cell−surface pits, type I procollagen is converted to type I collagen by the proteolytic removal of the mainly 'globular' N− and C− propeptides. Separate N− and C− proteinases catalyse these cleavages (see Heathcote and Grant, 1980; Bornstein and Sage, 1980; Kivirikko and Myllylä, 1982; Peltonen et al., 1985). Rid of the propeptides, the resultant rod−like collagen molecule, about 300 nm long and 1.4 nm in diameter, therefore consists of two $\alpha 1$ chains and one $\alpha 2$ chain, each consisting of a 1014−residue long triple helical domain flanked by the short, non−helical telopeptide domains. Non−covalent, charge−interactions and hydrophobic interactions between molecules lead to the self−assembly of collagen molecules into fibrils (see Fietzek and Kühn, 1975; Hofmann et al., 1978, 1980).

b) Cross−link formation

The collagen fibrils resulting from self−assembly are stabilised and strengthened by the formation of lysine− and hydroxylysine−derived covalent intermolecular cross−links (see review by Eyre et al., 1984). A copper−dependent enzyme, lysyl oxidase (Siegel, 1979), catalyses the oxidative deamination of some lysyl and hydroxylysyl residues of the collagen α chains to their corresponding aldehyde forms allysine and hydroxyallysine, which serve as precursors in two similar but separate routes of cross−link formation (fig. 1.6). The chemistry of collagen cross−links have been reviewed by several authors (see Bornstein and Traub, 1979; Harding, 1985) and will not be discussed here in great detail except to say that the primary cross−links are either Schiff−base products between the aldehyde group of one 'lysyl' residue and the ϵ−amino group of another 'lysyl' residue or aldol condensation products of two 'lysyl' aldehyde groups on separate collagen α−chains.

Hydroxyallysine−derived crosslinks are generally more stable than their allysine counterparts due their ability to spontaneously rearrange from the aldimine to the more



a)

Scheme outlining one of two routes of cross-linking in collagen: Allysine-based cross-links. LN, lysinonorleucine; HLN, hydroxylysinonorleucine; HMD, hydroxymerodesmosine; AH, aldolhistidine; HHMD, histidinohydroxymerodesmosine. The prefix Δ for dehydro signifies the natural, aldimine forms of the various compounds.

b)

Scheme outlining one of two routes of cross-linking in collagen: Hydroxyallysine-based cross-links. DHLN, dihydroxylysinonorleucine; HP, hydroxylysyl pyridinoline; LP, lysyl pyridinoline. The prefix Δ for dehydro, signifies the natural, aldimine forms of the various compounds.

Figure 1.6 Schemes showing the two routes for the formation of lysine-derived cross-links in collagen: a) the allysine route b) the hydroxyallysine route.
(From *Eyre et al., 1984*)

stable ketoamine forms. Initial crosslinks between two molecules may react with other co-aligned lysyl, hydroxylysyl residues to form poly-functional intermolecular cross-links such as pyridinoline (Fujimoto et al., 1978). Incorporation of histidine residues into cross-links has also been reported but it has been suggested that these are artifacts of sodium borohydride reduction (Robins and Bailey, 1977), a preparatory step in the isolation of collagen crosslinks. Intramolecular crosslinks have also been discovered (Harding et al., 1980), but their significance to fibril stability is not known.

II D. STRUCTURE-FUNCTION RELATIONSHIPS OF TYPE I COLLAGEN AND OF OTHER COLLAGEN TYPES

1. Type I collagen

Type I collagen is the most abundant collagen in vertebrates and is the major collagen of the cornea and many other tissues. One of the main features of type I collagen is its occurrence as fibrils.

a) Periodicity of fibrils

Staining of collagen reveals the gross molecular features of collagen (see fig. 1.7; Piez and Miller, 1974; Bornstein and Traub, 1979; Woodhead-Galloway, 1980). Negative-staining of collagen fibrils allows stain to fill any gaps between the constituent molecules. Viewed under the electron microscope, the stain-filled gaps show up as dark areas while the stain-excluding regions (collagen) show up as light areas, resulting in a characteristic striped appearance. Positive-staining of collagen fibrils results in the staining of charged amino acid residues. Dispersed collagen mollecules or α chains can be made to precipitate, in the presence of ATP, as lateral aggregates of parallel molecules, with their N- and C- terminal ends in register. These aggregates, known as segment-long-spacing (SLS) crystallites, can also be positively stained.

Electron microscopic examination of negative-stain and positive-stain patterns of native collagen fibrils reveal a consistent 'D'-periodicity along the fibril axis, where D

is a distance of 67 ± 3 nm (fig. 1.7). Comparison of the D period with the collagen molecular length of 300 nm, as indicated by examination of SLS patterns, prompted Schmitt et al. (1955) to propose the 'quarter−stagger' hypothesis for the intrafibrillar arrangement of collagen molecules. More careful examination of the data led to a modification of the 'quarter−stagger' hypothesis to the more accurate 'D−stagger' hypothesis according to Petruska and Hodge (1964).

According to the D−stagger hypothesis, depicted in figure 1.8, the type I collagen molecule can be divided into five segments, four of length D and a fifth, of length 0.4D. All molecules lie parallel to the fibril axis and point in the same direction. Axially, adjacent molecules lying on the same axis are separated by gaps of 0.6D. Laterally, adjacent molecules are staggered by length D. A side view of the fibril therefore shows 'overlap' regions of length 0.4D, devoid of axial gaps, alternating with 'gap'− containing regions of length 0.6D. It is these overlap and gap regions that show up as the light and dark bands respectively, of negatively stained native fibrils. Support for the D−stagger hypothesis came from demonstrations that synthesis of appropriately displaced SLS patterns showed good agreement with the native fibril pattern (Kühn et al., 1960). More recent X−ray diffraction analysis of collagen fibrils have led to the modification of the gap:overlap ratio to 0.53D:0.47D (Brodsky and Eikenberry, 1982). Incorporation of the Petruska−Hodge model into a three−dimensional fibril will be discussed later (Chapter 1.IIE).

b) Primary structure of type I collagen

The ordered D−stagger arrangement of collagen molecules in native and reconstituted type I fibrils implied that there are features of collagen primary structure which direct the self−assembly process.

Soluble type I collagen can be extracted for study, from tissues of young or lathyritic animals (i.e. tissues with decreased intermolecular cross−linking of collagen) in dilute acid or neutral salt solutions. Insoluble collagens can be partially solubilised by limited digestion with pepsin or other proteases, which digest the non−triple helical

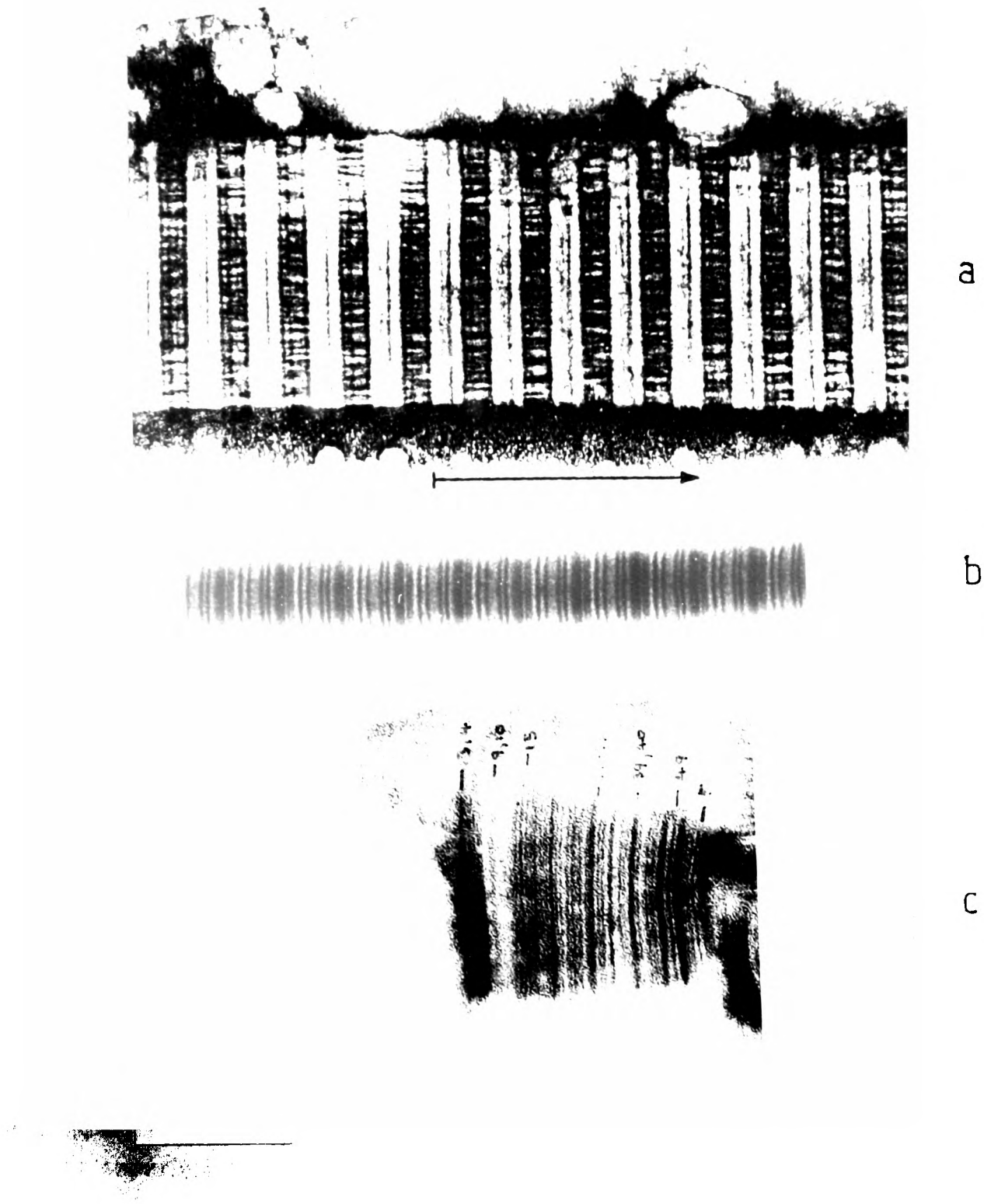


Figure 1.7 Staining of type I collagen

a) Electronmicrograph of negatively−stained fibril (from Piez and Miller, 1974); b) positively−stained reconstituted collagen (pepsin−solubilised corneal collagen) fibril and c) positively−stained segment−long−spacing (SLS) crystallite of corneal collagen (provided by *J.J. Harding*). The arrow marks the length of a collagen molecule (N'→C' terminal end). Note the 67nm periodicity in a) and b).

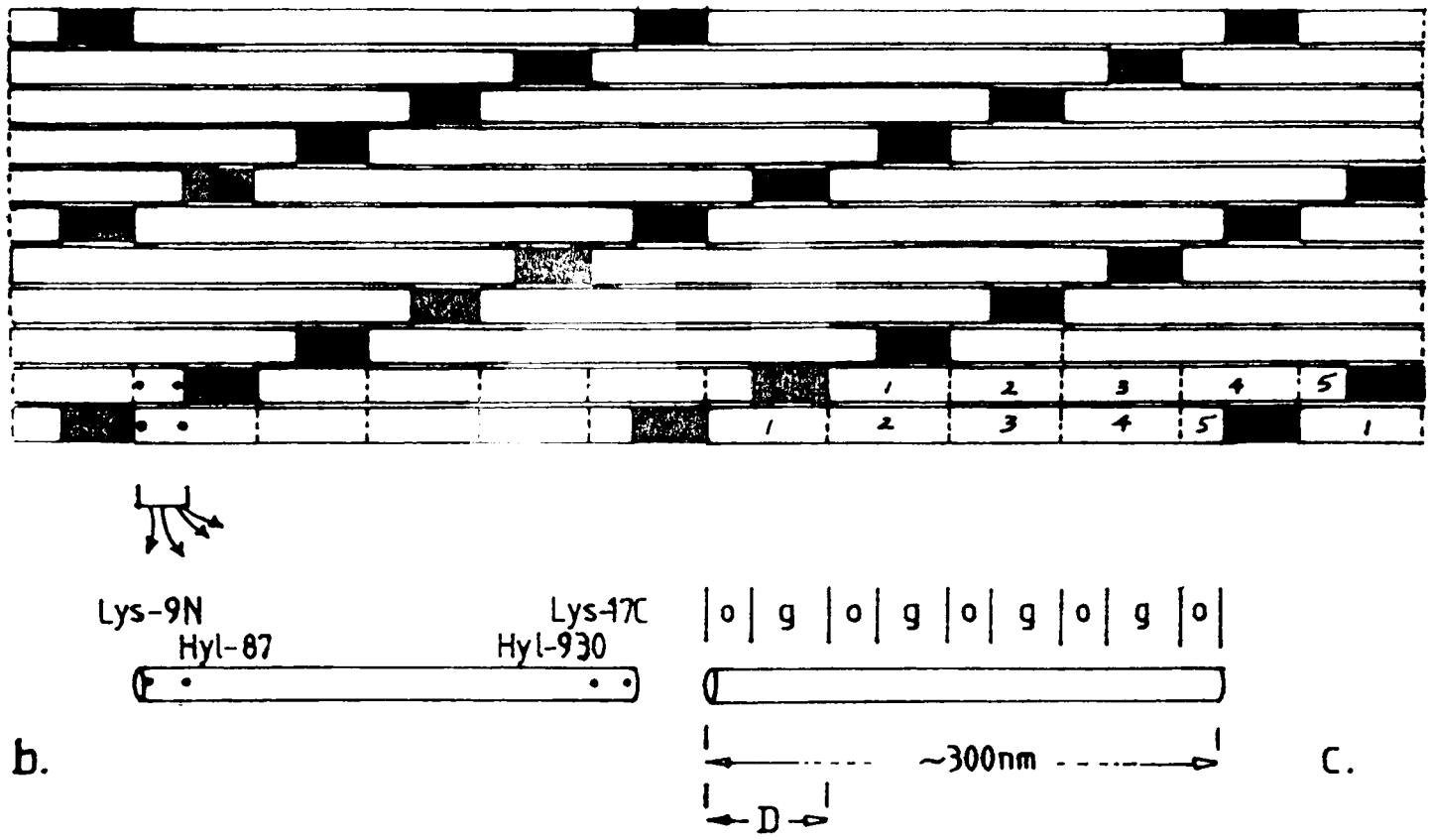


Figure 1.8 The packing of collagen molecules (type I) in 1D stagger according to the hypothesis of Petruska and Hodge (1964). a) two−dimensional representation: the black areas represent axial 'holes' or 'gaps' between the molecules (white blocks). Numbers indicate the five segments per molecule of length 4.4D within the D−stagger arrangement. Segment *1* is at the N−terminal end; b) principal cross−linking sites (in the α1 chain) forming 'head−to−tail' cross−links which stabilise the D−staggered fibril. Lys-9^{N} and Lys-17^{C} are aldehyde forming lysines in the N− and C− telopeptides respectively; c) the 'gap' and 'overlap' *regions* of the fibril superimpose four gap and five overlap *zones* onto each collagen molecule.

telopeptide regions containing the major sites of intermolecular cross-links (see Bornstein and Sage, 1980). Alternatively, insoluble collagen can be purified by extracting the non-collagenous tissue components into 4M guanidinium hydrochloride, pH 7.5 (Miller and Lunde, 1973). The different types of collagen can be selectively isolated from heterogeneous solutions of collagen by selective salt fractionation. This and other subsequent chromatographic procedures used in preparing collagens for study have been reviewed in depth by Miller and Rhodes (1982).

The systematic elucidation of the primary structure of type I collagen began with the separation of α1 and α2 chains by ion-exchange chromatography on carboxymethyl cellulose (Piez et al., 1963). The isolated chains were then cleaved, using agents that yielded a small number of large peptides to ease subsequent fractionation and characterisation (see Fietzek and Kühn, 1976). Cleavage on the carboxyl side of methionyl residues is a method of choice, as its high specificity means that a relatively small number of fragments are produced (fig. 1.9), making the isolation of pure fragments relatively easy (Butler, W.T. et al., 1967). Cyanogen bromide (CNBr) peptides were isolated by ion-exchange chromatography on carboxymethylcellulose and were assigned names such as α1-CB1, α1-CB2, α2-CB2, α2-CB3 and so forth, according to their chain of origin and the order in which they eluted during the chromatography. Large peptides were also obtained by digestion with collagenase and hydroxylamine. The order of these large peptides along the α chains was determined by analysis of uncleaved 'double' peptides and by pulse-labelling studies. Comparing the SLS patterns of large collagen peptides with those of intact α chains also revealed the positions of peptides along collagen α chains. Large collagen peptides were cleaved further with proteases such as trypsin, chymotrypsin, thermolysin and more recently, *Staphylococcal aureus* V8-protease (Glanville et al., 1983). Sequencing of these overlapping peptides has yielded the entire amino acid sequences of the α1(I) chain of chicken and calf skin and substantial portions of the rat α1(I) and bovine α2(I) chains. In addition, nucleotide sequencing of collagen genetic DNA and RNA has also been used to supplement the data. Whole or partial amino acid sequences for the α chains

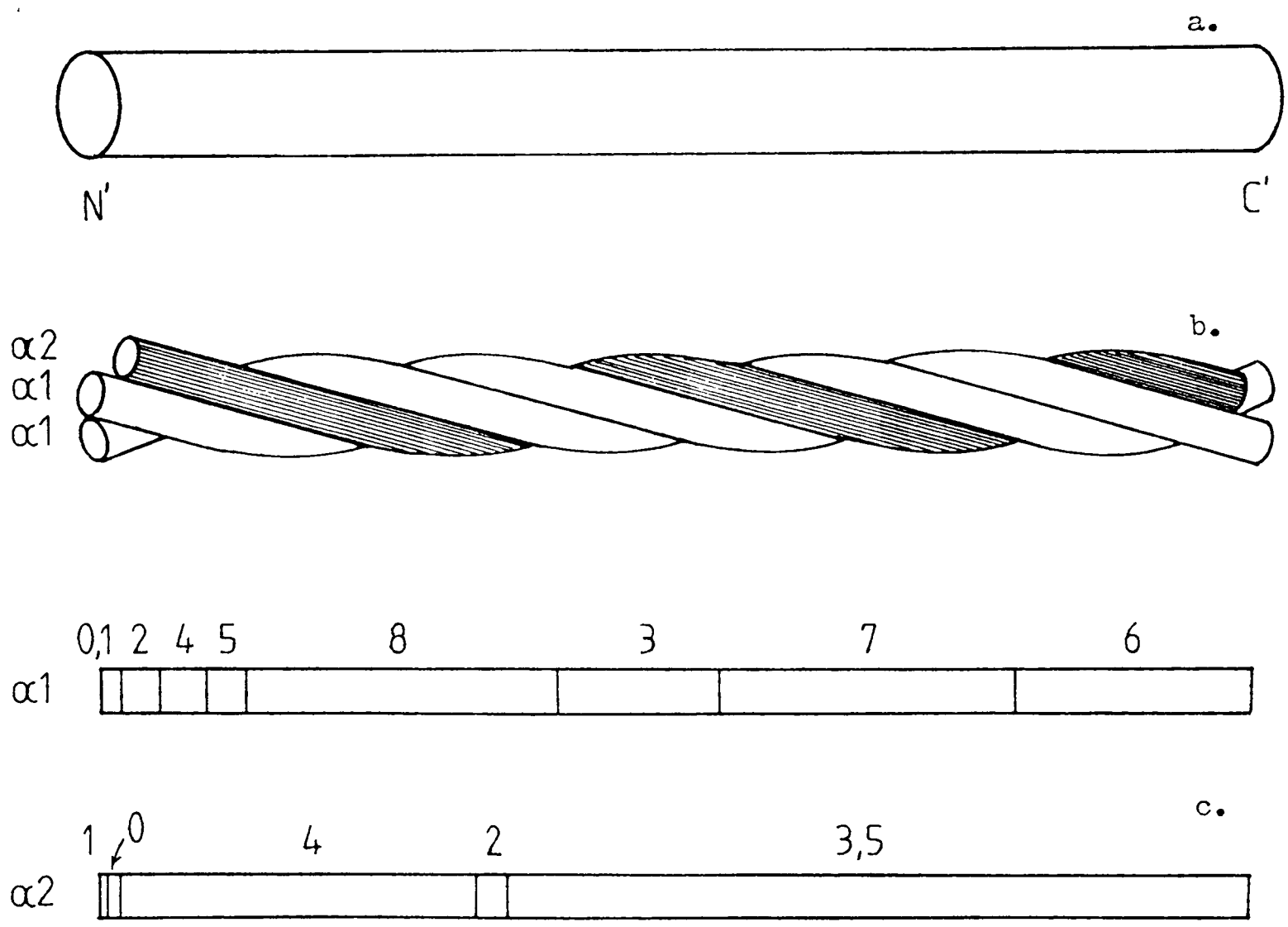


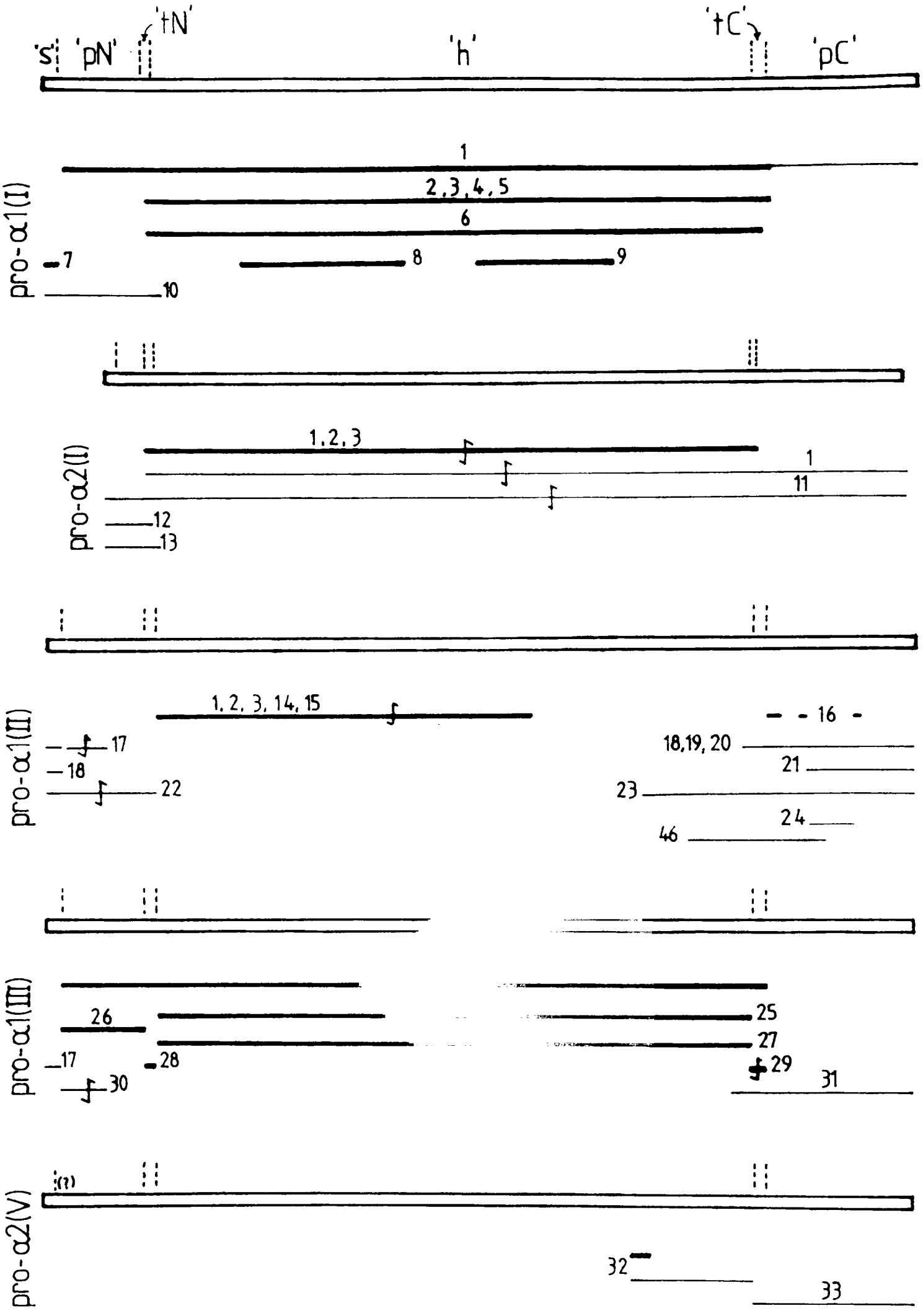
Figure 1.9 Diagram showing the rod−like type I collagen molecule 'a', to consist of two $\alpha 1$ chains and one $\alpha 2$ chain in a right−handed triple−helix as indicated in 'b'. In **'c'**, peptides derived from the α chains (bovine) by cleavage with cyanogen bromide (CNBr) at methionyl residues (vertical strokes) are indicated. *Note*: the collagen molecules are not drawn to scale, *the* 'true' proportions being a length of about 300nm and a diameter of about 1.4 nm. Data for CNBr peptides taken from *Fietzek and Kühn, 1976*.

of some of the other collagen types have also been elucidated. References for most of the known collagen amino acid sequences are listed in the 'Appendix' at the end of the thesis in conjunction with figure 1.10.

Comparison of the known amino acid sequences of type I, type II and type III collagen α chains show substantial interchain and interspecies homology. The sequence of the α1(I) chain (as well as the α1(III) chain) reveals the occurrence of glycine as every third residue in a stretch of about 1014 amino acid residues, otherwise known as the helical domain. Proline and 4-hydroxyproline occupy approximately 100 post-Gly and 100 pre-Gly positions respectively in this region. As mentioned previously, the occurrence of glycine and imino acid residues in these positions facilitate the formation of a stable and relatively rigid triple helical domain in the procollagen molecule and subsequently the collagen molecule.

There is a non-random distribution of hydrophobic and polar amino acids along the α1(I) and α2(I) chains. Examination of these collagen chain sequences (as well as that for the α1(III) chain) indicate that intermolecular polar and hydrophobic interactions are maximised between clusters of polar or hydrophobic amino acids when adjacent molecules are staggered by integral multiples of a period equivalent to 234 ± 2 amino acid residues (Fietzek and Kühn, 1976; Hofmann et al., 1980). Chapman and Hardcastle (1974), tested several possible staggers and found that a stagger of 232-233 residues gave optimum interactions between two molecules. The 233-residue period is synonymous with the 67 nm 'D' period (see fig. 1.8). Sub-D periodicities have also been noted within the α chains of the fibrillar collagen types (Hofmann et al., 1980).

Studies of collagen primary sequence have also revealed the occurrence of (4)-hydroproline and hydroxylysine residues in 'pre-Gly' positions of the helical sequence. Important sites of hydroxylysine have been located at positions 87 and 930 in the α chains of type I collagen, with identical or similar positions in collagen types II and III. These hydroxylysyl residues have been found to participate in intermolecular cross-links between adjacent molecules staggered by four D periods, involving aldehyde derivatives of lysyl residues occurring in the telopeptide regions (see Eyre et al., 1984;


's'
'pN'
'tN'
'h'
'tC'
'pC'
pro-α1(I)
1
2,3,4,5
6
7
8
9
10
pro-α2(I)
1,2,3
1
11
12
13
pro-α1(II)
1,2,3,14,15
16
17
18
18,19,20
21
22
23
24
46
pro-α1(III)
26
25
27
17
28
29
30
31
pro-α2(V)
32
33

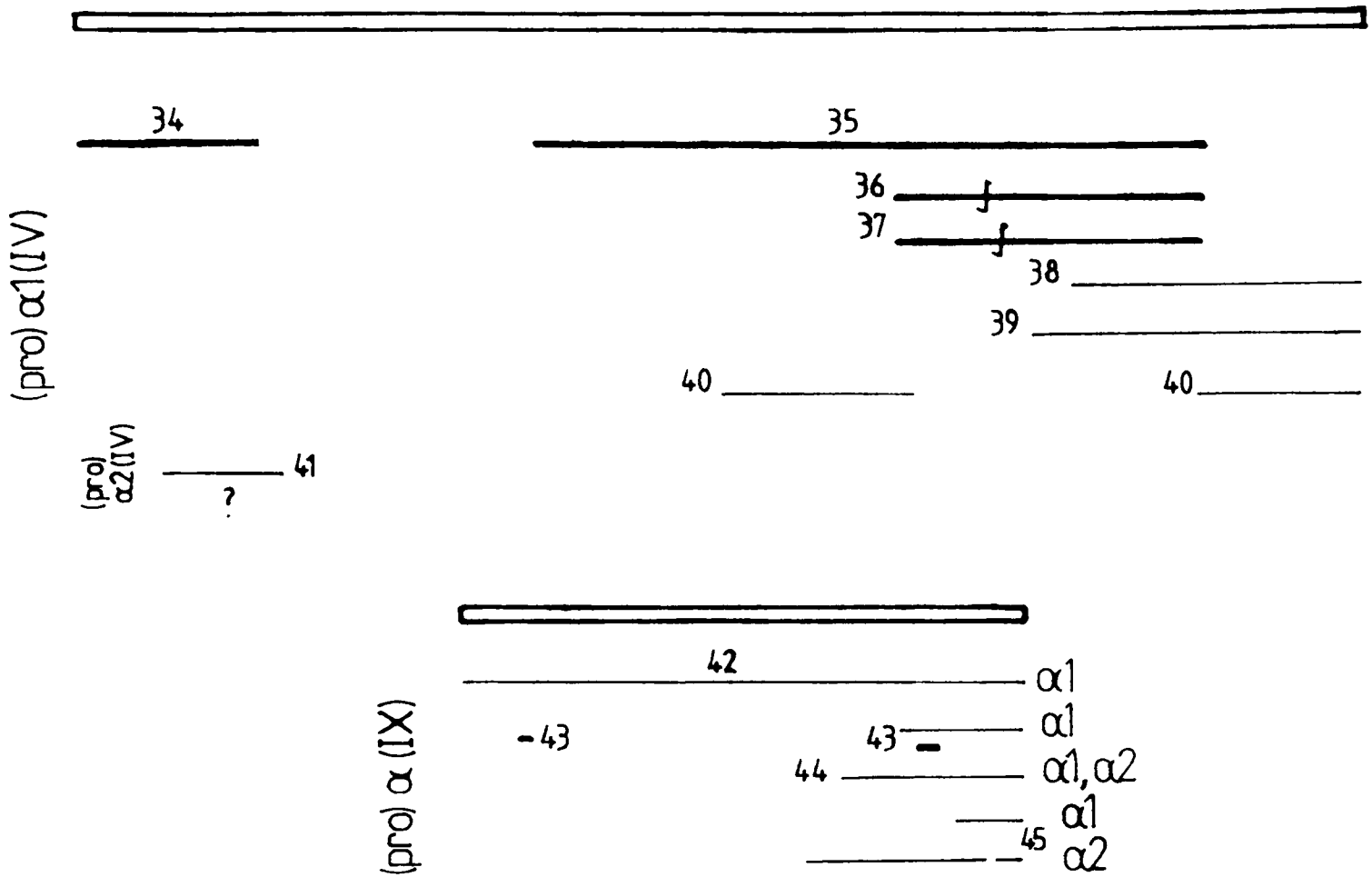


Figure 1.10 AMINO ACID SEQUENCES AVAILABLE FOR REGIONS FROM NINE (PRO)α CHAINS (including signal peptides).

'▭', represent the entire lengths of the pro−α1(I), pro α1(II), pro−α1(III) and pro−α2(V) chains, each consisting of about 1450 amino acid residues; of the pro−α2(I) chain of about 1300 residues, of the (pro)α1(IV) chain of about 1700 residues (minus signal peptide) and of the (pro)α(IX) chains of about 739 residues (minus signal peptide). The domains within the procollagen chains (except for types IV and IX) are demarcated by vertical broken lines and are similar relative to the domains of the pro−α1(I) chain (plus signal peptide): *'s'*, signal peptide (25 residues); *'pN'*, N−propeptide (139 residues); *'tN'*, N−terminal telopeptide (16 residues); *'h'*, 'helical' domain consisting of repeating 'Gly−X−Y' triplets (1014 residues); *'tC'*, C−terminal telopeptide (25 residues) and *'pC'*, C−propeptide (246 residues). The telopeptides and the N−propeptide of the *a*2(I) chain are shorter.

'——————' denotes amino acid sequences derived from *amino acid* sequencing.

'——————' denotes amino acid sequences deduced from *DNA* or *RNA−nucleotide* sequencing, which therefore do not reveal post−translational modifications of proline and lysine. ' ⤴ ' denotes an incomplete sequence. '?' denotes uncertainty as to the exact length of the collagen chain or the exact location of the elucidated sequence.

Numbers *1−46* are references for the respective sequences which are compiled in the *APPENDIX* at the end of this thesis. 'Bovine' sequences: refs. *8, 9, 14, 15, 21, 25, 26, 28, 29, 37.* 'Chick' sequences: refs. *6, 7, 11, 12, 19, 20, 30, 31, 43, 45, 46.* 'Human' sequences: refs. *10, 13, 18, 23, 24, 27, 32, 33, 34, 35, 39, 40.* 'Rat' sequences: ref. *22.* 'Mouse' sequences: refs. *17, 36, 37, 38, 41.* All other sequences are from either of the five species mentioned or are composite sequences. The references (see *APPENDIX*) are selected to give easy access to the literature. Some of these are compilations in reviews while some point the way to other related sequences that are not actually presented in the work referred to.

fig. 1.8). These 'head−to−tail' crosslinks are thought to be the major cross−links which stabilise the 'D−stagger' collagen fibril (see Miller, 1982). Some of the hydroxylysyl residues have also been found to be glycosylated. The possible effects of hydroxylysine glycosides on collagen fibril structure are discussed later (Chapter 1.IIE).

2. Other types of collagen

Like type I collagen, the collagen types II to X are trimers of three α chains, have triple helical domains consisting of repeating 'Gly−X−Y' triplets, and contain hydroxyproline and hydroxylysine. There are however basic differences in tissue distribution, molecular length, extent of triple helical conformation and their abilities to form supramolecular structures outside cells (see fig. 1.11).

a) Types II, III and V

Collagen types II and III are similar to type I collagen in being able to form fibrils and, together, are known as the 'fibrillar' or 'fibre −forming' collagens (see Brodsky and Eikenberry, 1982). Unlike type I molecules, type II and type III molecules are homotrimers of α1(II) and α1(III) chains respectively. Type II collagen is the major collagen of cartilage and also of the vitreous body of the eye while type III collagen is usually found in pliable tissues which also contain type I collagen such as skin, uterus and blood vessels (Bornstein and Sage, 1980). The entire human and calf skin α1(III) chain and about half of the bovine α1(II) chain have been sequenced (fig. 1.10). The 'D' and 'sub−D' periodicities of polar and hydrophobic amino acid clusters found in the α1(I) and α2(I) chains have also been shown in the α1(III) chain (Hofmann et al., 1980). Type III and type II collagens also form fibrils with the characteristic 67 nm 'D' period (Brodsky and Eikenberry, 1982). The α1(III) chain contains two consecutive cysteine residues in the C−terminal region which participate in intramolecular disulphide bonds (Allman et al., 1979).

The current knowledge of type V collagen indicates that it is similar to the fibrillar collagen types I, II and III. The constituent α chains, α1(V), α2(V) and α3(V) are about 1000 amino acids long, and combine in various trimeric permutations as the

A

Type	Chains	Characteristics	Aggregate form	Localization
I	α1(I), α2(I)	Most abundant	67 nm banded fibrils	Skin, bone, etc.
II	α1(II)	Most abundant	Small 67 nm banded fibrils	Cartilage, vitreous humor
III	α1(III)	Cartilage reticulin, abundant	Small 67 nm banded fibrils	Skin, blood vessels, internal organs, etc.
IV	α1(IV), α2(IV)	Basement membrane	Non-fibrillar network	All basement membranes
V	α1(V), α2(V), α3(V)	Abundant	Small fibers	Most interstitial tissues
VI	α1(VI), α2(VI), α3(VI)	Microfibrils	100 nm banded fibrils	Most interstitial tissues
VII	?	Long chain	FLS-like dimer*	Anchoring fibrils
VIII	α1(VIII)	Small helices linked in tandem	?	Some endothelial cells
IX	α1(IX), α2(IX), α3(IX)	Minor cartilage protein, contains attached glycosaminoglycan	?	Cartilage
X	α1(X)	Short chain	?	Hypertrophic and mineralizing cartilage

* The term FLS refers to collagen molecules arranged in an antiparallel, partially overlapping array.

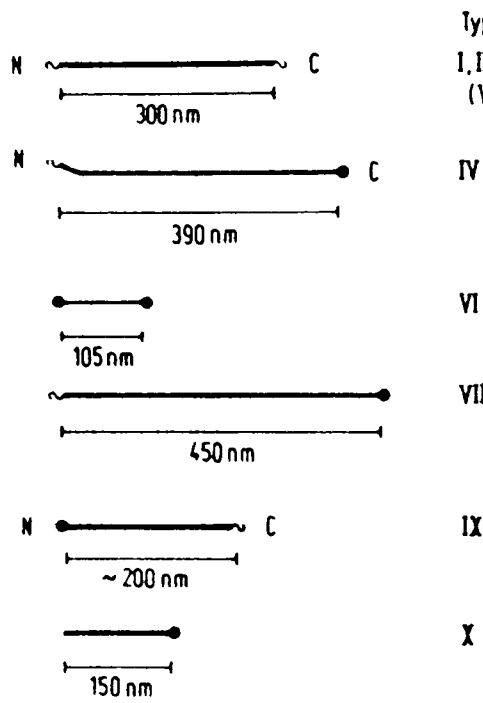


Fig. 1. Scheme of the molecular structures of different collagen types. The triple-helical regions are represented by bars, the non-helical areas by thin lines and the globular domains by circles. The pepsin resistant part of the collagen V molecules resembles the collagens I, II and III. There is, however, evidence that its genuine matrix form contains an additional non-collagenous segment at the N terminus . In the case of types IV and IX collagens, and perhaps others, there are interruptions in the helix which create bends, flexible sites and sites of increased proteolytic susceptibility.

B

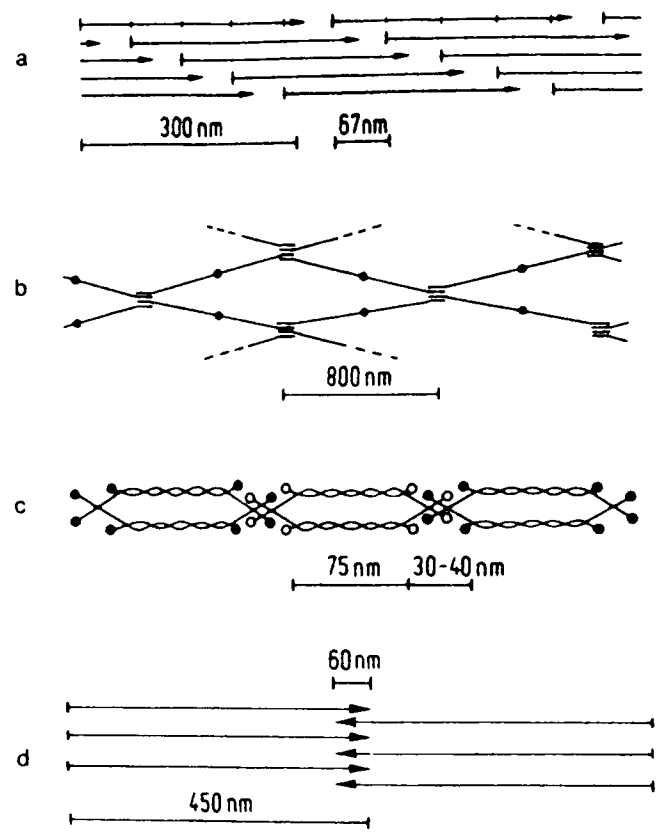


Fig. 2. Models of the macromolecular association of different collagen types. (a) The fiber forming collagens I, II and III. The 300 nm long rod-like molecules aggregate in a parallel and staggered fashion. These fibrils exhibit in the electronmicroscope a characteristic cross-striation period of 67 nm. (b) The basement membrane collagen IV forms a non-fibrillar network. The 400 nm long molecules aggregate only with their identical ends. Four molecules are held together via the triple-helical N-terminus, whereas the C-terminal globular domain connects two molecules with each other. (c) The microfibrillar collagen VI. The aggregation occurs in several steps. Dimers are formed from 105 nm long monomers by antiparallel and staggered alignment, where the 75 nm overlapping helical segments twist around each other forming coiled dimers. After a symmetrical association of dimers to tetramers, fibrillar structures are formed by end-to-end aggregation. The connection of monomers to dimers, tetramers and polymers occurs by disulfide bonds between triple helical segments and globular domains. (d) The long-chain collagen VII. The 450 nm long molecules associate into antiparallel dimeric structures which show a 60 nm overlap and which subsequently assemble laterally with their ends in register.

C

Fig. 1.11 Characteristics of the different collagen types
A. Chain compositions and general characteristics. B. Gross molecular structures. C. Models for their higher-order aggregate forms. (from *Martin et al., 1985*).

type V collagen molecules (see Martin et al., 1985). Type V collagen has been found in many tissues including the corneal stroma (Bornstein and Sage, 1980; Lee and Davison, 1984). The elucidated sequence of 225 amino acids in the 'helical' region and 270 amino acids in the C−terminal region of the pro−α2(V) chain shows some homology with the fibrillar collagen chains (Myers et al., 1985a, 1985b). These include uninterrupted 'Gly−X−Y' triplets, a rigid sequence rich in imino acids, a flexible sequence lacking imino acids, a potential cross−linking 'Lys−Gly−His−Arg' sequence and a similar distribution of polar, charged amino acids. In vitro aggregation studies suggest that type V collagen can form small diameter fibrous structures (Broek et al., 1985).

b) Type IV

Type IV, also known as basement membrane collagen, is present in kidney glomerular basement membrane, lens capsule and Descemet's membrane of the cornea. The type IV collagen molecule is about 400 nm long, being trimer permutations of 1700 residue−long α1(IV) and α2(IV) chains. About two−thirds of the human α1(IV) chain has been sequenced (refer to figure 1.10). At least twelve non−triplet sequences interrupt the 'helical' region. There is no distinct sequence homology with the fibrillar collagen α chains. Analysis of the α1(IV) gene indicates exons which do not seem related to 54 base−pair exons of the fibrillar collagen genes, in keeping with the discovery of interruptions in the 'helical' region of the α1(IV) chain (Kurkinen et al., 1985). The non−helical interruptions give rise to relatively flexible type IV molecules, with a distinctive kink towards the N−terminus, as revealed in electron micrographs (Hofmann et al., 1984; Kühn et al., 1985). The 400 nm long, kinked molecule is thought to be the unprocessed precursor 'procollagen' form which is deposited directly into the extracellular space (see Grant et al., 1981). Aggregation is thought to occur by dimeric association via the globular C−terminal regions and tetrameric association via the N− termini. Many disulphide and lysine−derived cross−links stabilise the tetrameric N−terminal associations (Glanville et al., 1985). The resultant aggregate is a

sheet−like structure described as being not unlike the appearance of 'chicken−wire' (fig. 1.11).

c) Types VI, VII, VIII, IX and X

Collagen types VI , VII, VIII, IX and X are recently discovered collagens (for recent reports see Ayad et al., 1985, 1986; Benya and Padilla, 1986; Kielty et al., 1985) which differ considerably from the fibrillar collagens with respect to α chain length, the occurrence of interruptions in their 'helical' regions and the nature of their aggregates (see Martin et al., 1985; Fleischmajer et al., eds., 1985). With the exception of type VII collagen which has a triple helical region 1.5 times that of the fibrillar collagens, the other collagen types are considerably shorter molecules. Of special note is type IX collagen which is thought to have covalently−attached chondroitin sulphate and dermatan sulphate and thus has also been called a proteoglycan (Bruckner et al., 1985). Characterisation of parts of type IX collagen genes indicates that, like the type IV collagen gene, it too contains exons which do not seem to be related in size, to the 54 base−pair exons of the fibrillar collagen genes (Olsen et al., 1985).

Further characterisation of these recently discovered collagens is required before their in vivo structures and functions and possible evolutionary relationships can be certified. Their basic characteristics, as presently understood, are indicated in figure 1.11.

d) Reassessment of the 'collagen' family

With increasing knowledge of the recently discovered collagens and advances in our knowledge of collagen gene structure it seems likely that in the near future the vertebrate collagen types I−X will fall into more well−defined 'groups' such as the 'fibrillar' collagens. Several collagenous proteins such as the $1\alpha 2\alpha 3\alpha$ 'collagen' (Burgeson and Hollister, 1979), type I trimer 'collagen' (see Bornstein and Sage, 1980) and a reported 55 kilodalton chain of lens capsule (Dixit et al., 1985) are yet to be accorded a collagen type number. It is likely that other proteins may be found to have

collagenous domains as has been found in component C1q of the complement cascade (Reid and Porter, 1981) and acetylcholinesterase (Mays and Rosenberry, 1981). A recently described protein, chondrocalcin, thought to play a role in the calcification of cartilage, has been indicated to be the C-propeptide of type II procollagen, having been cleaved off the helical body of type II procollagen (van der Rest et al., 1986).

In addition to the collagens and collagenous proteins of vertebrates is the separate class of invertebrate collagens (see Murray et al., 1982).

II E. FIBRIL MORPHOLOGY

With respect to the fibrillar collagens, especially type I collagen, several models have been proposed, which incorporate the Petruska-Hodge 'D-stagger' model (fig. 1.8) into the three-dimensional fibril. In addition, several factors may affect the packing of collagen molecules and thus play a role in regulating collagen fibril diameters. There are variations in the diameter of collagen fibrils with respect to tissue source, age, scarring and disease (see Parry et al., 1979; Harding et al., 1980; Eikenberry et al., 1982; Chapter 1.IIF). An understanding of these aspects of collagen structure may reveal the origins of the thin and uniform nature of collagen fibrils in the corneal stroma, features which seem to be essential for corneal transparency (Chapter 1.IB).

1. Packing of collagen molecules

Smith (1968) proposed a five-stranded microfibril which in cross-section shows five D-staggered molecules arranged in a regular pentagon around a central hole which runs through the entire length of the microfibril axis while Veis and Yuan (1975) proposed a four-stranded microfibril. The fibril was therefore thought to consist of lateral aggregates of microfibrils.

Meridional X-ray diffraction patterns have shown the 67nm axial periodicity of collagen fibrils (Brodsky and Eikenberry, 1982) while equatorial X-ray patterns of tendons of rat, rabbit and quail have shown a highly crystalline lateral order (Brodsky et al., 1982). The quasi-hexagonal model for collagen molecular packing (Hulmes and

Miller, 1979; Fraser et al., 1983) proposes a near-perfect hexagonal lattice of collagen molecules in the fibril cross-section. Electron microscopic and X-ray diffraction studies of cryo-sections from rat tail tendon lend support to this model (Chew and Squire, 1986). Too much time need not be spent choosing between the various models proposed as the models are basically exercises that serve to explain, firstly, the high degree of order indicated within collagen fibrils, and secondly, the known and predicted sites of lysine-derived intermolecular cross-links (see Miller, 1982). Geometrically speaking, it is easy to distort the microfibril and square lattice models to fit into a quasi-hexagonal arrangement (Piez and Trus, 1981; Lees, 1986; see fig. 1.12). What does seem important however, is firstly, whether any of the possible 1D, 2D, 3D and 4D intermolecular staggers are favoured which, would predict the importance of particular cross-linking regions. Cross-links in specific, preferred three-dimensional orientations could present particular constraints in processes such as bone mineralisation or on going from the hydrated to the dry fibril state (Lees, 1986). Secondly, if close-packing of molecules, as suggested by the quasi-hexagonal model, is necessary for fibril stability, any loosening of molecular packing would be expected to destabilise the fibril. In the loose-packing microfibril models, any potential obstructions to collagen molecular packing, up to a certain finite size, could in theory be accomodated in unoccupied spaces inside the fibril, occurring between and within microfibrils, as well as in the axial 'holes' of the D-stagger arrangement. The stark light and dark bands of negatively-stained fibrils seem to suggest that within the overlap regions at least, packing is close enough to prevent the penetration of stain. Inside a close-packed fibril, any potential obstructions to the packing of collagen molecules, up to a certain finite size, could not be accomodated within the fibril except in the axial 'holes' of the gap regions. A diagrammatic representation of a hexagonal-close-packed fibril is shown in figure 1.13.

2. Possible influences on fibril formation

It has been suggested that assembly of collagen molecules into fibrils occur within

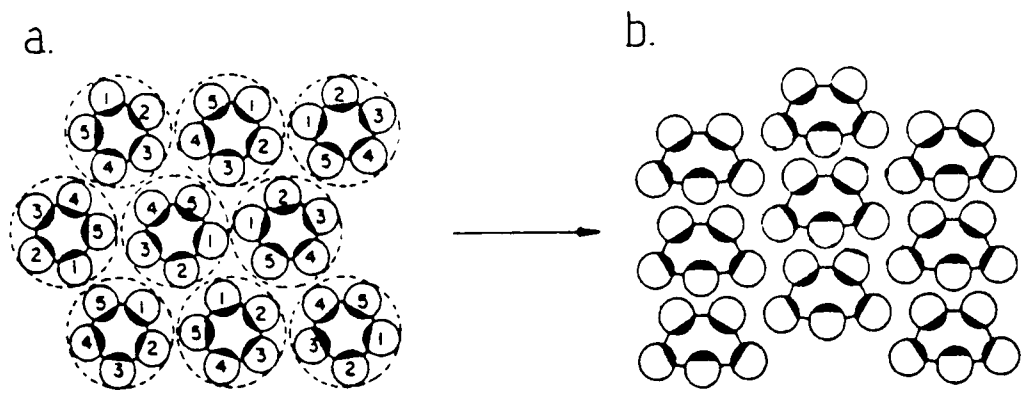


Figure 1.12 End−on views of the type I collagen fibril according to different packing models. a) the five−stranded microfibril model; b) the 'compressed microfibril' model; c) the quasi−hexagonal model: the interplanar spacings *x*, *y* and *z* are 1.33, 1.26 and 1.37nm respectively. Broken lines indicate the crystallographic 'unit−cell'. Numbers *1−5* denote the D−staggered segments as indicated in side−view in fig. 1.8(a). (modified from *Piez and Trus, 1981*).

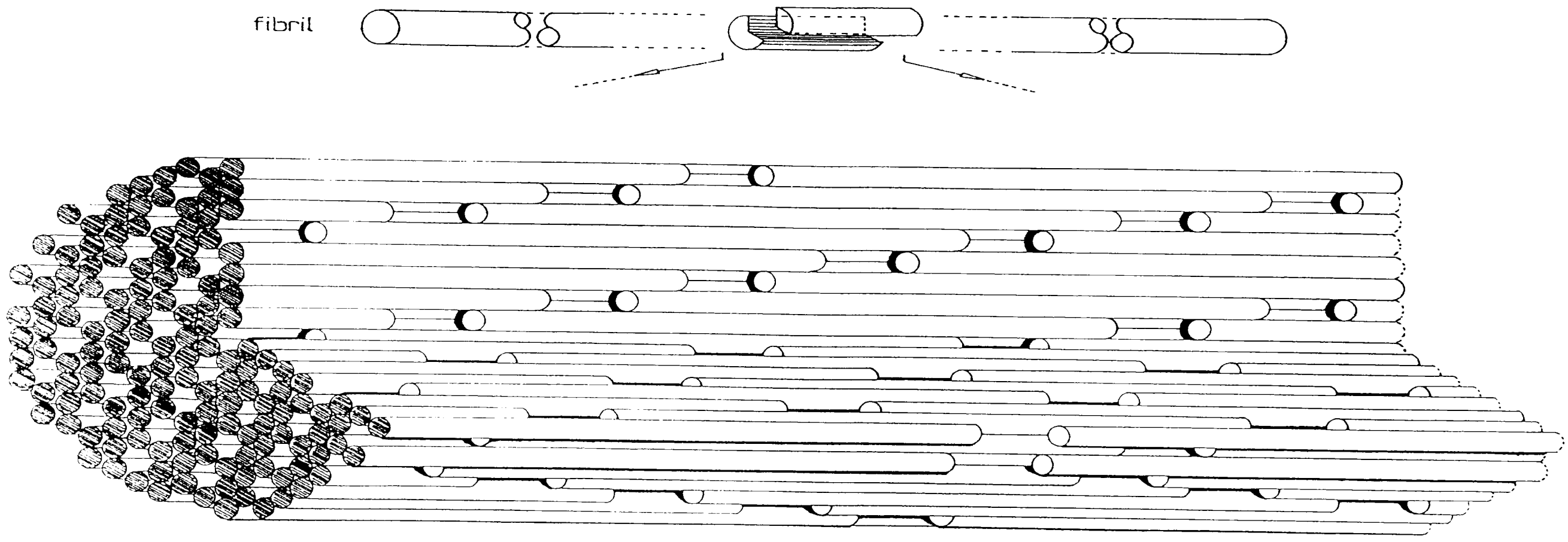


Figure 1.13 Three—dimensional adaptation of the 'D—stagger' arrangement of collagen molecules showing the cross—section and inside of a hypothetical hexagonal—close—packed fibril. Relative staggers between molecules are as shown in figure 1.12(c). The 'hatched' cross—section is through the beginning of a gap region (see fig. 1.8 for comparison). Thick lines are drawn merely to enhance the image, especially with respect to the axial holes which give rise to the *gap* regions of the fibril.

deep pits on the surface of tendon and corneal fibroblasts (Birk and Trelstad, 1984). Although the fusion of these cell−surface pits may determine the interfibrillar arrangement of collagen fibrils such as the lamellar structure of corneal stromal collagen, it seems less likely that the cell surface could exert influences on the extent of lateral fibril growth and hence control fibril diameter except perhaps, indirectly, by creating microenvironments in which other factors may exert their influences.

a) Kinetics of fibril formation

Reconstitution of fibrils from collagen molecules in solution results in an increase in turbidity of the solution. Turbidity−time data from in vitro experiments have led to the proposal of a two−step mechanism for collagen fibrillogenesis (Wood, 1960; Farber et al., 1986). An initial period showing very little change in turbidity, called the 'lag' phase, thought to represent the formation of thin, linear aggregates of collagen, is followed by a period showing a sharp increase in turbidity, or the 'growth' phase, signifying the lateral addition of collagen molecules or linear aggregates to the initial aggregates. The turbidity of the solution levels off at the end of the growth of the growth phase, signifying the end of fibrillogenesis (fig. 1.14). Factors which shorten the lag phase are thought to do so by promoting the formation of many initial fibril 'nuclei' in a short time. In a finite population of collagen molecules, this would lead to the generation of many fibrils of small diameter at the end of the growth phase. Factors which lengthen the lag phase, would lead to the formation of a small number of thick fibrils. Factors such as pH, temperature (Wood and Keech, 1960), molecular species such as proteoglycans (Oegema et al., 1975), fibronectin (Kleinman et al., 1981b; Brokaw et al., 1985) and glucose (Lien et al., 1984) have been found to affect the rate of fibrillogenesis in vitro. Use has been made of turbidity−time data to show that modifications of the N− and C− telopeptide regions or their removal with pepsin impairs fibril formation in vitro (Helseth and Veis, 1981; Capaldi and Chapman, 1982). This has led to suggestions that the telopeptides have essential conformation−dependent roles in facilitating linear aggregation of collagen molecules during fibrillogenesis.

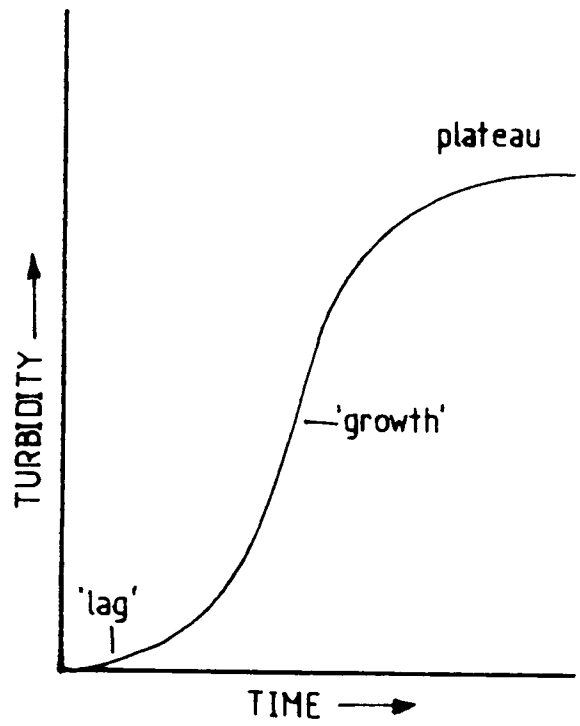


Figure 1.14 Typical turbidity–time plot used in examining the kinetics of collagen fibrillogenesis in vitro.

However, the kinetics of fibril formation in vitro need not reflect accurately the in vivo situation and should be interpreted with caution. For example, Birk and Lande (1981) presented data which showed that reconstituted corneal collagen fibrils were thinner than reconstituted scleral fibrils even though the corneal collagen took 6−7 times longer to reach final turbidity than the scleral collagen. Thus there may be other factors present during fibrillogenesis that may control the extent of fibril growth regardless of any effects that they may have on the kinetics of collagen self−assembly.

b) Proteoglycans

The detection of proteoglycans along collagen fibrils by staining techniques coupled with electron microscopy has led several workers to suggest the non−covalent binding of keratan sulphate and dermatan sulphate proteoglycans at particular sites in the gap regions of type I collagen fibrils (Scott and Haigh, 1985). Although it has been implied that variations in the types of proteoglycans may control fibril diameters by their different abilities to bind to collagen (Scott and Orford, 1981; Scott, 1984), it has also been observed that during wound healing in the rabbit cornea, significant changes in types and sizes of proteoglycans over a 1−year period corresponded to changes in interfibrillar spacing and not to the diameters of collagen fibrils (Hassell et al., 1983). Turbidity−time studies by Chandrasekhar et al. (1984) suggest very little difference between the abilities of corneal and of scleral proteoglycans in limiting collagen fibril diameters although it is known that, on average, scleral fibrils are thicker than corneal fibrils.

c) Intrinsic properties of collagen

c (i) Hydroxylysine glycosides

Harding (1965), in reviewing the data available on the carbohydrate of collagen, concluded that collagen contains covalently−bound hexose. Butler and Cunningham (1966) isolated a hexapeptide 'Gly−Met−Hyl(gal−glc)−Gly−His−Arg' from guinea pig skin collagen and thus demonstrated the covalent linkage of galactose and glucose to

hydroxylysine residues of collagen. The chemical structure of the diglycoside, glucosyl−galactosyl−hydroxylysine (GGH), was identified as 2−O−α−D−glucopyranosyl−O−β−D−galactopyranosyl−hydroxylysine (Spiro, 1967; fig. 1.15). Since then , the occurrence of GGH and the monoglycoside, galactosyl−hydroxylysine (GH), has been confirmed in many collagens (Morgan et al., 1970; Aguilar et al., 1973; Panjwani and Harding, 1978a; Kimura and Karasawa, 1985) and the collagenous complement sub−component C1q (Shinkai and Yonemasu; 1979). Hydroxylysine glycosides result from intracellular post−translational modifications i.e., the hydroxylation of lysine and subsequent glycosylation of hydroxylysine (Chapter 1.IIC).

There are different amounts of hydroxylysine glycoside in different collagens (table 1.1). There is an inverse relationship between fibril diameters and hydroxylysine glycoside content (Schofield et al., 1971). Plotting the available data has shown that this inverse relationship is linear (fig. 1.16). Of the tissue collagens consisting of mainly type I collagen, corneal collagen contains the most hydroxylysine glycoside, and is in the form of thin fibrils while skin, scleral and tendon collagens contain less hydroxylysine glycoside, and are in the form of thicker fibrils. A 'type−I−like' squid cartilage collagen which occurs as thin fibrils also has a high glycoside content (Kimura and Karasawa, 1985). Similarly, the collagens of heart valve and cartilage have high hydroxylysine glycoside contents and occur as thin fibrils while the non−fibrillar nature of lens capsule (type IV) collagen is consistent with its extremely high content of hydroxylysine glycoside (although the molecular shape may also be important in this respect − see Chapter 1.IID2). Vitreous humour type II collagen whose fibrils are thinner than those of cartilage type II collagen, has been found to have the higher hydroxylysine content of the two collagens (Ayad and Weiss, 1984).

The range of fibril diameters within a particular tissue also seem to correlate with glycoside content. Collagens with a high glycoside content and small mean fibril diameters such as those of cornea and cartilage also have a narrow range of fibril diameters. In contrast, collagens with a low glycoside content and large mean fibril

Table 1.1 *Total hydroxylysine-glycoside content, and fibre diameter of collagens from a number of tissues*

Tissue	Hydroxylysine glycoside residues per 1 000 amino acid residues	Fibre diameter (nm)
Calf anterior lens capsule		No fibres or very fine fibres
Bovine glomerular basement membrane	17	No fibres or very fine fibres
Chick cartilage	9	10-20 *
Human cartilage	1.9	10-20
Human cornea	6.4	14-32
Scarred human cornea		20-140
Rabbit cornea	5.8	14-32 *
Scarred rabbit cornea	4.2	up to 50
Calf cornea	5.0	30-35
Human heart valve	4.7	20-30
Rat skin	2.4	65-125 *
Adult human skin	1.4	65-125
Bovine skin	1.5	about 150
Bovine tendon	1.8	30-130 *
Human achilles tendon	1.3	30-130
Rat tail tendon	1.0	30-100
Human sclera	1.7	30-300
Rabbit sclera	1.0	up to 250

* Human.

adapted

(From *Harding et al., 1980*)

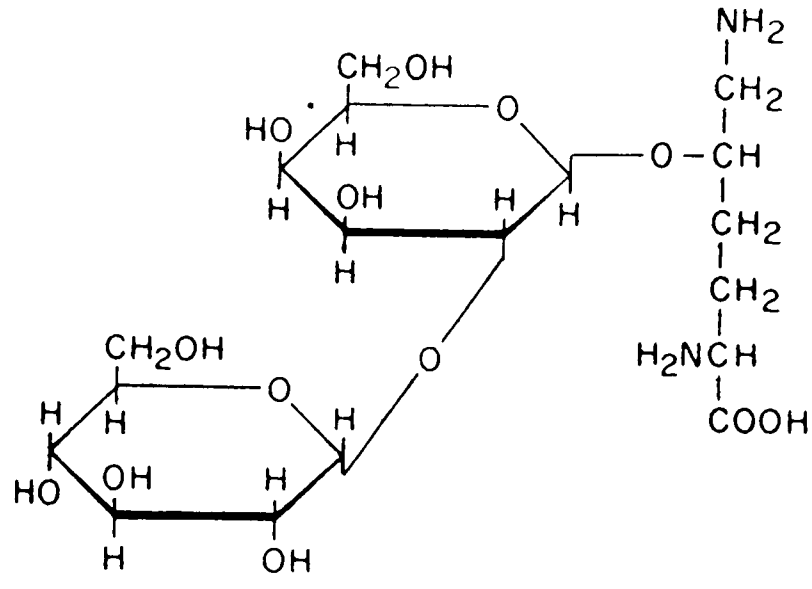


Figure 1.15

The structure of
2−O−α−D−glucopyranosyl−O−β−D−galactopyranosylhydroxylysine.
(From *Butler, 1982*)

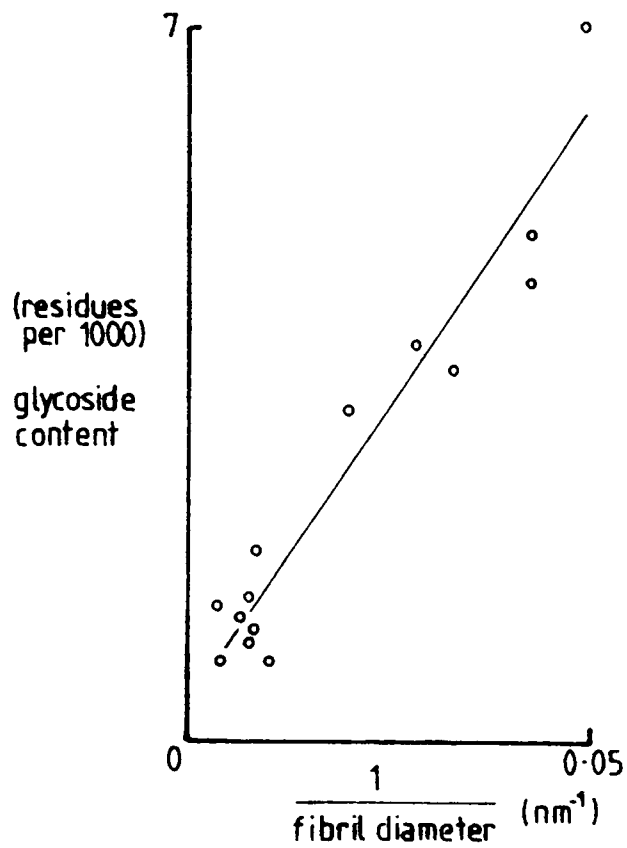


Figure 1.16 The relationship between the glycoside content and the reciprocal of the diameter of the largest collagen fibres in different tissues. Data taken from Table 1.1.
Harding & Panjwani (1980)

diameters, such as those of sclera and tendon, have a wide range of fibril diameters (table 1.1).

Morgan et al. (1970) postulated that hydroxylysine glycosides may inhibit lateral growth of collagen fibrils if they occur at positions along the collagen molecule which are situated in the overlap region of the fibril D−stagger arrangement. Several clinical and experimental observations lend support to the view that hydroxylysine glycosides may regulate fibril diameter. Electron microscopic examination has shown that corneal scars have thicker fibrils than the normal cornea (Schwarz and Graf Keyserlingk, 1969). Studies of rabbit corneal scars (Cintron, 1974) showed that the amount of GGH decreased from a value of 4.9 residues per 1000 amino acid residues in the normal cornea, to a value of 2.7 residues per 1000 amino acid residues in the 2−week old scar. The GGH content then increased to 3.6 residues per 1000 amino acid residues in the 6−week old scar, consistent with the observation that the degree of opacity of corneal scars reaches a maximum and then diminishes. Cetta et al. (1982) demonstrated that with aging, the hydroxylysine glycoside content of rat Achilles tendon decreases significantly, corresponding to an increase in mean collagen fibril diameter. In particular, there was a drop in the GGH:GH ratio with aging. These changes were particularly evident in the last period of pre−natal life and in the first months after birth, corresponding to a rapid increase in collagen fibril diameter. These findings thus also implied that the disaccharide moiety of GGH is more effective in limiting collagen fibril growth than is the smaller monosaccharide of GH. A specific $\alpha(1\rightarrow2)$ glucosidase activity, capable of converting GGH to GH, has been demonstrated in chick, rat and human tissues (Hamazaki and Hotta, 1979; Sternberg and Spiro, 1979; de Luca et al., 1983) which, if present in vivo, could regulate the GGH:GH ratio and hence exert fine control over fibril growth. It was also confirmed that corneal collagen contains more hydroxlysine glycoside than scleral collagen and moreover, that the GGH:GH ratio of corneal collagen is twice that of scleral collagen (de Luca et al., 1983). The data are consistent with the demonstration by Birk and Lande (1981), that corneal collagen forms fibrils 6 to 7 times more slowly than scleral collagen in vitro.

In addition to the steric hindrance of fibril growth expected from the bulk of hydroxylysine glycosides as suggested by Morgan et al. (1970), the hydrophilic nature of the saccharide moieties could favour collagen−water interactions and thus impair collagen−collagen interactions (Hayashi and Nagai, 1972, Lien et al., 1983). These effects may be especially important if they interfere with important cross−link sites in the D−stagger arrangement of collagen molecules. The growing fibril would thus be expected to grow away from glycosylated loci, resulting in a final fibril whose periphery is more glycosylated than its interior. The diameter of such a fibril would thus be determined firstly, by the proportion of molecules which are glycosylated and secondly, by the relative importance of particular glycosylated loci in impeding fibril growth.

c (ii) N−propeptides

Cattle suffering from the condition dermatosparaxis, have weak, fragile skins whose collagen fibrils are thin, corresponding to the occurrence of 'pN collagen', a procollagen−collagen intermediate which retains the N−propeptide (Lenaers et al., 1971). Fleischmajer et al. (1981, 1983), using immunoelectron microscopic techniques, reported the occurrence of N−propeptides on thin, but not thick fibrils of normal adult human and chicken skins and of embryonic chicken skin. It has thus been suggested that the retention of the mainly globular N−propeptides on collagen molecules on the fibril periphery may inhibit fibril growth by steric hindrance. The fibril would thus only grow laterally on removal of these N−propeptides by N−proteinase followed by the accretion of more pN collagen (Hulmes, 1983). It was however pointed out that if N−propeptides do control fibril growth by steric hindrance, they would not constitute the only form of control as the thicker fibrils would then be expected to have N−propeptides on their fibril peripheries too. In addition, Scott (1984) reported only a small quantity of N−propeptides in developing calf flexor tendons and concluded that only about 1% of the molecules on the fibril peripheries would be pN collagen, a proportion insufficient to impede fibril growth.

It is possible that removal of N−propeptides from type III procollagen is less complete than for type I procollagen as Fleischmajer et al. (1981) reported that labelling of skin collagen with antibodies to type III N−propeptides only labelled the thin fibrils. Similarly, it has been suggested that retention of N−propeptides on type V collagen may prevent it from forming thick fibrils (Broek et al., 1985).

c (iii) Hybrid fibrils

As the different collagen types seem to differ in their fibril−forming capabilities perhaps due to their different lengths, shapes and biosynthetic modifications such as glycosylation, 'hybrid' fibrils consisting of two or more collagen types could be expected to exhibit 'hybrid' characteristics. In vitro aggregation studies have led Lapière et al. (1977) to suggest that hybrid type I/type III fibrils exhibit intermediate fibril diameters. In the embryonic avian cornea, it was suggested that 'type I' fibrils also contain type V collagen molecules as type V collagen was reported to be 'unmasked' in immunohistological studies, by the action of acetic acid (Linsenmayer et al., 1983; Fitch et al., 1984). However, this unmasking effect was not observed in bovine cornea (Lee and Davison, 1984). It was also observed that an increase in type V collagen from 3% to 10% of the total stromal collagen in the developing fetal calf cornea occurred during a period in which there is no change in collagen diameter (Lee and Davison, 1984). Thus, either the inclusion of type V collagen into predominantly type I collagen fibrils does not affect fibril diameter or hybrid type I/ type V fibrils do not exist in the cornea.

Nevertheless, if more evidence such as the report of a type I to type III collagen crosslink (Henkel and Glanville, 1982) accumulates to verify the existence of hybrid fibrils, the modulation of collagen type synthesis as seen in development for example (see Bornstein and Sage, 1980; Müller et al., 1981), may be important in regulating the diameters of hybrid fibrils.

II F. HERITABLE COLLAGEN DISORDERS

So far, we have seen how the self−assembly properties of procollagen α chains and of collagen molecules are dependent on collagen primary structure and complex biosynthetic modifications. As collagen is the major structural protein of the body, genetic abnormalities which lead to defects in collagen primary structure or biosynthesis thus lead to many clinical disorders. These disorders mostly involve type I collagen while some involve type III collagen (see Prockop and Kivirikko, 1984; Sykes and Smith, 1985).

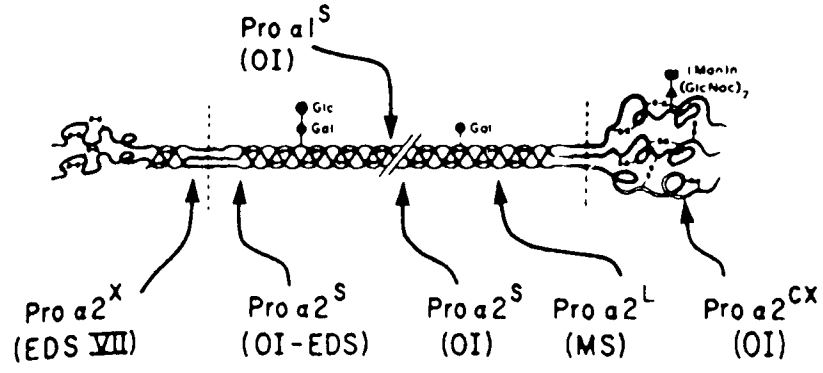
Osteogenesis imperfecta (O.I.) is a clinically, genetically and biochemically heterogeneous group of connective tissue diseases characterised by brittle bones and abnormalities in most other tissues containing type I collagen. Sykes et al. (1977) showed a decrease in the ratio of type I to type III collagen in skin of O.I. patients. They suggested that this was probably due to a decrease in the amount of type I collagen which, if extended to bone, would cause bone fragility. Since then, analyses of type I collagen genes and and gene products from O.I. patients have revealed mutations in collagen genes such as deletions (Barsh et al., 1985) and substitutions (Steinmann et al., 1984) which correspond to decreases or structural alterations in the gene products. The defective pro−α 1 or pro−α 2 chains may be shortened or altered within the helical or propeptide domains (see table 1.2, fig. 1.17). These defects could lead to impaired chain selection and triple−helix formation causing a decreased production of normal type I collagen or the secretion of abnormal collagen with poor fibril−forming properties. A deletion within one of the α1(I) alleles can lead to 75% of the type I procollagen molecules being defective as each molecule contains two pro−α1 chains. If these molecules are unstable, they may be degraded, leading to the deposition of only 25% of the normal amount of type I collagen with often, lethal consequences (see Sykes, 1985). Studies of O.I. mutant collagen genes by the characterisation of restriction enzyme sites in and around type I collagen genes, coupled with linkage studies (Sykes et al., 1986) as well as nuclease S1 mapping (Dickson et al., 1984) and 'R−loop' analysis are currently under way, with the

Table 1·2 Consequences of Molecular Defects in Four Heritable Diseases of Collagen.

Disease	Defect *	Consequences
Osteogenesis imperfecta		
Type I	Proα1(I)°	Half normal amount of Type I collagen
	Proα2(I)^{S}	Probably abnormal fibrils
	Other (unidentified)	
Type II	Proα1(I)^{S}	Unstable triple helix; increased synthesis of proα1(III)
	Proα2(I)^{S} and Proα2(I)°	Uncertain
	Other (unidentified)	
Type III	Proα2(I)^{CX}	Synthesis of proα1(I) trimers
	Proα1(I)^{CX} or Proα2(I)^{CX}	Increased mannose in C-propeptide and decreased solubility of Type I procollagen
	Other (unidentified)	
Variant with characteristics of Ehlers–Danlos syndrome	Proα2(I)^{S}	Resistance to procollagen N-proteinase and persistence of pNcollagen †
Marfan syndrome	Proα2(I)^{L}	Probably abnormal cross-linking
	Other (unidentified)	
Ehlers–Danlos syndrome		
Type IV	Proα1(III)$^{+}$	Marked decrease in Type III collagen
	Proα1(III)^{SM}	Unstable triple helix
	Proα1(III)^{X}	Decreased secretion rate of Type III collagen
	Other (unidentified)	
Type VI	Lysine hydroxylase deficiency	Hydroxylysine-deficient collagen and defective cross-linking
	Other (unidentified)	
Type VII	Procollagen N-proteinase deficiency	Persistence of pNcollagen †
	Proα2(I)^{X}	Resistance to procollagen N-proteinase and persistence of pNcollagen †
Type IX	Defective copper metabolism	Lysine oxidase deficiency and defective cross-linking
Menkes syndrome	Defective copper metabolism	Lysine oxidase deficiency and defective cross-linking

*Proα1(I)°, proα2(I)°, and proα1(III)$^{+}$ denote nonfunctional or inefficiently functioning alleles for proα chains; proα1(I)^{S} and proα2(I)^{S}, shortened proα chains; proα2(I)^{L}, lengthened proα chains; proα1(I)^{CX} and proα2(I)^{CX}, mutations altering the structure of the C-propeptides of proα chains; proα1(III)^{SM}, an altered proα1(III) chain that migrates slowly in electrophoretic gels; and proα1(III)^{X} and proα2(I)^{X}, poorly defined mutations altering the structure of proα chains.

†Intermediate in the conversion of procollagen that contains the N-propeptides but not the C-propeptides.



Approximate Locations of Mutations in the Structure of Type I Procollagen.

EDS denotes the Ehlers–Danlos syndrome, MS the Marfan syndrome, and OI osteogenesis imperfecta. For explanation of other abbreviations, see first footnote to Table 1.2

Figure 1.17

(From *Prockop and Kivirikko, 1984*)

possibility that pre−natal diagnosis of some forms of O.I. may soon be feasible.

Of interest to the current project, is the common finding, in some cases of O.I., that the expected delay in triple helix formation for example where glycine is replaced by cysteine in the triple helical region, leads to the secretion of collagen which is overhydroxylated and overglycosylated with respect to lysyl residues (Kirsch et al., 1983; Traub and Steinmann, 1986). Electron microscopic examination of these O.I. collagen fibrils indicate a decrease in average fibril diameters, from 80nm to 50 nm in human O.I. (Byers et al., 1983; Steinmann et al., 1984), as well as in bovine O.I. (Denholm et al., 1985). This is consistent with an inhibitory effect of hydroxylysine glycosides on lateral fibril growth as discussed previously.

The various types of Ehlers−Danlos syndromes (see Prockop and Kivirikko, 1984; Byers and Holbrook, 1985) are characterised by joint hypermobility and thin, stretchy skin. Ehlers−Danlos syndrome type VII has been correlated with incomplete cleavage of N−propeptides due to collagen−gene alterations in the propeptide cleavage site of the α2(I) chain (Eyre et al., 1985) or a deficiency of N−proteinase. Meanwhile, Ehlers−Danlos syndrome type VI has been correlated with a deficiency of the enzyme lysyl hydroxylase, and Ehlers−Danlos syndrome type IX , with a defective copper metabolism (see table 1.2). In the former case, the clinical symptoms would thus result from a decrease in hydroxylysine−derived cross−links, while in the latter case, a decrease in the activity of the copper−dependent enzyme lysyl oxidase leads to a decrease in all lysine− and hydroxylysine− derived cross−links.

II G. COLLAGEN OF THE CORNEAL STROMA

The complex biochemistry of collagen discussed in this chapter so far, also apply to corneal collagen. The thin, uniform nature of collagen fibrils in the corneal stroma is essential for corneal transparency. Particular features of corneal (stromal) collagen will now be dicussed.

Collagen fibrils of the corneal stroma, when stained and examined under the electron microscope, show the same pattern of cross−striations shown by other

collagens with the characteristic 67 nm axial periodicity. The axial periodicity has been confirmed by X−ray diffraction studies of fresh cornea (fig. 1.7; Goodfellow et al., 1978; Meek et al., 1981, 1983). Within corneal stromal fibrils, collagen molecules are therefore packed in the 'D−stagger' arrangement. Itoi (1961) and Lewis (1967) ascertained the size and shape of the corneal collagen molecule as being identical to that from other fibrillar collagens.

The possibility that the corneal stroma contains more than one genetic type of collagen has attracted considerable attention. In characterising corneal stromal collagen, it has to be ensured that that the collagen being analysed does not include that of Descemet's membrane (which is mainly type IV collagen), epithelium and endothelium, nor that of contaminating sclera (the sclero−corneal junction contains type III collagen). As most of corneal collagen is not extractable with mild acid, most biochemical analyses have been made on pepsin−solubilised collagen (e.g. see Welsh et al., 1980). It is thus important to note the proportion of collagen solubilised as well as the possibility of artifacts arising from degradation by pepsin.

Identification of collagen types in the cornea has involved examination of solubility properties, electrophoretic and chromatographic properties (of collagen α chains and cyanogen bromide peptides), amino acid compositions and immunological staining properties. In examining the results however, it is important to consider whether for example, all the analysed sample enters the gel during gel electrophoresis, as high molecular weight cross−linked α chains or peptides may be excluded. Similarly, in immunological staining studies, the proportion of collagen molecules accessible to the antibodies need to be considered and perhaps, even whether the antibodies prepared are entirely type−specific, considering the high degree of homology between the fibrillar collagens (see von der Mark, 1981).

Many reports agree that type I collagen is the major collagen type in mammalian corneas (Praus et al., 1979; Tseng et al., 1982; Newsome et al., 1982; Lee and Davison, 1984). About 95% of rabbit (Welsh et al., 1980) and 85% of bovine (Lee and Davison, 1984) corneal stromal collagens is type I collagen, the remainder being

type V collagen with trace amounts (less than 1%) of type III collagen. The proportion of type I collagen in the bovine cornea was arrived at by semi−quantitative optical density scanning of gel−electrophoretic patterns. The chick corneal stroma also contains mainly type I collagen as shown by immunofluorescence studies which also indicate the absence of collagen types II and III (von der Mark et al., 1977). Nakayasu et al. (1986) suggest a significant occurrence of type III collagen in the human corneal stroma on the strength of immunological studies, with the implication of the existence of hybrid type I/type III fibrils. The immunohistological and immunoelectron microscopic evidence presented however, should perhaps be verified by other means. A recent report (Zimmerman et al., 1986) suggesting that type VI collagen may constitute as much as a quarter of human corneal collagen, should be treated with caution as large amounts of pepsin (about 30% of the corneal dry weight) were used in the extraction procedures, which, apparently only solubilised 5 to 10% of the corneal dry weight. Furthermore, it was not indicated whether the Descemet's membrane, epithelial and endothelial layers were removed. Extrapolation of these results to the whole cornea may thus have been misguided and prone to misinterpretation of possible artefacts. There have also been suggestions that corneal type I collagen may be coded by genes other than those coding for type I collagens in other tissues (Kao and Foreman, 1980; Church, 1981) stemming especially from somatic cell hybridisation studies which showed that the corneal procollagen gene was on chromosome 7 while that of other type I procollagens was on chromosome 17. We now know that chromosomes 17 and 7 contain the pro−α1(I) and pro−α2(I) genes respectively and the suggestion of a genetically distinct corneal type I collagen has been withdrawn (Kao, 1985). Corneal scars also appear to contain mainly type I collagen (Cannon and Cintron, 1975) in contrast with other scars where increases of type III collagen have been found.

Stabilisation of collagen molecules within bovine corneal collagen fibrils seem to occur via some cross−linking sites other than or in addition to those found in other type I collagens, involving peptides α2−CB3,5 and α1−CB6. A cross−linked dimer of

peptide α1−CB6 has also been isolated from bovine cornea but this most probably represents an intramolecular cross−link (see Harding et al., 1980).

The high content of hydroxylysine glycoside in corneal stromal collagen relative to other type I collagens such as tendon and sclera has been discussed in a previous section (Chapter 1.IIE2).

III. INTRODUCTION TO THE PRESENT WORK

The possibility that hydroxylysine glycosides limit the growth of collagen fibrils is the central issue of the current work. Morgan et al. (1970) postulated that glycosides occurring in the 'overlap' regions of the fibril D−stagger arrangement can impair the lateral packing of collagen molecules by steric hindrance while those in the 'gap' regions can fit into the axial holes. It is thus important to know the positions of glycosides along collagen molecules in relation to the overlap and gap regions of the fibril. It is also important to know the nature or form of the glycoside i.e. whether the covalently attached saccharide moiety is the monosaccharide, galactose, or the disaccharide, glucosylgalactose, as the larger disaccharide may be expected to have the greater inhibitory effect on fibril growth. Furthermore, post−translational hydroxylation of lysine and glycosylation of hydroxylysine do not necessarily go to completion at any particular position along collagen α chains (see Bornstein and Traub, 1979; Panjwani and Harding, 1978b). Thus, a particular position may consist of 20% lysine, 20% unglycosylated hydroxylysine, 20% galactosylhydroxylysine (GH) and 40% glucosylgalactosylhydroxylysine (GGH). This means that in a population of collagen molecules, two out of every ten α chains contain lysine (unmodified) at that position and so forth. It is thus important to determine 'percentage' glycosylation in addition to the nature and position of the glycoside.

With the increasing use of nucleotide sequencing techniques to determine protein primary structure, it is important to note that, as hydroxylysine glycosides result from the post−translational modifications of lysyl residues, nucleotide sequencing would only

reveal the ultimately modified residues as 'lysine'. In addition, the current capabilities of Edman degradation in amino acid sequencing do not facilitate positive identification of hydroxylysine glycosides (Butler, 1982), as their hydrophilic nature probably results in the inability to extract their derivatives from the 'spinning−cup' of the sequencing apparatus (Schuppan et al.; 1982, Babel and Glanville, 1984). However, with prior knowledge of the collagen amino acid sequence, it is possible to locate the sites of hydroxylysine glycosides. Successive fragmentations of the collagen α chains by the use of cyanogen bromide and proteolytic enzymes yield relatively small peptides which, when isolated, can be identified by matching their amino acid compositions to the known amino acid sequence. The presence of hydroxylysine can be ascertained by amino acid analysis of acid hydrolysates of the peptides and that of hydroxylysine glycosides by amino acid analysis of the alkaline hydrolysates of the peptides (Butler, 1982). The need to analyse both acid and alkaline hydrolysates stems from the fact that glycosidic bonds are cleaved during acid hydrolysis but not, during alkaline hydrolysis (Spiro, 1967). Alkaline hydrolysis however, destroys some of the common amino acids (Hill, 1965) but not others including hydroxylysine. Having determined the hydroxylysine glycoside within the identified peptide, the exact position of the hydroxylysine glycoside can thus be placed in the known sequence by deduction.

In the α1(I) chain of calf skin collagen, there are 0.3 moles of GH and 0.3 moles of GGH at position 87, and 0.3 moles of GH at position 684 (positions are counted from the start of the 'helical' sequence). In the α2(I) chain, 0.6 moles of GH and small amounts of GGH occur at positions 87 and 174. These all occur in the gap regions of the D−stagger fibril structure and thus could fit into the axial holes. The remainder 0.7 moles of GH per molecule, are still to be located among seven hydroxylysyl residues of peptide α2−CB3,5 (see Fietzek and Kühn, 1976).

In bovine corneal collagen, Panjwani and Harding (1978b) have located 0.2 moles of GH and 0.1 moles of GGH at position 408 of the α1(I) chain, occurring in a 'gap' region. More important, they also located 0.17 moles GH and 0.41 moles of GGH at position 531 of the α1(I) chain, which is in an 'overlap' region. This is the only

hydroxylysine glycoside found in the overlap region of a type I collagen.

The current work continues the identification of the sites of hydroxylysine glycosides in bovine corneal collagen initiated by Panjwani and Harding (1978b). Bovine corneal collagen was purified and then fragmented specifically using cyanogen bromide. Some of the fragments were isolated and characterised (Chapter 3). Particular attention was given to cyanogen bromide peptide α1(I)−CB7 (Chapter 4), a 271−residue peptide which accounts for about 26% of the α1(I) chain. General examination of the covalent structure of peptide α1−CB7 of corneal collagen is also of importance, as this peptide also contains the single Gly−Ile peptide bond cleaved by animal collagenase and the binding site for fibronectin (see Kleinman et al., 1981a). The amino acid sequence of peptide α1−CB7 (of type I collagen, unless otherwise stated), of calf skin collagen was originally reported by Fietzek et al. (1973) while that of chick skin collagen was reported by Highberger et al. (1975). These sequences have since been corrected to include three amino acids which were missed in the original reports (Kleinman et al., 1978; Highberger et al., 1978).

In this work, precise location of hydroxylysine glycosides in peptide α1−CB7 was facilitated by the isolation of small peptides resulting from digestions with trypsin and V8−protease (from *Staphylococcus aureus*, strain V8). These overlapping peptides were characterised to reveal the exact locations of the glycosides GH and GGH as well as the 'percentage−glycosylation' at these sites (Chapter 4). In the final chapter of this thesis, the discovered hydroxylysine glycosides will be discussed with respect to their possible role in controlling the diameter of corneal collagen fibrils.

CHAPTER 2

EXPERIMENTAL MATERIALS AND METHODS

I. MATERIALS

A. CORNEAS AND LENSES

Bovine eyes were obtained from J. Meade Abbatoir (Slough) and F.M.C. Harris (Thame) Ltd. They were removed soon after slaughter and stored at –20°C. At the laboratory, these were thawed and the corneas and lenses were removed for storage at –20°C until further use.

B. CHROMATOGRAPHIC MATERIALS

Carboxymethyl cellulose (CM 52) cation exchanger was bought from Whatman Biochemicals Ltd. (Springfield Mill, Kent).

Sephadex G−50, superfine grade was bought from Pharmacia (UK) Ltd.

Bio−Rad AG50W−X4, minus 400 mesh cation−exchange resin was bought from Bio−Rad Laboratories (Watford).

Ion−exchange 'amino acid resin' D−X8.25 size 8±1 micron, used in the separation of amino acids and hydroxylysine glycosides on the amino acid analyser was bought from Benson Company (Reno, U.S.A.).

Buffers used for amino acid analysis were bought from LKB Instruments Ltd. (Surrey, England).

C. ENZYMES AND CHEMICALS

Cyanogen bromide, citraconic anhydride, bovine trypsin (type XIII, TPCK−treated), collagen amino acid standards and dansyl amino acid standards were bought from Sigma Chemical Co. Ltd. (Poole).

V8−protease (from *Staphylococcus aureus*, strain V8) was bought from Miles Laboratories Ltd. (Slough).

Other chemicals were bought from 'Sigma' (Poole) and BDH Chemicals Ltd. (Atherstone) and were of AnalaR or Aristar grade.

II. METHODS

II A. PREPARATIONS OF COLLAGEN FROM CORNEAS AND LENS CAPSULES

Bovine corneas, stored at –20°C, were thawed at room temperature. Any remaining scleral tissue on the corneal periphery was trimmed off with a pair of scissors. Placing the cornea on a glass plate, the epithelium and endothelium were scraped away with a scalpel. A shallow cut was made, just deeper than the Descemet's membrane, dividing the membrane into two halves. The membrane was then effectively peeled off the stroma, one half at a time, with a firm scraping action from the cut to the periphery. The remaining stroma was then cut with a pair of scissors into pieces about 2 × 2 mm. The diced stroma was then placed in 15 volumes of 4M guanidinium hydrochloride/50mM Tris, pH 7.5 and stirred over a magnetic stirrer for a period of 18 hours (Miller and Lunde, 1973) to extract proteoglycans and other soluble components into solution. During the 18–hour extraction, the solution was changed five times. The corneal stroma was then washed with distilled water at 4°C until free of guanidinium hydrochloride. The stroma was assessed as 'completely washed' when the discarded wash no longer gave a white precipitate when a drop of 2% aqueous silver nitrate was introduced. After draining, the washed stroma was freeze–dried on a freeze–dryer (Birchover Instruments Ltd., Letchworth). Batches of 100 corneas were processed in the manner described to yield 5 to 6 gram stocks of freeze–dried corneal stromal collagen.

Bovine lenses, stored at –20°C, were also thawed at room temperature. The lens capsules were removed, scraped gently on both sides and prepared in the same way as the corneas but without cutting them into small pieces.

Freeze–dried corneal (stromal) collagen and lens capsule collagen were stored at –20°C until further use.

II B. CLEAVAGE OF CORNEAL COLLAGEN WITH CYANOGEN BROMIDE

Freeze-dried corneal collagen (2g) was immersed in 70% formic acid (600ml), flushed with nitrogen. Cyanogen bromide (3g) was introduced and the reaction flask was incubated at 40°C inside a fume-cupboard. The flushing with nitrogen was intensified to keep the mixture stirring. At 3¾ hours after the introduction of cyanogen bromide (Panjwani and Harding, 1978a), the cleavage reaction was arrested by diluting the mixture five-fold with distilled water pre-cooled to 4°C and storing this at 4°C. Two rotary evaporators were then operated over several hours at 40°C to remove cyanogen bromide, formic acid and water from the mixture, evaporating 300ml at a time down to a small volume and repeating this until all of the mixture was concentrated in two flasks. The concentrate was then diluted with 300ml distilled water and rotary-evaporated to a small volume. This was repeated until the final mixture was free of cyanogen bromide and formic acid as tested by smell (with caution). The cyanogen bromide digest was then freeze-dried.

II C. CLEAVAGE OF PEPTIDE α1-CB7 WITH PROTEOLYTIC ENZYMES

1. Cleavage with trypsin

a) Citraconylation of peptide α1-CB7

Prior to tryptic cleavage of purified peptide α1-CB7, 'citraconylation' was carried out to mask the lysyl residues from tryptic cleavage and hence restrict the cleavage to the carboxyl side of arginine residues (Dixon and Perham, 1968).

Peptide α1-CB7 (60mg) was dissolved in 0.5M potassium phosphate buffer, pH 8.2 (Slaby and Holmegren, 1975) and stirred with a 30-molar excess of citraconic anhydride, added in four 20μl aliquots. The pH was maintained at 8.1-8.2 by the addition of 5M potassium hydroxide. When the pH had stabilised after the final 20μl aliquot was added, the mixture was stirred for 1 hour at room temperature and then dialysed against five changes of 1-litre volumes of distilled water (brought to pH 8 with ammonia) to remove potassium phosphate and unreacted citraconic anhydride.

b) Tryptic cleavage of citraconylated peptide α1−CB7

Freeze−dried, citraconylated peptide α1−CB7 (63mg) was dissolved in 11ml of 200 mM ammonium bicarbonate/1mM calcium chloride pH 7.9 (Kang et al., 1975) and brought to 37°C in a shaking water bath. To this, 120μl 0.5% trypsin in 0.11mM HCl was added to start the cleavage. After 2 hours, another 120μl of the trypsin solution was added, the reaction being allowed to proceed for another 2 hours. The 4−hour cleavage was stopped by acidifying to pH 3.25 with 1M acetic acid. The tryptic digest was then freeze−dried.

Prior to chromatography on Sephadex G−50, the tryptic digest was stirred in 5ml pyridine acetate buffer (0.2M pyridine), pH 3.1 at room temperature for 3 hours in order to remove citraconyl groups from the tryptic peptides (Panjwani and Harding, 1978b). The tryptic digest was then rotary−evaporated and freeze−dried.

2. Cleavage with V8−protease

a) In 100mM ammonium bicarbonate, pH 7.8

Cleavage with V8−protease (also known as staphylococcal protease) in this buffer restricts cleavage to peptide bonds on the carboxyl side of glutamic acid residues (Drapeau, 1977).

Purified peptide α1−CB7 (71mg) was dissolved in 4.7ml 100mM ammonium bicarbonate/2mM EDTA pH 7.8. EDTA was included to inhibit any neutral proteases present (Arvidson, 1973). Cleavage was commenced in a shaking water bath at 37°C by the addition of 2.4mg V8−protease (1:30 enzyme:substrate ratio). At 15½ hours, the reaction mixture was acidified to pH 4.8 with glacial acetic acid and loaded onto the Sephadex G−50 column.

In an earlier exploratory cleavage, a 28mg batch of peptide α1−CB7 was cleaved with V8−protease using the same enzyme:substrate ratio.

b) In 50mM potassium phosphate buffer, pH 7.8

A sub−peptide of peptide α1−CB7 (11.5mg) was dissolved in 820μl 50mM potassium phosphate/2mM EDTA, pH 7.8 and 380μg of V8−protease (1:30 enzyme:substrate ratio) was introduced. In this medium, cleavage occurs at peptide bonds on the carboxyl side of aspartic acid and glutamic acid residues (Drapeau, 1977). The cleavage reaction was carried out in a shaking water bath at 37°C for 18 hours after which the V8−protease digest was brought to pH 4.8 with glacial acetic acid and loaded onto the Sephadex G−50 column.

In an earlier exploratory cleavage, a 5.2mg batch of peptide was digested using the same enzyme:substrate ratio.

II D. SEPARATION OF COLLAGEN PEPTIDES BY COLUMN CHROMATOGRAPHY

1. Ion−exchange chromatography of CNBr peptides on CM 52

a) In 20mM sodium citrate buffer, pH 3.6

This was performed essentially as described by Volpin and Veis (1973) except for a few changes designed to increase the yield of peptide α1−CB7.

Carboxymethyl cellulose (CM 52) was equilibrated with 200mM sodium citrate buffer (ten times the concentration of the eluting buffer), pH 3.6. This was degassed and a 70% slurry was prepared. A glass tube measuring 145 × 1.9 cm, within a water−jacket kept at 40°C, was set upright with an extension tube attached to the top. Degassed equilibrating buffer was poured in to a height of 10 cm. Air bubbles at the bottom column piece were removed through the outlet tubing by suction into a syringe. The slurry was poured down the extension tube and the resin was allowed to pack under a head of pressure controlled by raising or lowering the end of a length of extension tubing connected to the outlet. The rate of effluent flow was kept at about 1.5 ml per minute during column−packing. When the resin had packed to a height of 140 cm, the column extension was removed and 2 litres of starting buffer

(see below) was pumped through the column at 120 ml per hour to ensure, initially, a constant bed height and secondly, that the pH and ionic strength of the column were identical to that of the starting buffer at the start of chromatography. More of the starting buffer was pumped until the pH and conductivity of the effluent equalled that of the starting buffer.

The eluting buffer system used was 20mM sodium citrate buffer pH 3.6 with a linear sodium chloride gradient using two identical flasks as the gradient chambers. Two different sodium chloride gradients were used on separate occasions:

[i] A linear gradient of 40−140 mM sodium chloride over a total volume of 4 litres.

[ii] A linear gradient of 60−100 mM sodium chloride over a total volume of 2 litres.

The column was washed through with 200ml of 10mM sodium hydroxide, re−equilibrated with equilibrating buffer and the appropriate starting buffer between fractionations.

Freeze−dried CNBr digest of corneal collagen (1.6g) was suspended in the appropriate starting buffer (24ml) and centrifuged at 10,000 × g for 20 minutes. The supernatant was loaded onto an almost drained column−bed and gradient elution at 40°C was carried out at a flow−rate of 100 ml per hour.

b) **In 30mM sodium acetate buffer, pH 4.8**

Further purifications of peptides α1−CB7 and α1−CB8 were carried out on a CM 52 column, 40 × 1.9 cm. The equilibrating buffer was 300mM sodium acetate buffer, pH 4.8 while the eluting buffer was 30mM sodium acetate, pH 4.8 with a sodium chloride gradient (Volpin and Veis, 1973).

Two different sodium chloride gradients were used on separate occasions, both over a volume of 1200ml between two identical 1−litre flasks:

[i] A linear gradient of 30−120 mM sodium chloride for the purification of peptide α1−CB7.

[ii] A linear gradient of 35−120 mM sodium chloride for the purification of peptide α1−CB8.

Preparation of the sample and chromatography were carried out essentially as for the fractionations at pH 3.6 (see above) except for the difference in buffers used.

Fractions collected in all the fractionations on CM 52 were monitored at 234 nm against distilled water in a spectrophotometer (Unicam SP500) in order to determine the light absorbance (A_{234}). Using the resulting elution profiles, peak fractions were pooled accordingly, dialysed against four changes of 5−litre volumes of distilled water and freeze−dried.

2. Gel chromatography on Sephadex G−50_S

Tryptic and V8−protease digests of peptide α1−CB7 were initially fractionated on Sephadex G−50 (superfine grade) before being chromatographed on the high resolution AG50W−X4 cation−exchange column. Two columns were packed and used on separate occasions, one being 105 × 2.3 cm, the other, 128 × 1.4 cm. Chromatography was carried out once in 100mM acetic acid and thereafter in 30mM ammonium acetate pH 4.8 as described later in Chapter 4. All fractionations were carried out at 37°C.

3. High resolution ion−exchange chromatography on AG50W−X4 resin

Separation of tryptic and V8−protease peptides of peptide α1−CB7 was achieved by fractionation on a column of AG50W−X4 resin, 100 × 1 cm. The column was set up as follows (Schroeder, 1967):

Resin (125ml) was washed with 750ml of 2M sodium hydroxide in a sintered glass funnel over a Buchner flask without applying suction, to extract the maximum of fines and contaminants. Applying suction, the resin was then washed serially with 750ml distilled water, 750ml 3M HCl, 750ml distilled water and finally, with 500ml 2M pyridine. The resin was re−equilibrated with 1.5 litres of starting buffer, 0.2M pyridine acetate pH 3.1 (0.2M pyridine, pH adjusted with glacial acetic acid) and, as a thicker slurry, was poured into a jacketed glass tube (kept at 40°C) 105 × 1 cm. The resin was packed under gravity and then pumped at 18 ml per hour with starting buffer

until a constant bed height of 100 cm was reached and until the pH and conductivity of column effluent equalled that of starting buffer.

Freeze−dried tryptic and V8−protease peptides of peptide α1−CB7 were dissolved in distilled water (1.7 mg/ml), acidified to pH 2.5 with 6M HCl and loaded onto an almost drained column bed. Elutions were carried out at 40°C with sequential pyridine acetate buffers as described below:

[a] 70ml 0.2M pyridine acetate, pH 3.1 (buffer A).

[b] A linear gradient between 500ml of buffer A and 500 ml 1.1M pyridine acetate, pH 4.2 (buffer B).

[c] A linear gradient between 250 ml of buffer B and 250 ml 2M pyridine acetate, pH 5.0.

(Molarities of the buffers refer to the pyridine concentrations)

Elutions were performed at a flow−rate of 12 ml/hr, collecting 3ml fractions. The fractions were assayed by a 'ninhydrin' assay. As regards the first two fractionations of tryptic peptides on AG50W−X4, 600μl portions from test−tube fractions were evaporated to dryness in an oven at 110°C. To each of these, 600μl of ninhydrin solution (1:1:1 v/v/v distilled water, 10% aqueous pyridine and fresh 2% aqueous ninhydrin) was added. The mixtures were then placed in a boiling water bath for 20 minutes. After cooling, 500μl of 10% aqueous pyridine was added and mixed with the coloured reaction product. The light absorbance of the mixture was then determined at 570 nm (A_{570}) within an hour of removal from the water bath.

As regards the third fractionation of tryptic peptides and all the fractionations of V8−protease peptides on AG50W−X4 resin, the ninhydrin solution used for the 'ninhydrin' assay was the same as that used for amino acid analysis (see section IIG) which was merely more convenient to use. 80μl or 160μl portions from the 3ml fractions were dried down in small hydrolysis tubes. The sensitivity of the assay was increased by hydrolysing the dried portions in 50μl 6M HCl, carried out within the evacuated and flame−sealed hydrolysis tubes. After 18−hour hydrolysis at 110°C, the tubes were opened and the acid was evaporated off, also at 110°C. Ninhydrin solution

(600μl) was introduced and the reaction was carried out in a boiling water bath for twenty minutes. After cooling, 500μl distilled water was mixed into the coloured reaction product. The light absorbance of the mixture was then measured at 570nm in a spectrophotometer.

II E. SODIUM DODECYL SULPHATE – POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS–PAGE).

Cyanogen bromide peptides of corneal collagen were examined by SDS–PAGE on slab gels using a discontinuous buffer system (Laemmli and Favre, 1973: see Hames, 1981).

A 12% 'resolving–gel' solution (36ml) was prepared by mixing together 30% acrylamide/0.8% bis–acrylamide (14.4ml), 1.5M Tris pH 8.8 (9ml), 10% SDS (0.36ml) and distilled water (12.12ml). Polymerisation was set off by mixing in 12μl N N N'N'–tetramethylethylenediamine (TEMED) and 120μl of fresh 10% ammonium persulphate. The mixture was poured into a vertical cavity formed between two rectangular glass plates (one of which has a cut–out edge) separated by three well–greased plastic separators, one sealing the cavity at the bottom and two sealing the cavity at the sides. The solution was poured to within 3cm from the cut–out edge, overlayed with 0.1% SDS and allowed to set.

A 6% 'stacking–gel' solution (36ml) was prepared by mixing together, 30% acrylamide/0.8% bis–acrylamide (7.2ml), 0.5M Tris pH 6.8 (9ml), 10% SDS (0.36ml) and distilled water (19.3ml). About 6ml of this solution was required for each slab gel, the rest being stored at 4°C.

The top edge of the polymerised resolving gel (polymerisation takes about 10 minutes) was rinsed clean with distilled water and blotted dry. TEMED (5μl) and 10% ammonium persulphate (20μl) were mixed into the stacking–gel solution to set off polymerisation. The mixture was then overlayed onto the resolving–gel. A well–former was inserted into this overlay. When the gel had set, the well–former was carefully removed and the sample wells were rinsed and blotted dry. The bottom separator was

removed from between the glass plates and the gel was mounted, within the glass plates, onto a vertical slab gel unit (Shandon). The bottom reservoir was filled with running buffer (0.6% Tris, 2.88% glycine, 0.1% SDS) and any air bubbles were removed from the bottom edge of the gel using a pasteur pipette with a curved tip.

Samples were prepared by dissolving 1mg of cyanogen bromide peptide(s) in 100μl sample buffer (0.125M Tris pH 6.8, 2% SDS, 10% glycerol, 0.001% bromophenol blue) and heating this in a boiling water bath for 10 minutes. After cooling, 10μl of the sample was carefully loaded into a sample well using an automatic pipette ('Gilson'), increasing the amount of sample where more than one peptide band was expected at the end of electrophoresis. Any trace of sample was washed of the glass plates into the corresponding sample wells with running buffer from a pasteur pipette, eventually topping up the sample wells with running buffer without disturbing the samples beneath.

The top reservoir of the unit was filled with running buffer and the connecting hood was set in place. Electrophoresis was carried out at 15mA constant current (delivered from a power pack) until the bromophenol blue marker line enters the resolving-gel, after which the current was increased to 28mA. When the blue marker line had reached the bottom of the gel, the apparatus was switched off.

The gel was immediately removed from between the glass plates and placed in 'fixer' (7:50:50 v/v/v acetic acid:methanol:distilled water) for 45 minutes or overnight. The gel was then transferred into Coomassie Blue stain (2.5 g Coomassie Blue in 1 litre of 25% methanol / 10% acetic acid) and stained for 30-40 minutes. Destaining of the gel to reveal the peptide bands was achieved by placing the gel in 5% methanol / 7% acetic acid within a destaining tray (a plastic box with a lid) in a gently shaking water bath initially at 60°C for 3-4 hours and thereafter at 40°C until optimum destaining was achieved. Strips of polyurethane foam were placed in the destaining tray to soak up the stain and hence accelerate the destaining process.

The destained gels were then placed on a flat white surface and photographed from above using 'Ortho type 3' film (Kodak), and overhead lighting.

II F. DETERMINATION OF N-TERMINAL HYDROXYPROLINE OF PEPTIDE α1-CB7.

1. Dansylation (Woods and Wang, 1967; Gray, 1972)

Putative peptide α1-CB7 (400-500μg) was dissolved in 100μl of 1% SDS in a boiling water bath for 4 minutes. On cooling, 100μl of N-ethylmorpholine was added, followed by 150μl dansyl chloride solution (2.5mg per 100μl dimethyl formamide which had been kept dehydrated over calcium hydride). Dansylation was allowed to proceed for 1¼ hours in the dark after which the dansylated sample was precipitated by the addition of 1ml acetone. This mixture was centrifuged in a bench centrifuge, discarding the supernatant. The precipitate was then washed twice by repeatedly resuspending in acetone, centrifuging and pouring off the acetone. The residue was dried under a gentle stream of nitrogen, transferred into a small hydrolysis tube, and dissolved in 200μl 6M HCl. This tube was evacuated, flame-sealed and placed in an oven at 110°C. After 6 hours of hydrolysis, the hydrolysate was dried down by rotary-evaporation and dissolved in 20-30μl 50% aqueous pyridine. This was stored cool and in the dark until analysis by thin layer chromatography.

Glycine and 4-hydroxyproline (chromatographically homogeneous) standards were also dansylated for use as markers. They were dissolved separately in 0.5M sodium bicarbonate to give a final concentration of 1.2 mg/ml. To 100μl of this, 100μl of dansyl chloride solution (3.5 mgper 100μl acetone) was added. Dansylation was carried out in the dark for 2 hours. The reaction mixture was then extracted three times with 0.5ml water-saturated ethyl acetate, keeping the upper layer at each extraction. The upper layers were pooled and dried at 40°C under a gentle stream of nitrogen. Finally, the residue was dissolved in 20-30μl of 50% aqueous pyridine for chromatography.

2. Thin layer chromatography (Woods and Wang, 1967)

The dansylated sample hydrolysate was chromatographed by ascending flow chromatography on 4.5cm-square double-sided polyamide sheets. Sample (0.5μl of the

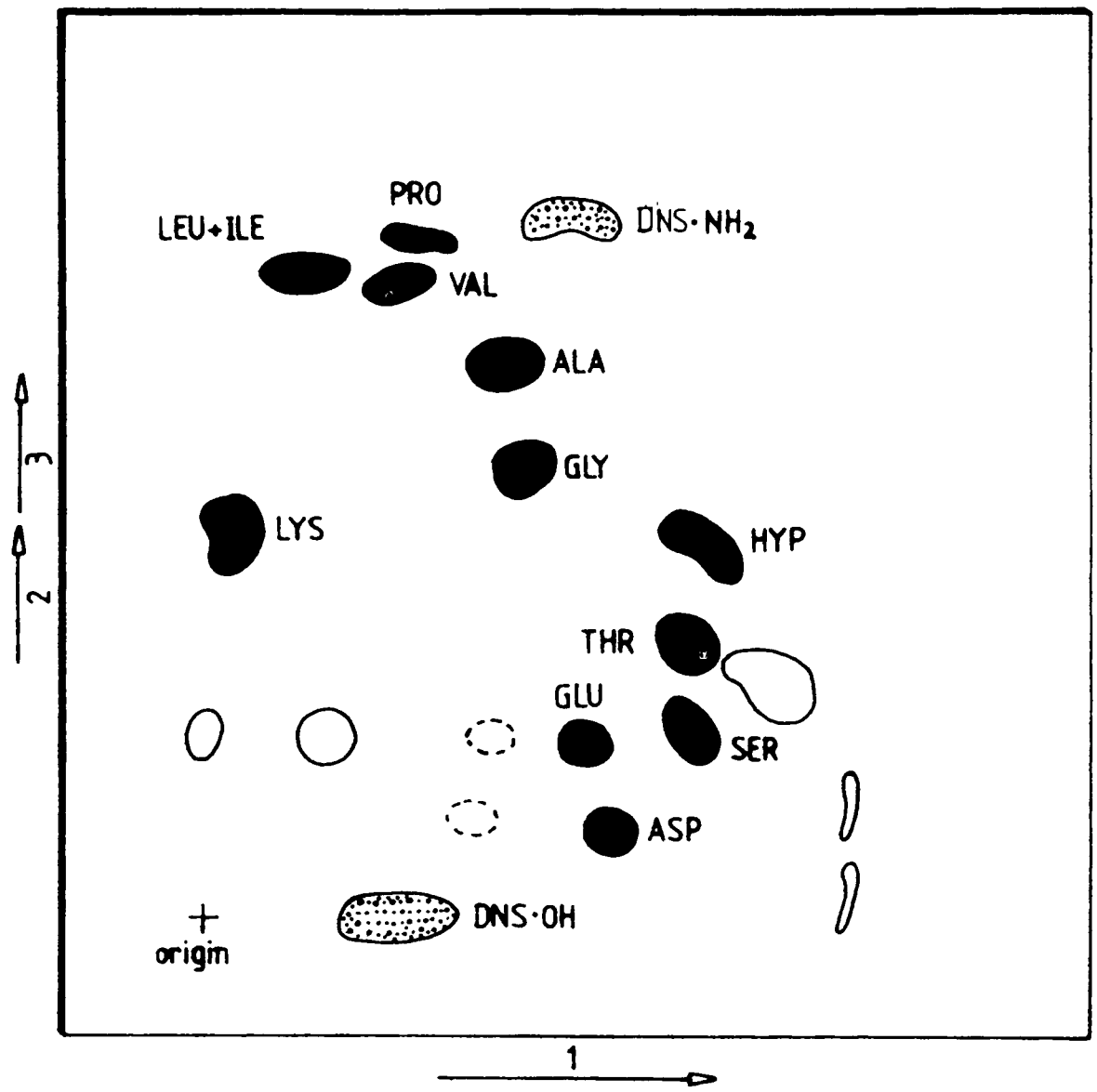


Figure 2.1 Separation of dansyl–amino acids by two–dimensional thin layer chromatography on polyamide sheets. Solvents: *1*, water:90% formic acid (200:3 v/v) in the first dimension and *2*, toluene:acetic acid (180:20 v/v) followed by *3*, ethyl acetate:methanol:acetic acid (20:1:1 v/v) in the second dimension. Spots not labelled were not identified. Faint spots are indicated by broken lines. *See Woods and Wang, 1967; Croft, 1980 for a comparison of dansyl amino acid spots.*

aqueous pyridine solution) was spotted and dried in a corner of a polyamide sheet, while on the exact reverse, dansyl−hydroxyproline, dansyl−glycine or both (0.5−1 μl of the aqueous pyridine solution) or a mixture of other dansyl−amino acids was likewise applied. Ascending−flow chromatography was then carried out in small rectangular jars, in two dimensions using the three following solvents (see fig. 2.1):

[a] First dimension: water/90% formic acid (200:3 v/v).

[b] Second dimension (solvent 1): toluene/acetic acid (180:20 v/v). (Croft, 1972)

[c] Second dimension (solvent 2): ethyl acetate/methanol/acetic acid (20:1:1 v/v).

The sheets were dried thoroughly for 10−15 minutes after each step with a hair dryer and were finally examined under an ultraviolet lamp (Hanovia Lamps, Slough) at 254nm to enable visualisation of dansyl−amino acids (fig. 2.1).

II G. AMINO ACID ANALYSIS

Amino acid analyses of collagen and collagen peptides were carried out on an LKB Biochrom 4400 automated amino acid analyser (LKB Instruments Ltd., Surrey). An attached Trilab Computing Integrator (Trivector Systems Ltd., England) was used for the integration of amino acid peak areas. The resin column used for the separation of amino acids was 24 × 0.45 cm. Buffer flow−rates used were about 26 ml/hour.

Collagen amino acid standards were analysed with every fresh batch of ninhydrin reagent to determine the colour factors used in the calculation of amino acid concentrations.

1. Ninhydrin reagent

'Acetate' buffer was first prepared, by dissolving into de−ionised distilled water, anhydrous potassium acetate (993.6g), sodium acetate.$3H_2O$ (459g), tri−potassium citrate (13.5g) and glacial acetic acid (337.5ml) to less than 3375ml. The pH was adjusted so that a 1:3 dilution (of a small test volume) with water gave a pH of 5.21 ± 0.02. The buffer was eventually made up to 3375ml with de−ionised distilled water and filtered through a 0.22 micron filter (Millipore Ltd., Harrow) into a Buchner flask

under suction.

2. Acid hydrolysis

Collagen and its peptides (250μg) were hydrolysed in 125μl 6M HCl (Aristar grade) in evacuated, flame-sealed hydrolysis tubes at 110°C for 18 hours. The hydrolysate was rotary-evaporated to dryness and dissolved in 270μl 0.2M (Na^+) sodium citrate buffer pH 2.2. After filtration through a 0.22 micron filter (Millipore), a portion of the sample (10μl to 80μl) was then applied to the amino acid analyser.

II H. QUALITATIVE AND QUANTITATIVE DETERMINATION OF HYDROXYLYSINE GLYCOSIDES

Hydroxylysine glycosides were determined by amino acid analysis. However, as glycosidic bonds are destroyed by acid hydrolysis, analyses were carried out on alkaline hydrolysates of collagen and its peptides (Spiro, 1967). Elution conditions were different to those used for amino acid analysis of acid hydrolysates.

1. Alkaline hydrolysis

Corneal collagen (5-20 mg/ml), collagen peptides (250μg-1mg per ml) and lens capsules (5-20 mg/ml) were hydrolysed in 2M NaOH (Aristar grade) at 110°C for 24 hours in tightly-capped polypropylene tubes, kept in a clamp. The hydrolysate, kept on ice, was acidified to below pH 2.5 (estimated by spotting onto pH paper) with 2M HCl (Aristar), filtered and analysed on the amino acid analyser without prior desalting.

2. Mild acid hydrolysis : elution positions of hydroxylysine glycosides

Some alkaline hydrolysates of lens capsule and corneal collagen were dried down by rotary-evaporation and re-hydrolysed in 100mM (H^+) sulphuric acid for 1½, 3, 5, 9 and 24 hours at 110°C in evacuated, flame-sealed glass tubes. The hydrolysates were then dried down by rotary-evaporation, dissolved in distilled water (270μl), filtered through a 0.22 micron filter and analysed on the amino acid analyser.

The gradual breakdown of glucosylgalactosylhydroxylysine (GGH) to

galactosylhydroxylysine (GH) and of GH to free 'unglycosylated' hydroxylysine (FH) with increasing acid hydrolysis result in corresponding changes in the sizes of the GGH, GH and FH peaks on amino acid analysis (Spiro, 1967; Odell et al., 1974), thus revealing the identities and positions of the respective peaks in the analysis chromatograms (see fig. 2.2).

3. Eluting conditions

In addition to hydroxylysine glycosides and amino acids, the alkaline hydrolysates prepared for analysis also contain amino acid breakdown products (Hill, 1965). It is necessary to use eluting conditions different to those for ordinary amino acid analysis in order to ensure the resolution of GGH, GH and FH from all the other amino acids and amino acid breakdown products (for access to the various conditions used by other authors, see Spiro, 1972; Odell et al., 1974; Panjwani and Harding, 1978a; Butler, 1982; Hennecke and Plapp, 1984; Tang and Williams, 1984).

In this work, a step−wise 3−buffer system, consisting of two eluting buffers and one equilibrating buffer was developed to facilitate quick and accurate determination of hydroxylysine glycosides on the automated amino acid analyser. The column used was the same as that used for the separation of amino acids in acid hydrolysates, as were the flow−rates of buffers and ninhydrin reagent.

Aliquots of the hydrolysates (5−100 μl) were applied to the amino acid analyser and the pre−equilibrated column was put through its elution cycle consisting of two elution steps, one 'wash' step and one equilibrating step:

Step 1: 26−minute elution with 0.1M (Na^+) sodium citrate buffer, pH 5.4
(obtained by diluting the equilibrating buffer 1:1 with water,
and adjusting the pH with NaOH).

Step 2: 26−minute elution with 1.2M (Na^+) sodium citrate buffer, pH 6.4.

Step 3: 3−minute 'wash' with 0.4M NaOH.

Step 4: 20−minute equilibration with 0.2M (Na^+) sodium citrate buffer, pH 3.2.
(After steps 1−4 the column is ready for the next sample)

Figure 2.2 **Elution positions of the diglycoside glucosylgalactosylhydroxylysine (GGH), the monoglycoside galactosylhydroxylysine (GH) and free, unglycosylated hyroxylysine (FH) in automated amino acid analysis.** The relevant parts from amino acid analysis chromatograms are shown with the elution position of phenylalanine (Phe) indicated for reference. Samples: *a*) alkaline hydrolysate of bovine lens capsule; *b*), *c*) and *d*) were 1½−, 5− and 24−hour mild acid hydrolysates* of sample a) respectively. *Hydrolysis in 100mM (H^+) H_2SO_4 at 110°C.

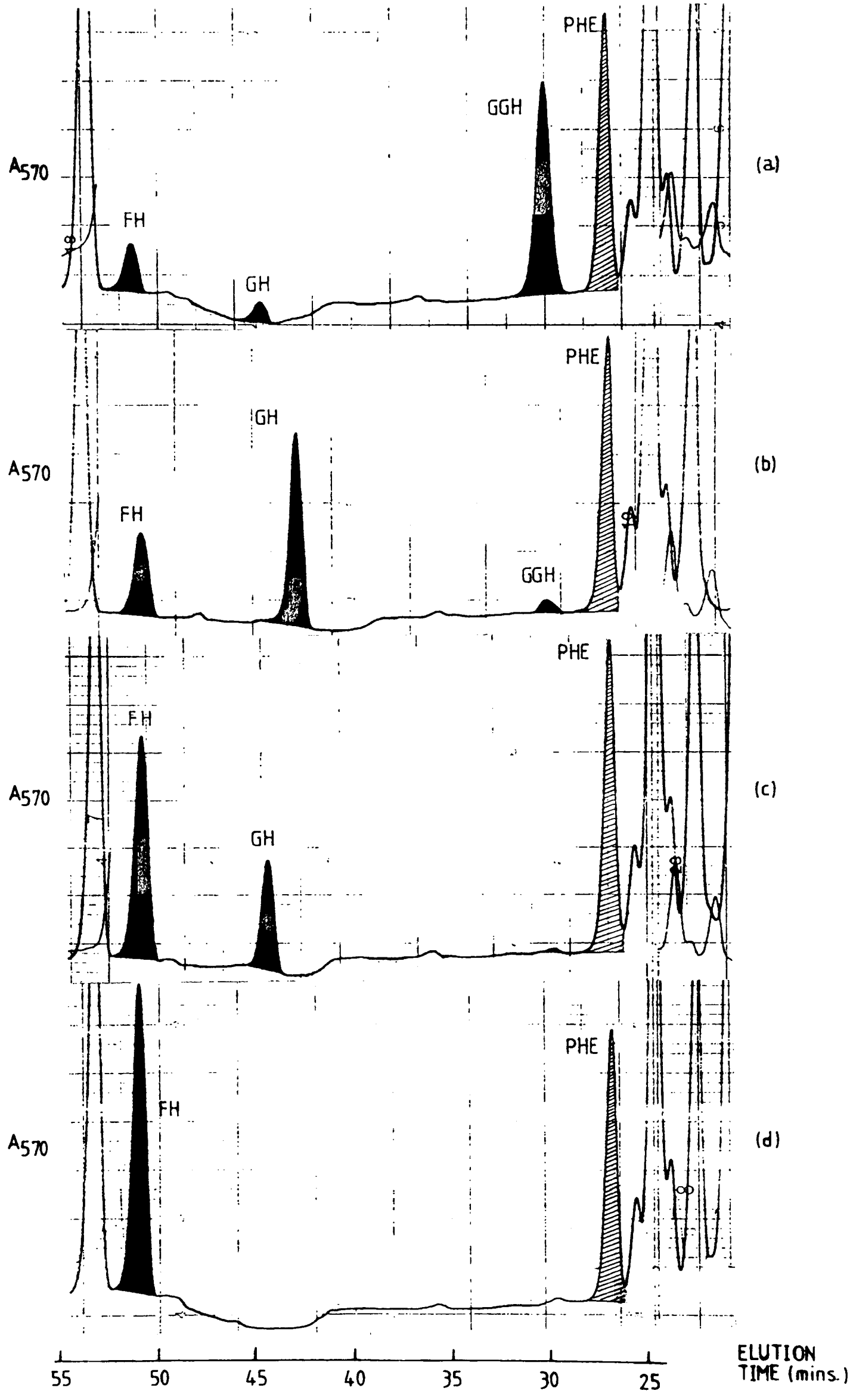

A570
FH
GH
GGH
PHE
(a)
A570
FH
GH
GGH
PHE
(b)
A570
FH
GH
PHE
(c)
A570
FH
PHE
(d)
55
50
45
40
35
30
25
ELUTION TIME (mins.)

Fig. 2·2

Using the described elution cycle, most of the amino acids eluted in the first part of the chromatogram followed by well-resolved peaks for phenylalanine at 27 minutes, GGH at 30 minutes, GH at 44 minutes and FH at 52 minutes (fig. 2.2).

The colour factors used to calculate the relative concentrations of the glycosides were 0.87 for GGH and 1.05 for GH relative to that of leucine (Odell et al., 1974). In this work, these factors were adapted to relate to that of phenylalanine which was known relative to that of leucine. The calculated relative proportions of GGH, GH and FH were correlated with the known concentration of 'total' hydroxylysine as determined by prior analysis of sample acid hydrolysates. If the sample also contained phenylalanine, the GGH, GH and FH concentrations were also calculated relative to phenylalanine. GGH and FH determined were not corrected for losses of about 15% incurred during hydrolysis (Spiro, 1967).

CHAPTER 3

ISOLATION AND CHARACTERISATION OF SOME CYANOGEN BROMIDE PEPTIDES OF BOVINE CORNEAL COLLAGEN

Cyanogen bromide specifically cleaves peptide bonds on the carboxyl side of methionyl residues of bovine type I collagen to yield thirteen peptides as indicated in figure 1.9. This chapter initially describes the purification and characterisation of insoluble bovine corneal collagen and subsequently, the characteristics of some of its cyanogen bromide (CNBr) peptides isolated in the present work.

I. PURIFICATION AND CHARACTERISATION OF CORNEAL COLLAGEN

I A. PURIFIED INSOLUBLE CORNEAL COLLAGEN

Insoluble corneal collagen was prepared as described in Chapter 2. Amino acid analysis of the insoluble corneal collagen gave an amino acid composition similar to those previously reported for collagens derived from cornea, skin and tendon (Table 3.1). Table 3.1 shows the high glycine, proline and hydroxyproline contents which are hallmarks of collagen. The absence of cystine from the collagen prepared for this work indicated the absence of collagen types III and IV. The collagen prepared in this work was thus considered to be reasonably pure for a study of its CNBr peptides.

Table 3.1 also shows the higher hydroxylysine content in corneal collagen (including that of the present work) when compared to skin and tendon collagens.

Hydroxylysine glycoside analysis of the purified collagen indicated the presence of 3.2 residues of glucosylgalactosylhydroxylysine (GGH), 2.0 residues of galactosylhydroxylysine (GH) and 4.0 residues of free, 'unglycosylated' hydroxylysine (FH) per 1000 amino acid residues (Table 3.2). The sum of these values, at 9.2 residues is close to the value of 9.6 residues 'total' hydroxylysine per 1000 amino acid residues found in the acid hydrolysate of purified corneal collagen (Table 3.1).

Table 3.1

Amino acid compositions of bovine corneal collagen prepared in this work and those previously reported for bovine corneal, skin and tendon collagens. Values are expressed as residues per 1000 amino acid residues.

	Present work		*Panjwani & Harding (1978a)*	*Volpin & Veis (1973)*	*Harding & Wesley (1968)*
	Cornea	corneal CNBr digest*	Cornea	Skin	Tendon
Hydroxyproline	93.5	92.6	82	94	98.3
Aspartic acid	50.5	49.9	48	46	49.9
Threonine	17.6	16.7	18	17	17.3
Serine	34.9	33.8	33	36	33.9
Glutamic acid	75.3	77.6	77	72	75.8
Proline	122	125	112	127	127.2
Glycine	329	331	334	326	337.3
Alanine	105	101	115	112	106.6
Valine	22.7	22.3	22	23	18.9
Methionine	5.5	2.8†	4.4	6.0	4.7
Isoleucine	11.5	12.1	12	11	11.8
Leucine	28.5	27	30	25	26.7
Tyrosine	3.3	3.2	5.0	3	4.1
Phenylalanine	13.7	14.4	14	13	13
Hydroxylysine	9.6	9.7	8.3	6	7.6
Histidine	4.4	3.8	4.4	5	5.3
Lysine	24.8	25.7	29	27	17.2
Arginine	48.2	51.4	52	51	44.4

**soluble in 20mM sodium citrate, pH 3.6*
†Homoserine: not corrected for losses due to homoserine lactone

Table 3.2

Hydroxylysine glycoside and hydroxylysine contents of purified bovine corneal collagen and of its CNBr digest.

	Present work		*Panjwani & Harding (1978b)*
	purified collagen	CNBr digest*	purified collagen
GGH	3.2	3.4	3.1
GH	2.0	2.2	1.9
FH	4.0	3.7	2.9
Total 'total' Hyl†	9.2	9.3	7.9
	9.6	9.7	8.3

GGH, *glucosylgalactosylhydroxylysine*
GH, *galactosylhydroxylysine*
FH, *free (unglycosylated) hydroxylysine*

**soluble in 20mM sodium citrate, pH 3.6*
†*found in the acid hydrolysate (see Table 3.1)*

I B. CLEAVAGE WITH CYANOGEN BROMIDE

Cleavage with cyanogen bromide at 40°C (Panjwani and Harding, 1978a) solubilised most of the collagen, with little residue left in the reaction flask. About 80–85% of the freeze–dried CNBr digest was soluble in 20mM sodium citrate buffer pH 3.6 which was the starting buffer for the subsequent ion–exchange fractionation on carboxymethyl cellulose. Methionine was not found on amino acid analysis of the citrate–soluble (Table 3.1) and citrate–insoluble fractions, indicating complete cleavage of the corneal collagen by CNBr. The 'expected' recovery of homoserine (CNBr digestion converts methionine to homoserine but some of this is not detected due to formation of the 'lactone') in the amino acid compositions accounts for approximately all the pre–existing methionyl residues.

The study of corneal collagen peptides in the rest of this work was carried out on citrate–soluble CNBr digests.

Hydroxylysine glycoside analysis of citrate–soluble CNBr digests showed the presence of 3.4 GGH, 2.2 GH and 3.7 FH residues per 1000 amino acid residues (Table 3.2). The sum of these values, at 9.3 residues, is close to the value of 9.7 'total' hydroxylysine residues per 1000 amino acid residues found in the acid hydrolysates of citrate–soluble CNBr digests.

Panjwani and Harding (1978a) showed that electrophoresis of CNBr digests of bovine corneal collagen by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) on 'rod' gels gave the characteristic pattern obtained for type I collagen. In this work, use of the 'slab' gel method as described by Laemmli and Favre (1973) facilitated better resolution of peptide bands and enabled the comparison of the more detailed gel staining–pattern with those reported for CNBr digests of other collagens. The staining patterns obtained in the present work (fig. 3.1) are similar to those obtained for CNBr digests of type I collagen by other workers (see figure 3.14 for example) and supports the view that bovine corneal (stromal) collagen is mainly type I collagen. No obvious 'marker' peptide bands for collagen types II, III, IV, V (see Bornstein and Sage, 1980) or VI (Zimmerman et al., 1986) were seen.

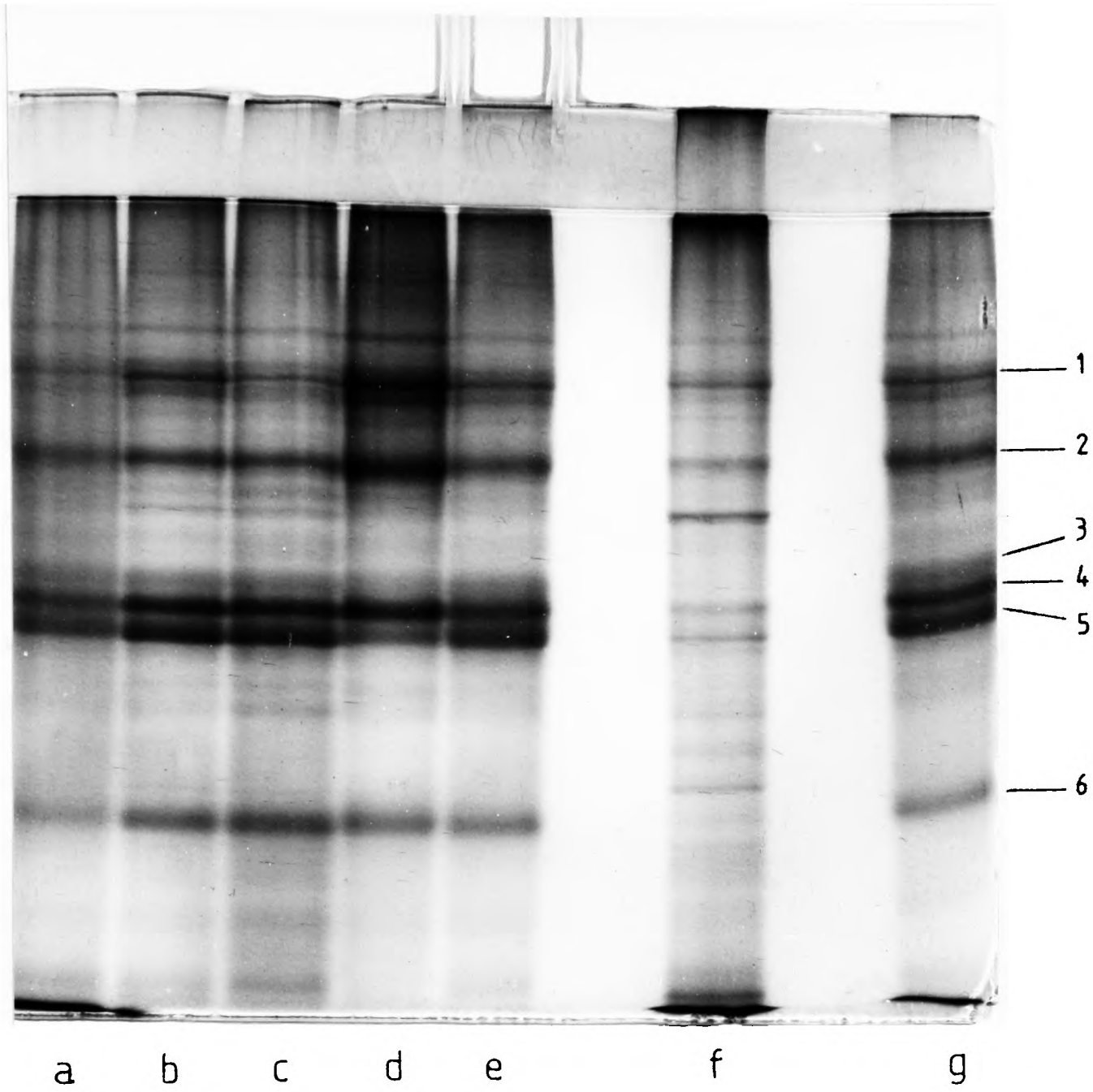


Figure 3.1 Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS−PAGE) of cyanogen bromide (CNBr) digests of bovine corneal collagen. Tracks ***a−e*** : five typical CNBr digests prepared and used in the present work. Track ***f*** : the portion of CNBr digest which was insoluble in 20mM sodium citrate buffer, pH 3.6. Track ***g*** : the portion of CNBr digest which was soluble in the citrate buffer. Identities of the peptide bands : *1*, α2−CB3,5; *2*, α1−CB6 dimer; *3*, α2−CB4; *4*, α1−CB7; *5*, α1−CB8; *6*, α1−CB3 (assignments according to *Laurent et al., 1981; Sage and Bornstein, 1979* except for α1−CB6 dimer which is according to *Harding and Crabbe, 1979*).

II. CYANOGEN BROMIDE PEPTIDES OF BOVINE CORNEAL COLLAGEN

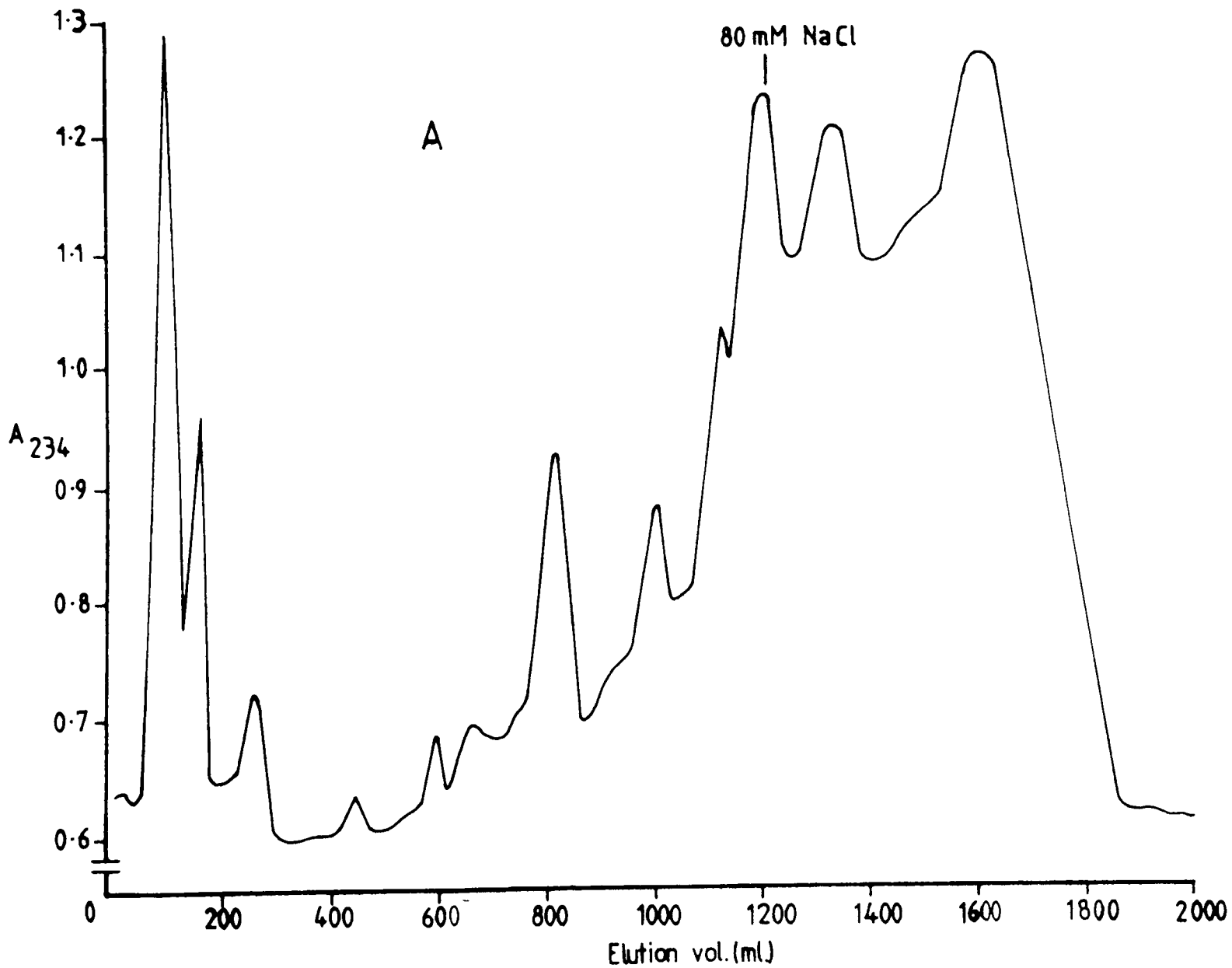
II A. INITIAL FRACTIONATION OF CYANOGEN BROMIDE PEPTIDES

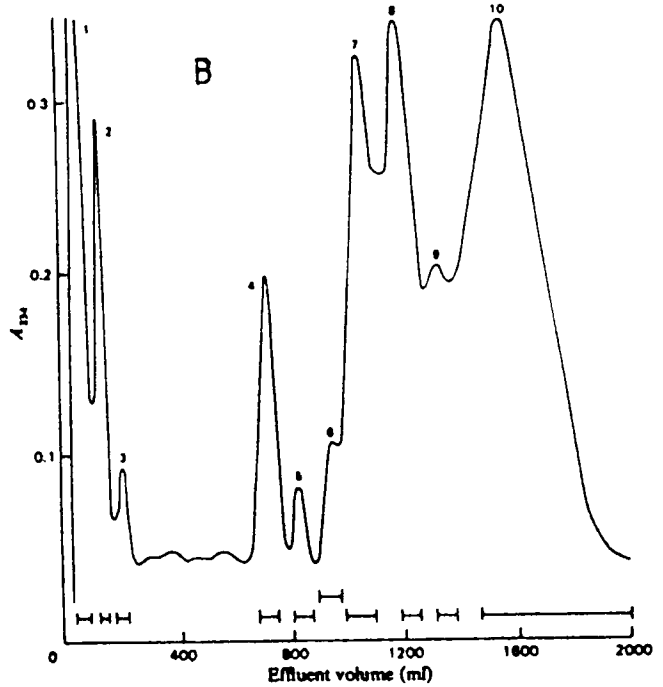
Early on during this project, fractionations of 800mg batches of corneal collagen CNBr digests were carried out on a carboxymethyl cellulose (CM 52) column 40 × 1.9 cm, utilising a linear 0−140mM sodium chloride gradient over 2 litres of 20mM sodium citrate buffer, pH 3.6. A typical chromatogram is shown in figure 3.2(A). The elution profile was similar to those reported previously for bovine corneal collagen CNBr digests (Panjwani and Harding, 1978a) and bovine skin collagen digests (Volpin and Veis, 1973). This supports the view that collagen of the corneal stroma is mainly type I collagen.

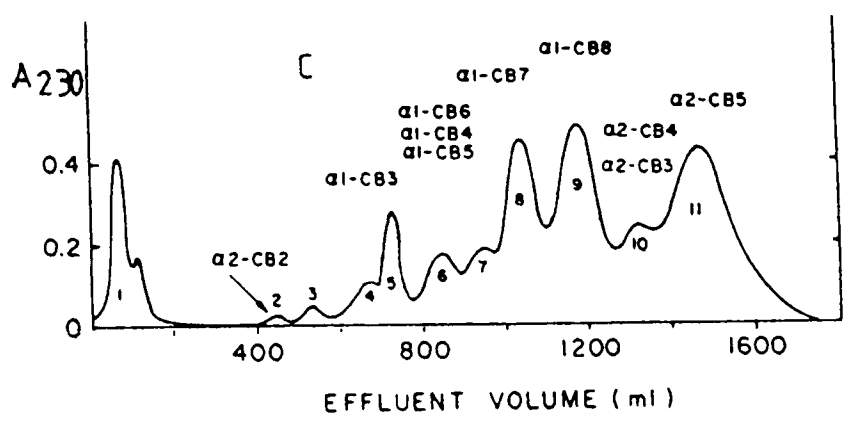
It soon became apparent that it was necessary to carry out this initial fractionation step on a larger scale for the purpose of the first stage of the project, which was the isolation and characterisation of CNBr peptide α1−CB7. The less than optimal resolution of peak 7, eluting at a sodium chloride concentration of about 80 mM, from the adjoining peaks (fig. 3.2) meant that pooling of fractions was necessarily stringent in order to minimise the inclusion of other CNBr peptides in the 'crude' preparation of peptide α1−CB7 recovered from peak 7. Consequently, an inconveniently large number of these 'pH 3.6' fractionations would have had to be performed in order to yield enough crude peptide α1−CB7 as starting material for this project.

Fractionations of CNBr digests on a larger scale (1.6−2.4g per fractionation) were carried out on a longer CM 52 column, 140 × 1.9 cm, utilising a shallower linear gradient of sodium chloride (40−140 mM) over a volume of 4 litres buffer (see Chapter 2.IID). Four such fractionations were carried out. A typical chromatogram is shown in figure 3.3. The peak fractions were pooled as shown, dialysed and freeze−dried. Portions of the freeze−dried products were examined by SDS−PAGE. The products recovered from peaks 1 and 2 (fig. 3.3) were expected to contain CNBr

Figure 3.2 CM−cellulose chromatograms for the fractionation of CNBr peptides of type I collagen at pH 3.6. Elutions were performed at 40°C with a linear gradient of 0−140mM NaCl in a total volume of 2 litres 20mM sodium citrate pH 3.6. *A.* Chromatography of bovine corneal collagen CNBr digest (800mg) in the present work. Flow−rate was 100ml/hr and tube−fractions of 7.2ml were collected. The column was 40 × 1.9 cm. *B.* Similar chromatography of bovine corneal collagen CNBr digest reported by *Panjwani and Harding (1978a).* *C.* Similar chromatography of CNBr digest of bovine insoluble skin reported by *Volpin and Veis (1973).*







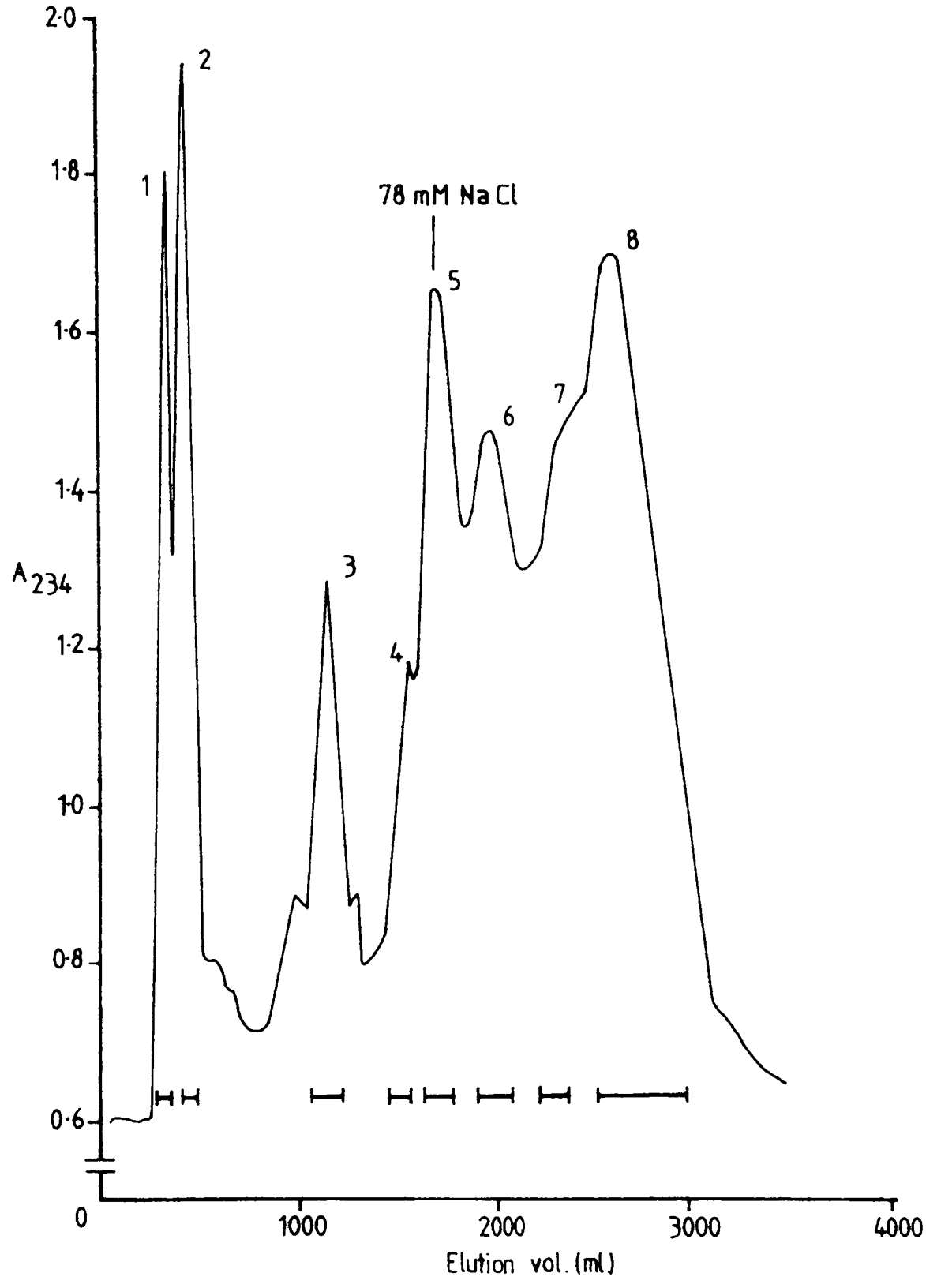


Figure 3.3 Chromatography of a CNBr digest of bovine corneal collagen (2.4g). on CM−cellulose . Column: 140 × 1.9 cm. NaCl gradient: 40−140mM over 4 litres of 20mM sodium citrate pH 3.6. Fraction size: 10.2ml. Other conditions were as indicated in fig. 3.2. Bars indicate the fractions pooled.

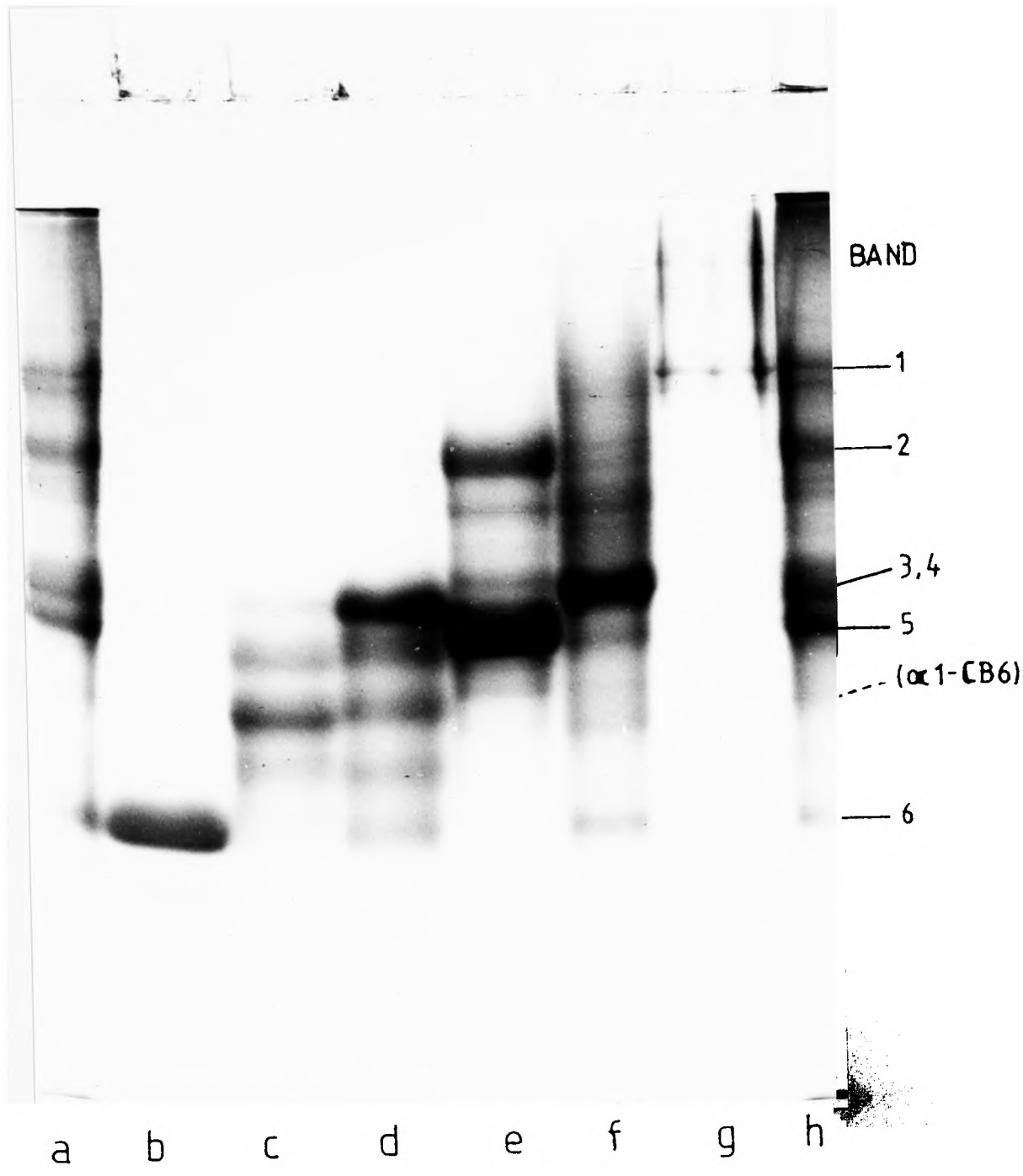


Figure 3.4 SDS−PAGE of peptide components from the CM−cellulose fractionation of a CNBr digest of bovine corneal collagen shown in fig. 3.3. Tracks *a*, *h* : total CNBr digest; *b*: peak 3; *c*, peak 4; *d*, peak 5; *e*, peak 6; *f*, peak 7 and *g*, peak 8. Assignment for peptide bands are as indicated for fig. 3.1.

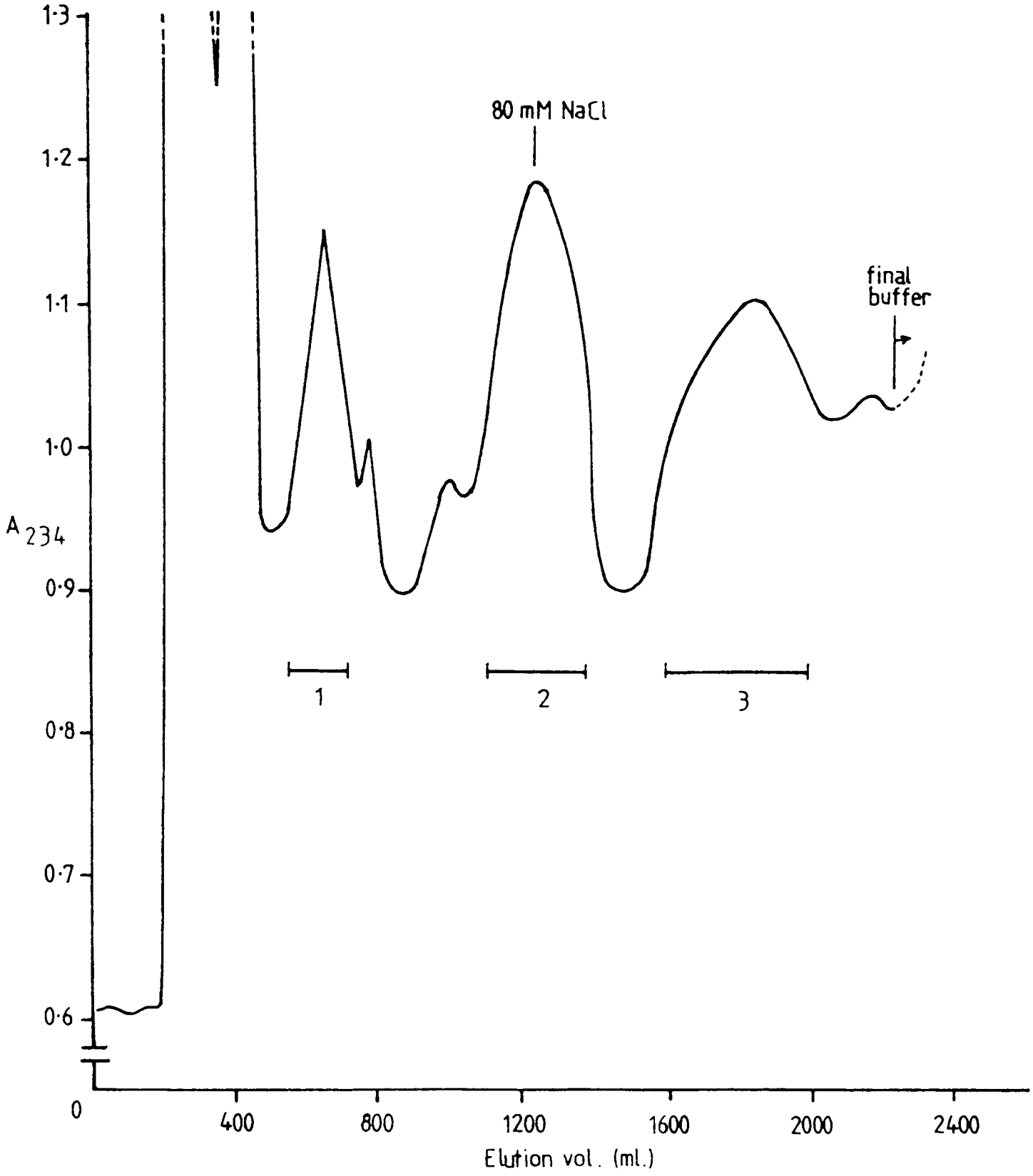


Figure 3.5 CM−cellulose chromatogram of a CNBr digest (1.6g). The elution profile shows the peptide peaks eluted utilising a linear gradient of 60−100mM NaCl over 2 litres of 20mM sodium citrate pH 3.6. Peptides remaining in the column after this mode of elution were washed through with a 'final' buffer containing 200mM NaCl. Fraction volume: 8.3ml. Other conditions were as indicated in fig. 3.3. Bars indicate the fractions pooled.

peptides smaller than peptide α1−CB3, and were not detected on SDS−gels perhaps because they were not retained in the gels due to their small sizes. The SDS−PAGE pattern corresponding to the peptides contained in peaks 3 to 8 of figure 3.3 is shown in figure 3.4.

The staining pattern of the slab gel shown in figure 3.4 indicated the elution of CNBr peptides α1−CB3, α1−CB6, α1−CB7, α1−CB8, α2−CB4 and α2−CB3,5 in peaks 3, 4, 5, 6, 7, and 8 respectively of the 'pH 3.6' fractionation shown in figure 3.3. None of these pooled fractions were shown to be homogeneous by SDS−PAGE except perhaps the fraction corresponding to peptide α1−CB3 (see fig. 3.4). The assignment of particular collagen CNBr−peptide identities to peaks eluted from 'pH 3.6' fractionations as practised by some authors (see fig. 3.2) thus merely indicate the major components in what are essentially crude preparations. A CNBr digest of type I collagen thus consists of a complex mixture of peptides which include cross−linked peptides such as the dimer of peptide α1−CB6 (Harding and Crabbe, 1979).

Eventually, the 'pH 3.6' fractionations of CNBr digests were carried out with an even shallower linear gradient of sodium chloride (60−100mM), over a volume of 2 litres buffer. This resulted in better resolution of the 'crude α1−CB7' fraction from the other fractions (fig. 3.5 shows a typical chromatogram) and also shortened chromatography time. As a result of this more selective elution however, peptides α2−CB4 and α2−CB3,5 which require sodium chloride concentrations greater than '100mM' to elute off the column, were still retained in the column at the end of the gradient elution. These were 'washed' off the column with a 'final buffer' containing 200mM sodium chloride, followed by 10mM sodium hydroxide. This selective fractionation of CNBr digests was performed three times. Examination of fractions 1, 2, and 3 (fig. 3.5) by SDS−PAGE (not shown) showed them to be enriched in peptides α1−CB3, α1−CB7 and α1−CB8 respectively, yielding SDS−PAGE patterns similar to tracks b, d and e respectively of figure 3.4. Although these fractions were as crude as those obtained earlier, they were recovered in much greater quantities as a result of the less stringent pooling of fractions which were more well−resolved from

each other.

After examination by SDS−PAGE, similar crude fractionations of particular CNBr peptides of corneal collagen were pooled together and stored for further purification. In the present work, priority was given to the eventual purification of peptides α1−CB7 and α1−CB8. Crude preparations of these particular peptides, obtained from these 'pH 3.6' fractionations gave the SDS−PAGE patterns shown in figure 3.6. The major contaminant in these crude preparations was always the α1−CB6 'dimer'.

II B. PURIFICATION OF PEPTIDE α1−CB7

Crude preparations of peptide α1−CB7 were pooled together and fractionated on a shorter CM 52 column (40 × 1.9 cm), this time, in 30mM sodium acetate buffer, pH 4.8, utilising a linear gradient of 30−120mM sodium chloride over a volume of 1200ml. This was performed twice, yielding enough of pure peptide α1−CB7 for the work presented in Chapter 4. The chromatograms for these two fractionations are shown in figures 3.7 and 3.8. Peptide fractions recovered from these fractionations were examined by SDS−PAGE (figs. 3.9, 3.10) which indicated the homogeneity of the main α1−CB7 fractions (peak 6 in fig. 3.7 and peak 4 in fig. 3.8).

In addition, the fractions following these main peaks in figures 3.7 and 3.8 were also shown on the SDS−gels to contain mainly peptide α1−CB7. Peptide α1−CB7 is thought to elute as two separate peaks in these fractionations at pH 4.8 due to the partial occurrence of its carboxy−terminal homoserine residue as homoserine−lactone at pH 4.8, resulting in two populations of peptide α1−CB7 with different electrostatic charge (Rauterberg and Kühn, 1971).

The amino acid sequence of peptide α1−CB7 from calf skin (Fietzek et al., 1973) indicates 4−hydroxyproline as being the N−terminal amino acid residue. Of the other CNBr peptides of type I collagen, most of them contain N−terminal glycine and none contain N−terminal hydroxyproline. Portions of both fractions 6 and 4 from the fractionations shown in figures 3.7 and 3.8 respectively, were dansylated for N−terminal determination (Chapter 2.IIF). After thin layer chromatography of

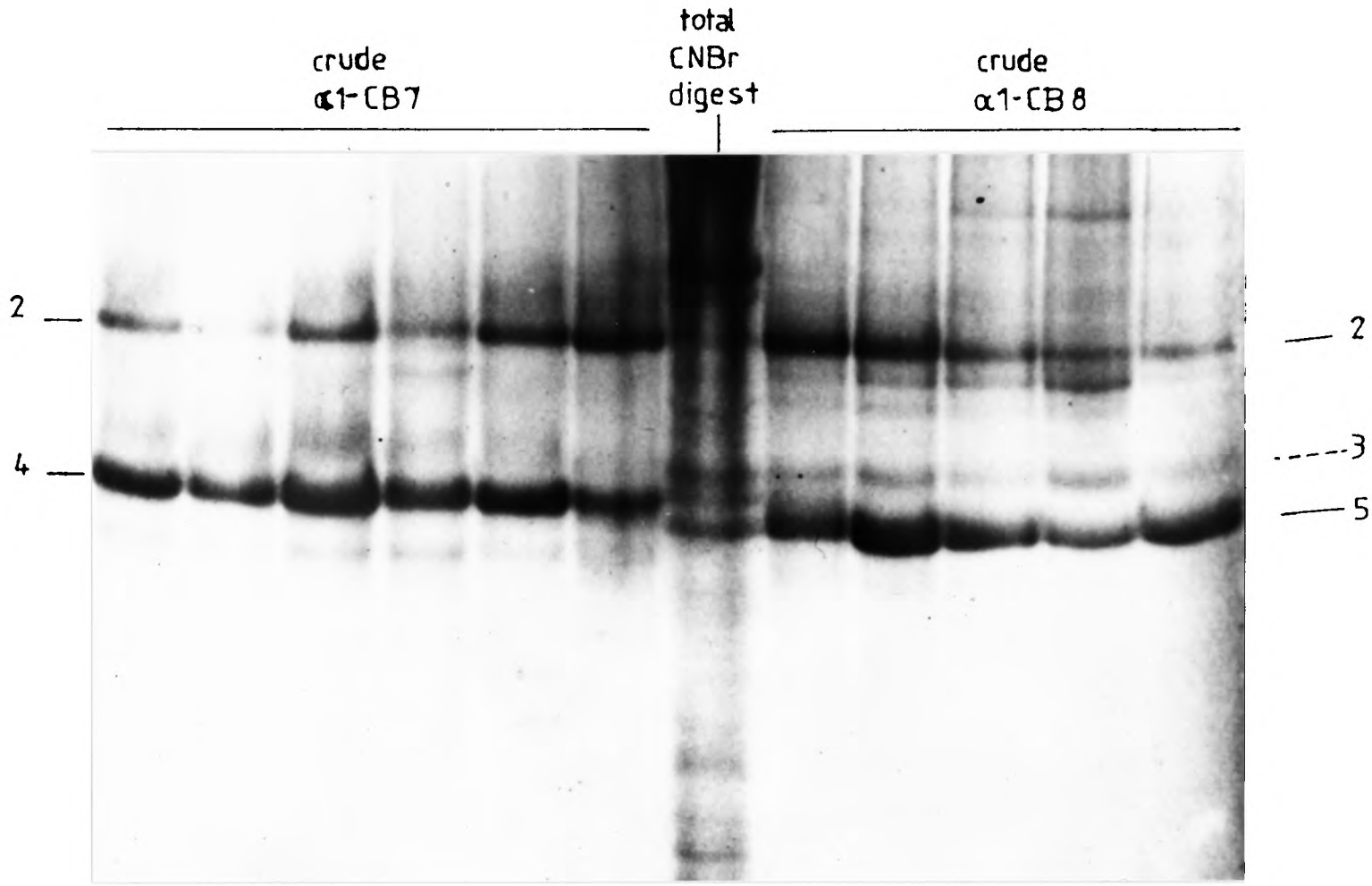


Figure 3.6 SDS−PAGE of some of the 'crude' peptide α1−CB7 and 'crude' peptide α1−CB8 fractions recovered from CM−cellulose ion−exchange chromatography of bovine corneal collagen CNBr digests at pH 3.6. Peptide bands are assigned as indicated in fig. 3.1.

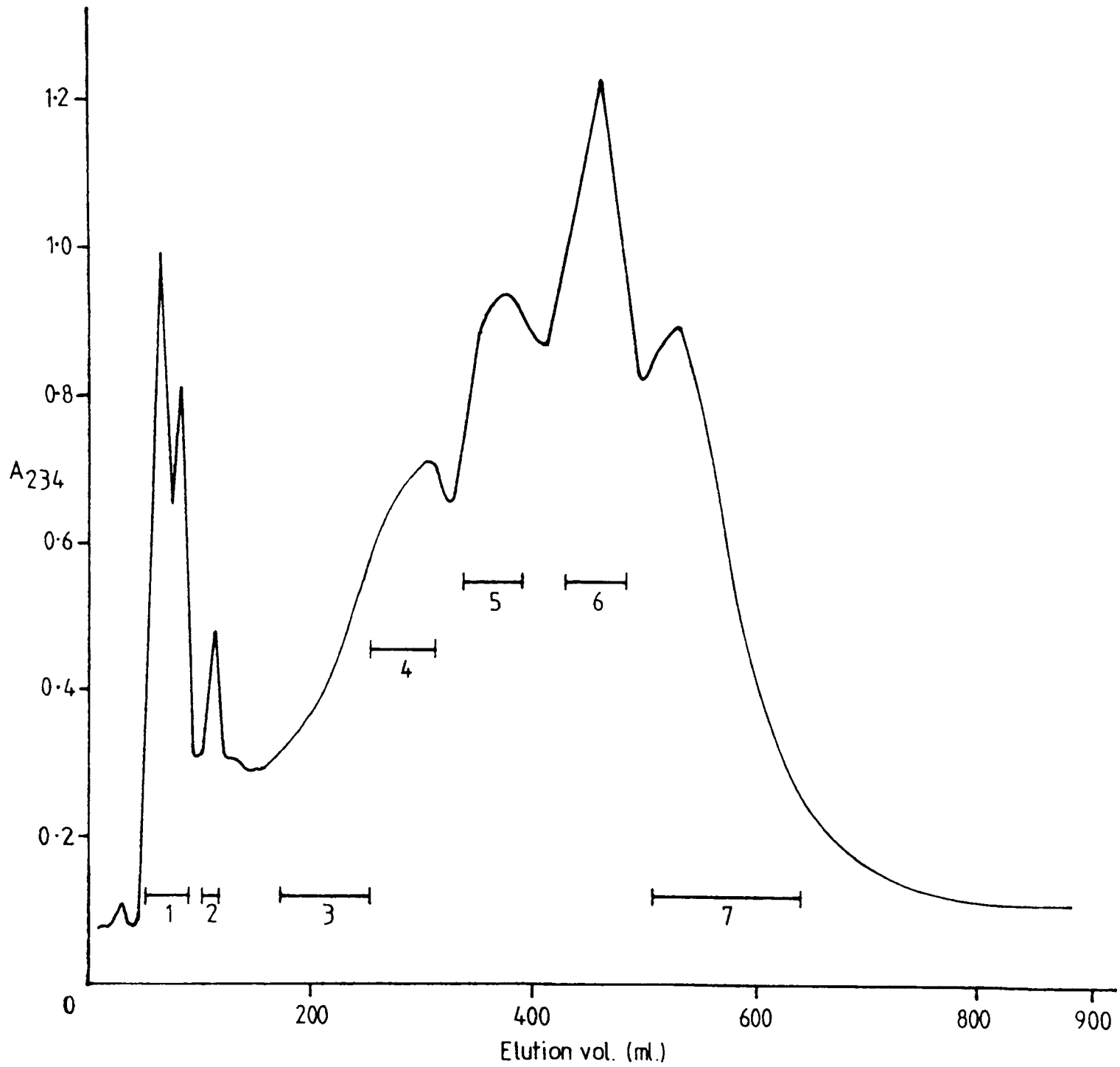


Figure 3.7 Purification of peptide α1−CB7 by CM−cellulose chromatography of 'crude' preparations of peptide α1−CB7 (440mg) in 30mM sodium acetate pH 4.8. Column: 40 × 1.9 cm. NaCl gradient: 30−120mM over 1200ml buffer. Flow−rate: 100ml/hr. Fractions: 7.9ml. Column temperature: 40°C. Bars indicate the fractions pooled.

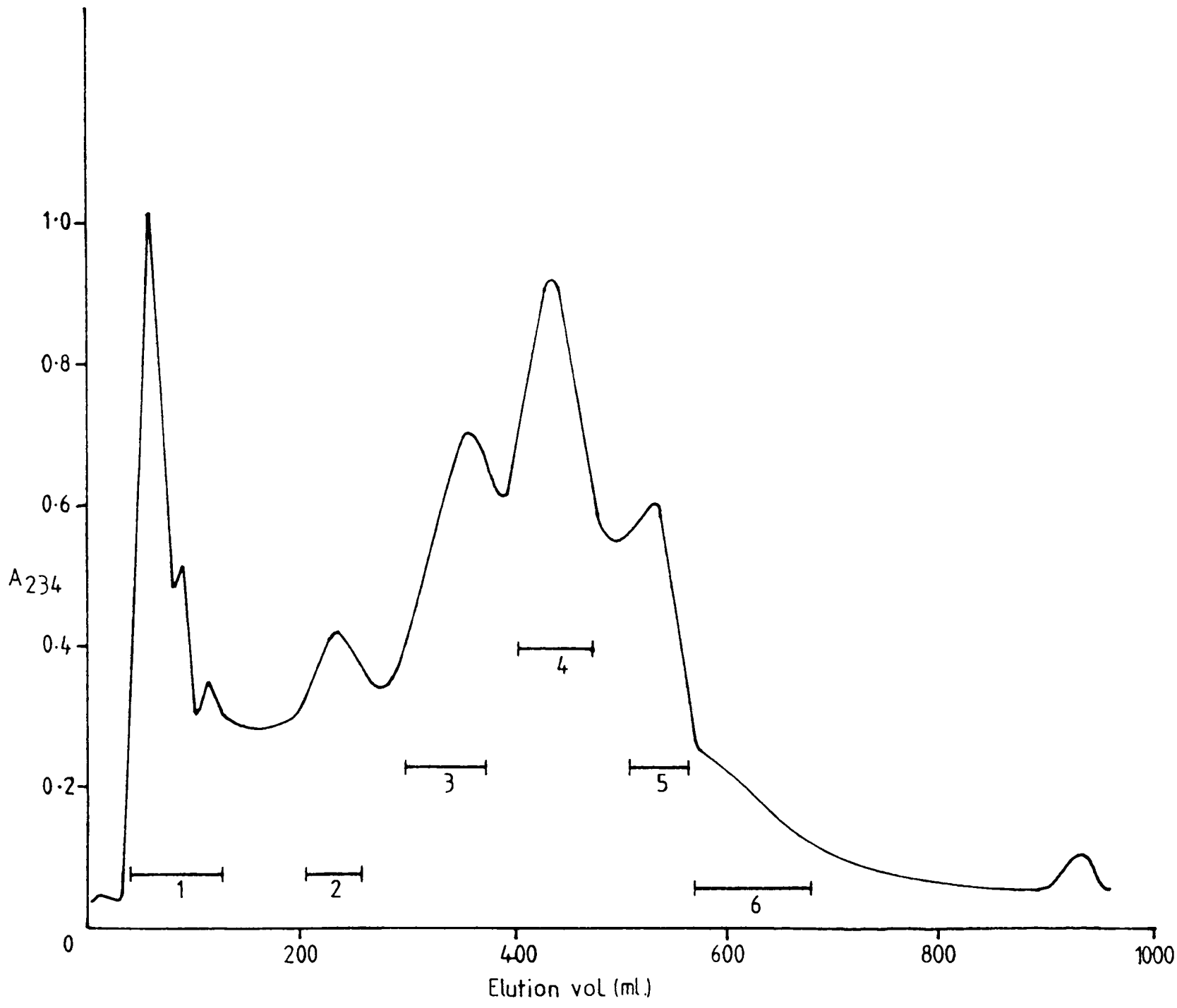


Figure 3.8 Purification of a second batch of peptide α1−CB7 from crude preparations (330mg). Chromatography was performed as described in fig. 3.7. Bars indicate the fractions pooled.

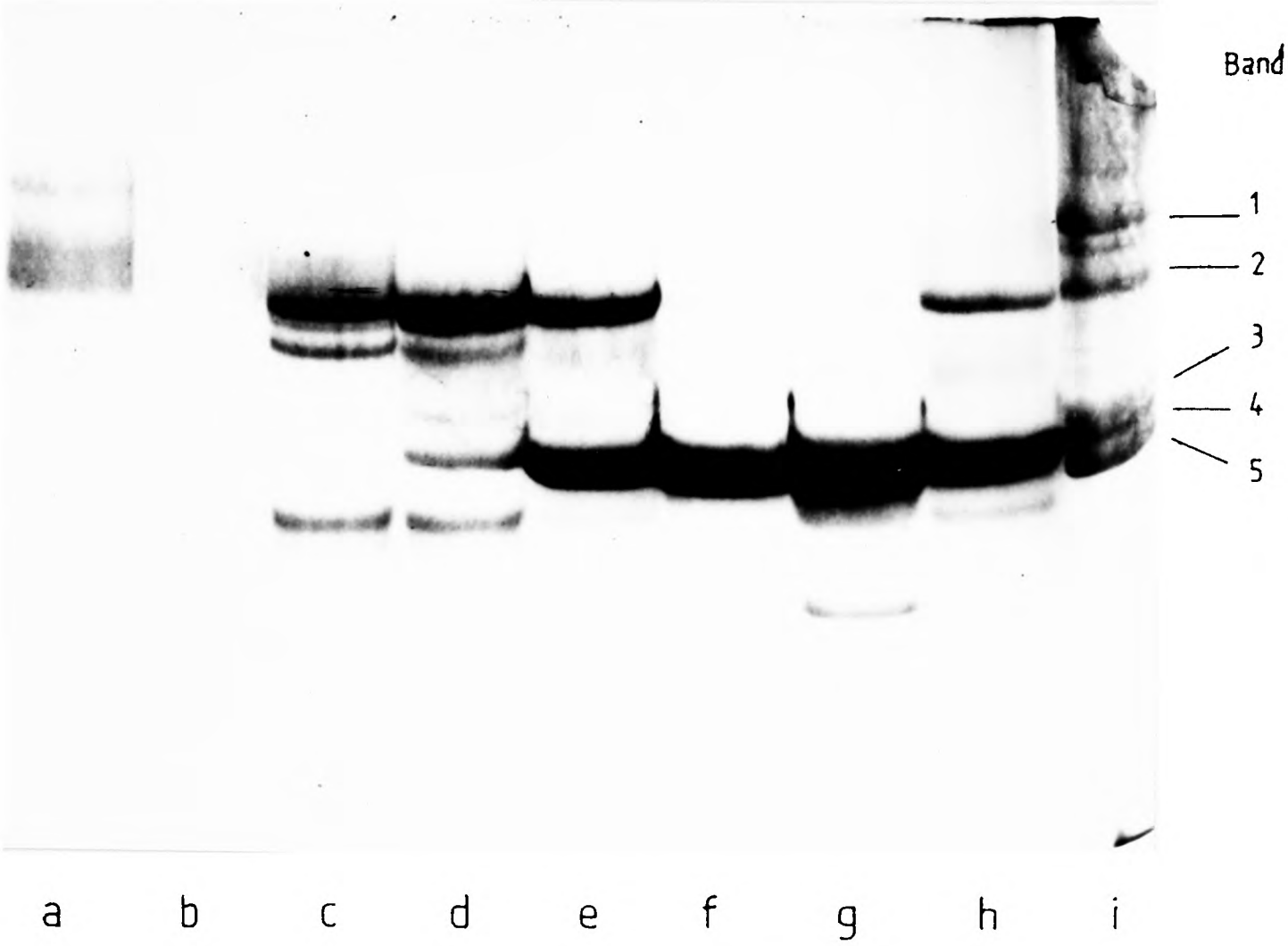


Figure 3.9 SDS−PAGE of fractions obtained from the fractionation of 'crude' peptide α1−CB7 (fig. 3.7). Tracks *a*−*g* correspond to fractions *1*−*7* respectively while track *h* shows the pattern for a crude preparation of peptide α1−CB7 from the 'pH 3.6' fractionations. Track *i* represents total CNBr digest. Peptide−band assignments are as indicated in fig. 3.1.

Figure 3.10 SDS−PAGE of fractions obtained from the fractionation of a second batch of 'crude' peptide α1−CB7 (fig. 3.8). Tracks *a*−*f* correspond to fractions *1*−*6* respectively. Track *g* represents total CNBr digest and track *h*, a 'crude' preparation of peptide α1−CB7. Peptide−band assignments are as indicated in fig. 3.1.

Table 3.3

Amino acid composition of peptide α1-CB7 from bovine corneal collagen and that calculated from the known sequence of peptide α1-CB7 from calf skin. Values are expressed as residues per peptide.

	Present work	
	bovine cornea*	calf skin†
Hydroxyproline	28.0	28
Aspartic acid	12.2	12
Threonine	5.3	5
Serine	7.0	7
Glutamic acid	16.1	16
Proline	36.1	36
Glycine	89.8	90
Alanine	36.0	36
Valine	5.8	6
Isoleucine	2.8	3
Leucine	4.1	4
Phenylalanine	3.1	3
Hydroxylysine	1.8	1.0‡
Lysine	9.6	9.7‡
Arginine	12.9	13
Homoserine	0.4§	1
total	271.0	270.7

mean of two preparations; †from Kleinman et al. (1978) except ‡ from Rexrodt et al. (1973); §not including that lost due to homoserine lactone

*samples: see track 'f' (Fig 3.9) and track 'd' (Fig 3.10)

hydrolysates of these dansylated fractions, examination of the polyamide sheets under ultra−violet light (254nm) showed single fluorescent spots for these samples which coincided only with the 4−hydroxyproline standard on the exact reverse of the sheets.

Amino acid analysis of the two putative 'α1−CB7' fractions gave very similar amino acid compositions. The mean composition was close to that obtained for peptide α1−CB7 from calf skin collagen (Rexrodt et al., 1973; Kleinman et al., 1978) except that the corneal peptide contains 1.8 hydroxylysine residues and 9.6 lysine residues per peptide (Table 3.3) while the calf skin equivalent contains 1.0 hydroxylysine and 9.7 lysine residues per peptide. My findings are in agreement with the data presented by Panjwani and Harding (1978a), which indicated the presence of 1.5 hydroxylysine and 8.6 lysine residues per peptide α1−CB7 of bovine corneal collagen.

The elution positions with regard to fractionation on CM 52, the peptide band position on SDS−gels, the identity of the N−terminal amino acid plus the amino acid composition data together confirm the identity of the purified peptide as pure peptide α1−CB7 of type I collagen. This confirmation also shows that the higher level of lysyl hydroxylation found in corneal collagen and corneal peptide α1−CB7 relative to lysyl hydroxylation in the calf skin equivalents is not the result of contamination with collagen types II, IV and V which are known to have high hydroxylysine contents (see Miller and Gay, 1982).

The purity of the fractions as indicated by SDS−PAGE, N−terminal determination and amino acid analysis also meant that fractions 6 and 4 from the fractionations shown in figures 3.7 and 3.8 respectively were of the desired purity for further, more detailed characterisation of peptide α1−CB7 as described in the next chapter.

The 271−residue peptide α1−CB7 represents about 26% of the 1055−residue long α1 chain of type I collagen.

II C. PURIFICATION OF PEPTIDE α1−CB8

Crude preparations of peptide α1−CB8 from the 'pH 3.6' fractionations were

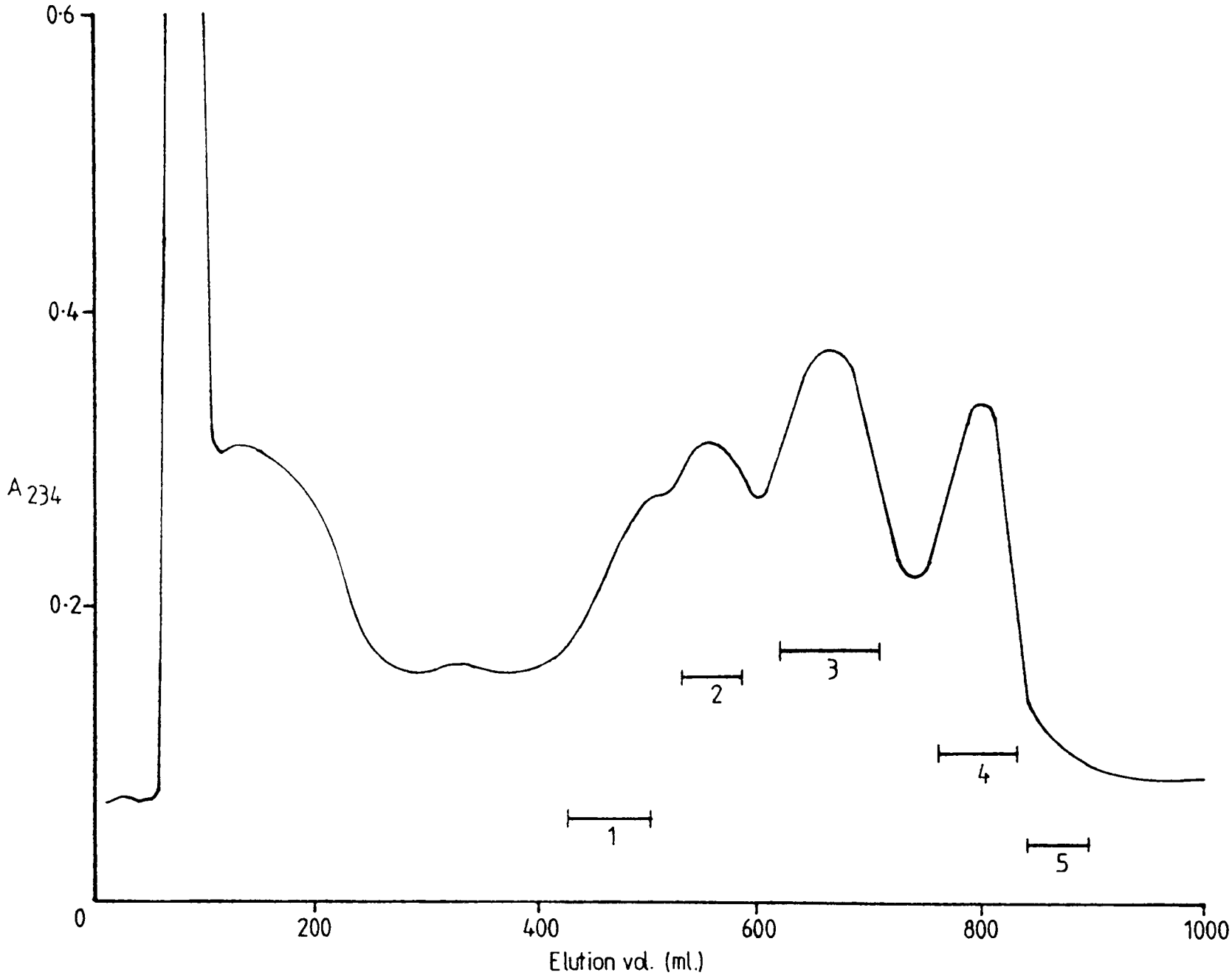


Figure 3.11 Purification of peptide α1−CB8 by CM−cellulose chromatography of crude preparations of peptide α1−CB8 (300mg). Chromatography was carried out as for peptide α1−CB7 except that the eluting gradient was 35−120mM NaCl.

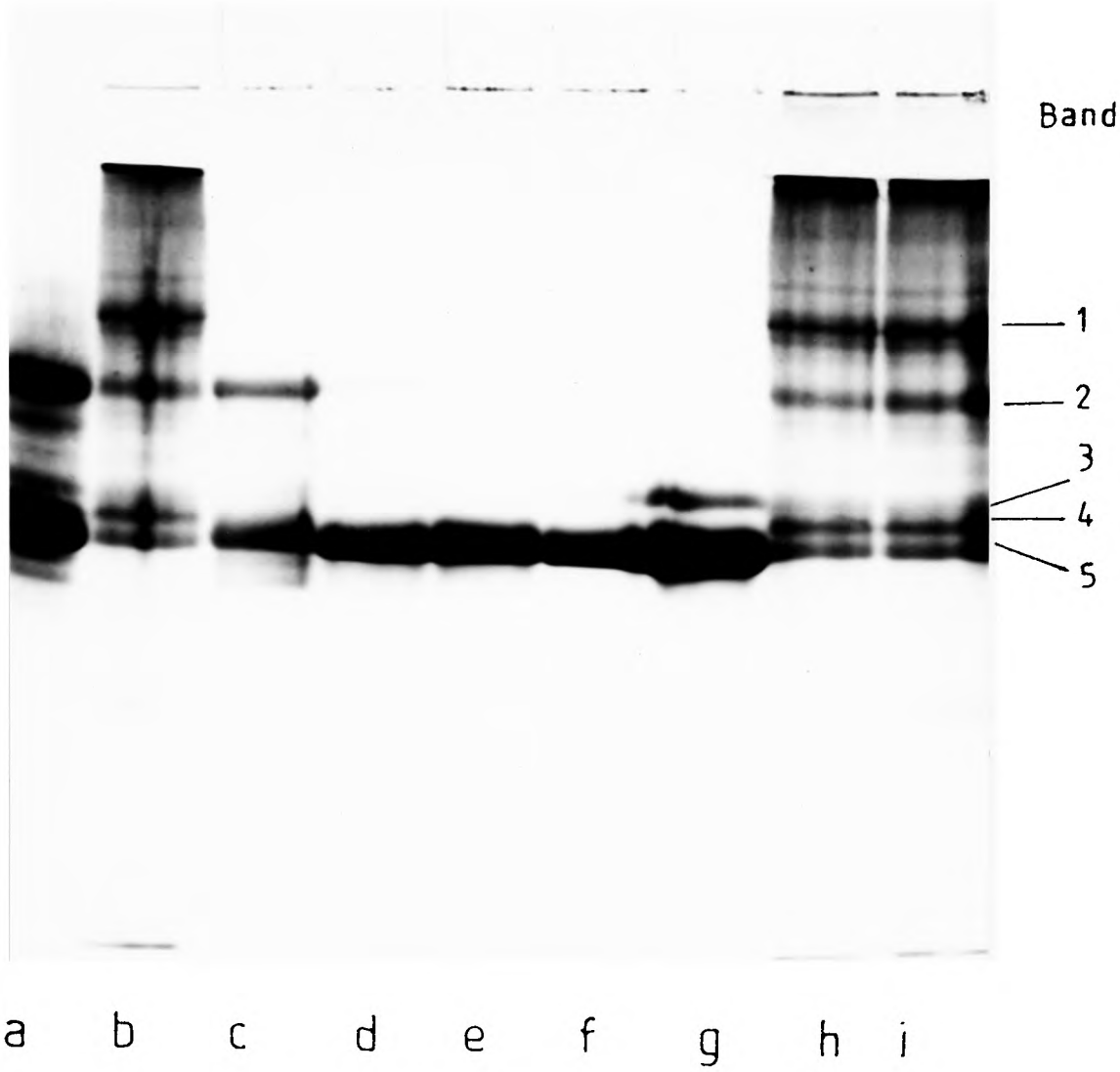


Figure 3.12 SDS−PAGE of fractions obtained from the fractionation of crude αl−CB8 fractions at pH 4.8 (fig. 3.11). Tracks *c*, *d*, *e*, *f* and *g* correspond to fractions *1*, *2*, *3*, *4* and *5* respectively. Track *a* represents a crude αl−CB8 preparation. Tracks *b*, *h* and *i* represent total CNBr digests of bovine corneal collagen. Peptide−band assignments are as indicated in fig. 3.1.

Table 3.4

Amino acid composition of peptide α1-CB8 from bovine corneal collagen and that calculated from the known sequence of peptide α1-CB8 from calf skin. Values are expressed as residues per peptide.

	Present work	
	bovine cornea	calf skin*
Hydroxyproline	31.0	33
Aspartic acid	10.6	11
Threonine	5.4	6
Serine	8.3	8
Glutamic acid	20.7	20
Proline	30.6	30
Glycine	95.0	96
Alanine	38.6	37
Valine	4.5	4
Isoleucine	1.5	2
Leucine	4.5	4
Phenylalanine	2.8	3
Hydroxylysine	2.5	0.9†
Lysine	8.2	9.1†
Arginine	14.3	15
Homoserine	0.5§	1
total	279.0	279.0

**from the sequence except †from the amino acid composition (Glanville et al. (1983)*

§not including that lost due to homoserine lactone

pooled together and re−fractionated on CM−52 at pH 4.8 as just described for crude peptide α1−CB7 except that the elution was achieved with a linear gradient of 35−120mM sodium chloride. The chromatogram is shown in figure 3.11. Examination of the five pooled fractions by SDS−PAGE showed peak fraction 3 to be a homogeneous preparation of peptide α1−CB8 (fig. 3.12). This peptide was also the major components of peaks on either side of peak 3. The elution of peptide α1−CB8 in more than two peaks at pH 4.8 cannot therefore be attributed solely to the formation of homoserine−lactone as mentioned previously with respect to peptide α1−CB7. With respect to peptide α1−CB8 derived from chick skin, Highberger et al. (1982) also reported multiple forms during fractionation at pH 4.8 and suggested the possibility of minor sequence differences as a possible explanation although this seems unlikely. They also suggested the possibility of contamination with type III collagen peptides. Although type III collagen does feature significantly in skin, it represents less than 1% of bovine corneal collagen (Lee and Davison, 1984) and is thus unlikely to be the source of the multiple forms of 'α1−CB8' found during this project.

Amino acid analysis of fraction 3 (fig. 3.11) gave an amino acid composition (Table 3.4) close to that obtained for peptide α1−CB8 of calf skin collagen except that the corneal peptide contained 2.5 hydroxylysine and 8.2 lysine residues per peptide while the calf skin peptide contains 0.9 hydroxylysine and 9.1 lysine residues per peptide (Glanville et al., 1983). These values for peptide α1−CB8 of bovine corneal collagen are the first reported for this CNBr peptide of corneal tissue. They are consistent with the observation that corneal type I collagen has a higher hydroxylysine content than skin type I collagen.

The 279−residue peptide α1−CB8 is similar in length to peptide α1−CB7 and also represents about 26% of the α1(I) chain.

II D. HYDROXYLYSINE GLYCOSIDE CONTENTS OF PEPTIDES α1−CB7, α1−CB8 and α1−CB3.

Hydroxylysine glycoside analysis was carried out on purified peptides α1−CB7 and

α1−CB8 as well as a preparation of peptide α1−CB3 which gave a single stained band when examined by SDS−PAGE (see track b, fig. 3.4).

An acid hydrolysate of the putative peptide α1−CB3 was analysed, giving the amino acid composition shown in Table 3.5. The agreement with the compositions reported by other workers (Table 3.5) indicated that the single−step fractionation of the CNBr digest at pH 3.6 (fig. 3.3) yielded a fraction corresponding to peptide α1−CB3 of reasonable purity. The slight discrepancies between some of the values e.g. those for aspartate, threonine and leucine were probably due to slight contamination of my peptide α1−CB3 preparation with peptides smaller than peptide α1−CB3, which, due to their size, were not detected on SDS−gels. They are possibly peptides α1−CB4 and α1−CB5 which are indicated as eluting soon after peptide α1−CB3 in 'pH 3.6' fractionations (see fig. 3.2). Due to the suspected contamination, the preparation of peptide α1−CB3 was considered to be 'crude' although reasonably pure for gross characterisation. The 149−residue peptide α1−CB3 (Fietzek et al., 1972) represents about 14% of the 1055−residue long α1(I) chain.

Hydroxylysine glycoside analysis of peptides α1−CB7, α1−CB8 and α1−CB3 (crude) gave the results summarised in Table 3.6.

Peptide α1−CB7 contained 0.40 residues of glucosylgalactosylhydroxylysine (GGH), 0.57 residues of galactosylhydroxylysine (GH) and 0.78 residues of free hydroxylysine (FH) per peptide. The sum of these values, at 1.75 residues, is close to the 1.8 residues of 'total' hydroxylysine per peptide as deduced from analysis of the acid hydrolysate (see Table 3.3). These values are similar to those reported by Panjwani and Harding (1978a) for this peptide which were 0.52 GGH, 0.61 GH and 0.45 FH residues per peptide except for the difference in FH content.

It is difficult to guess exactly where the 0.40 GGH, 0.57 GH and 0.78 FH residues reside in peptide α1−CB7 as eight out of the eleven lysyl residues of this peptide occur in 'pre−Gly' positions (see Chapter 4), signifying eight possible sites. The characterisation of the extent of hydroxylation and glycosylation at all the eleven lysyl positions of peptide α1−CB7 from bovine corneal collagen was the ultimate aim

Table 3.5

Amino acid composition of partially purified peptide α1−CB3 from bovine corneal collagen, that calculated from the known sequence of peptide α1−CB3 from calf skin and that previously determined for bovine corneal peptide α1−CB3. Values are expressed as residues per peptide.

	*Present work**	*Panjwani & Harding (1978a)*	*Fietzek et al. (1972)*
	bovine cornea		calf skin†
Hydroxyproline	14.1	13	17
Aspartic acid	7.1	6.3	6
Threonine	0.5	0.3	0
Serine	3.7	3.0	3
Glutamic acid	14.5	16.1	16
Proline	16.2	15.8	13
Glycine	50.0	51.7	50
Alanine	20.1	22	22
Valine	3.8	3.6	4
Isoleucine	0.4	0.4	0
Leucine	3.7	3.1	3
Phenylalanine	3.1	2.8	3
Hydroxylysine	1.2	0.9	0.3‡
Lysine	4.2	4.2	4.8‡
Arginine	6.2	6.2	6
Homoserine	0.3§	0.4§	1
total	149.1	150.0	149.1

*only partially purified
§not including homoserine lactone
†composition calculated from the sequence except for ‡ from the composition of bovine skin α1-CB3 (Volpin and Veis, 1973)

Table 3.6

Hydroxylysine glycoside and hydroxylysine contents of peptides α1−CB7, α1−CB8 and partially purified peptide α1−CB3 of bovine corneal collagen as determined in the present work. Values are expressed as residues per peptide.

	α1-CB7	α1-CB8	α1-CB3
GGH	0.40	0.25	0.58
GH	0.57	0.96	0.42
FH	0.78	1.20	0.36
total	1.75	2.41	1.36
'total' hydroxylysine*	1.8	2.5	1.2

GGH, *glucosylgalactosylhydroxylysine*; GH, *galactosylhydroxylysine*
FH, *'free' (unglycosylated) hydroxylysine*
**hydroxylysine found in the acid hydrolysate*

of this project. This is described in the next chapter.

The hydroxylysine glycoside values obtained for peptide α1−CB8 represent the first reported for this peptide in bovine corneal collagen. Peptide α1−CB8 was found to contain 0.25 GGH, 0.96 GH and 1.2 FH residues per peptide (Table 3.6). The sum of these values, at 2.41 residues, is close to the value of 2.5 'total' hydroxylysine residues per peptide as deduced from analysis of the acid hydrolysate (see Table 3.4). As mentioned for peptide α1−CB7, it is also difficult to guess exactly where the hydroxylysine glycosides reside in peptide α1−CB8 as the equivalent calf skin sequence shows that eight out of the ten lysyl residues of this peptide occur in 'pre−Gly' positions, signifying eight possible targets for hydroxylation and subsequent glycosylation. However the sequence studies of Glanville et al. (1983) indicated that lysyl positions 174, 219, and 342 in peptide α1−CB8 of calf skin collagen are 40%, 30% and 20% hydroxylated. Perhaps these lysyl positions are the major glycoside sites in peptide α1−CB8 of bovine corneal collagen.

Peptide α1−CB3 (crude) contained 0.58 GGH, 0.42 GH and 0.36 FH residues per peptide (Table 3.6). The sum of these values, at 1.36 residues, is close to the 1.2 'total' hydroxylysine residues per peptide (Table 3.5). These values are close to those of 0.54 GGH, 0.36 GH and 0.32 FH residues per peptide reported by Panjwani and Harding (1978b) whose detailed characterisation of peptide α1−CB3 of corneal collagen also pin−pointed the glycoside sites, at positions 408 and 531, the only two 'pre−Gly' lysyl positions.

III. RELATIVE ELECTROPHORETIC MOBILITIES OF PEPTIDES α1−CB7 AND α1−CB8 ON SDS GELS.

Before going on to the next chapter, I would like to clarify a small but important matter encountered in the course of the present work. This concerns the electrophoretic mobilities of the CNBr peptides α1−CB7 and α1−CB8 on SDS−gels.

From the time that CNBr peptides were first examined by SDS−PAGE to the recent present, there has been a certain inconsistency in published reports with regard

to the relative electrophoretic mobilities of these peptides. Let us retrace in chronological order, the development of SDS−PAGE as a method for the study of type I collagen CNBr peptides.

In 1966, the first cleavage of collagen (of rat skin) with CNBr was reported (Bornstein and Piez, 1966). In the years that followed, intensive research was carried out on CNBr peptides of the α1(I) chain from rat, chick and calf skin (see Butler, W.T. et al., 1967; Kang et al., 1969; Rauterberg and Kühn, 1971). The CNBr peptides derived from the α1(I) chain were named in the order in which they eluted during fractionations on carboxymethyl cellulose at pH 3.4−3.8. By a combination of biochemical analyses and electron microscopy, the peptides were ordered along the length of the α1(I) chain. Their molecular weights were determined by sedimentation equilibrium, gel chromatography, and amino acid analysis. On average, peptide α1−CB8 was estimated to be about five residues or 500 daltons larger than peptide α1−CB7.

In 1971, SDS−PAGE, as a method for molecular weight estimation, was applied to collagen α chains and their CNBr peptides for the first time (Furthmayr and Timpl, 1971). The use of a continuous buffer system and 6.5cm−long 'rod' gels did not result in the separation of peptides α2−CB4 α1−CB7 and α1−CB8 from one another. These and the other peptide bands were assigned their identities according to their expected relative mobilities as deduced from their accepted molecular weights. For example, a peptide band assigned as the uncleaved double−peptide "α1−CB3,7" was not identified as peptide α1−CB3,7 by other means such as amino acid analysis of the purified peptide. The authors' main conclusion was that SDS−PAGE was a feasible method for molecular weight estimations of collagen α chains and their CNBr peptides except that their electrophoretic mobilities are about 67% of those expected for globular proteins of similar molecular weights. This, they thought was due to the high imino acid content of collagen (proline and hydroxyproline represent about 20% of the amino acid composition of collagen). The imino acid residues are thought to confer a degree of rigidity to peptide sequences, even in the presence of SDS (sodium dodecyl sulphate).

Thus the authors suggested that the α1(I) chain had a lower electrophoretic mobility relative to the α2(I) chain because of its slightly higher imino acid content even though the molecular weights of the two chains were roughly equal (Furthmayr and Timpl, 1971).

Between 1970 and 1975, intensive amino acid sequencing studies yielded the amino acid sequences of all the CNBr peptides, albeit from more than one animal species, comprising the entire α1 chain and partial sequences of the α2 chain of type I collagen (see Fietzek and Kühn, 1976). Peptide α1−CB8 was shown to be 279 residues long while peptide α1−CB7 was shown to be 268 residues long. Peptide α2−CB4 was also determined to be 321 residues long. A subsequent discovery of three previously 'missed' amino acids led to an amendment of the length of peptide α1−CB7 to 271 residues (Kleinman et al., 1978; Highberger et al., 1978).

In 1976, Scott and Veis (1976a, 1976b) characterised the CNBr peptides of acid insoluble bovine type I collagens on SDS gels by the method of Furthmayr and Timpl (1971) except that longer rod gels (11.5cm) were used. From optical density scans of the peptide banding patterns, the authors calculated, by planimetry, the relative proportions of particular CNBr peptides within their samples. From their calculations, the authors deduced the extent of CNBr cleavage at particular methionyl bonds and the identities of putative cross−linked peptides. The longer rod gels used, at times facilitated a slight resolution of the " α2−CB4, α1−CB7, α1−CB8 " group of peptides. The authors assigned identities to the peptide bands, indicated as peaks in the optical density scans, from the knowledge of molecular weights and amino acid sequences and indicated that peptide α1−CB7 had a higher electrophoretic mobility than peptide α1−CB8 (see fig. 3.13[a]). The identities of the assigned peptide bands were not verified by other methods.

In 1977, using a flat−bed (slab) SDS gel method described by Sykes and Bailey (1971), Herbage et al. (1977) obtained better resolution of CNBr peptides of type I collagen. Their assignment of peptide bands indicated that peptide α1−CB8 had a higher electrophoretic mobility than peptide α1−CB7, despite the available knowledge

that the former had the higher molecular weight of the two and despite the inconsistency with the previous assignments by Scott and Veis (1976b). Burke et al. (1977) agreed with the assignments of Herbage et al. (1977) with respect to peptides α1−CB7 and α1−CB8.

From 1978 to 1985, researchers working on CNBr peptides of type I collagen continued presenting in their published work, optical density scans or photographs of rod or slab gels showing the relative positions of peptide bands. During this time, the increasing use of slab gels and the discontinuous buffer system described by Laemmli and Favre (1973) resulted in improved resolutions of CNBr peptide bands. For the purpose of the current discussion, the authors fall into two groups (Table 3.7). The first group comprises those who indicated that peptide α1−CB7 migrates faster than peptide α1−CB8 in analytical SDS−PAGE as exemplified by Scott and Veis (1976b). The second group comprises those who indicated the reverse, as exemplified by Herbage et al. (1977) and Burke et al. (1977). The second group has by far turned out to be the larger group and only some selected authors are shown within this group in Table 3.7.

Very few of these authors indicated that they performed SDS−PAGE on purified fractions of peptides α1−CB7 and α1−CB8. Thus, their assignments of CNBr peptide band identities probably relied mainly on results of previous authors and also on the available molecular weight data.

Dabbous et al. (1981) indicated a faster−migrating peptide α1−CB7 (relative to peptide α1−CB8) quoting Scott and Veis (1976b). Interestingly enough, in the same year "Scott" himself reversed his earlier assignment of these peptide bands (Scott and Edwards, 1981; see fig. 3.13[b]). Perhaps these inconsistencies were brought to light most clearly by the assignments used by Light and Bailey (1979, 1980). Although in 1979 they labelled the faster−migrating peptide as peptide α1−CB8, they also presented the same gel in 1980 with the reversed assignments for the bands concerned (see fig. 3.14). In fact the legend (not shown) describing the gel shown here as figure 3.14[a] also indicated that peptide α1−CB7 had the higher electrophoretic mobility,

Table 3.7

Some authors who have analysed cyanogen bromide (CNBr) peptides of type I collagen by sodium dodecyl sulphate-polyacrylamide gel electropheresis, indicating the relative electrophoretic mobilities of peptides α1-CB7 and α1-CB8.

	Electrophoretic mobility	
	α1-CB7 ↘ α1-CB8	α1-CB8 ↘ α1-CB7
Year		
1976	Scott and Veis*	
1977		Herbage *et al.*; Burke *et al.*
1978		Eyre *et al.*; Davison
1979		Sage and Bornstein; Light and Bailey.
1980	Light and Bailey	
1981	Dabbous *et al.*	Scott† and Edwards; Laurent *et al.*
1982		Kenney *et al.*; Tseng *et al.* Newsome *et al.*
1984		Heidemann and Linnert
1985	Kent, Light and Bailey	O'Driscoll *et al.*
1986		*J. Ibrahim (present work)*

*(1976b)
†*of " Scott and Veis* "*

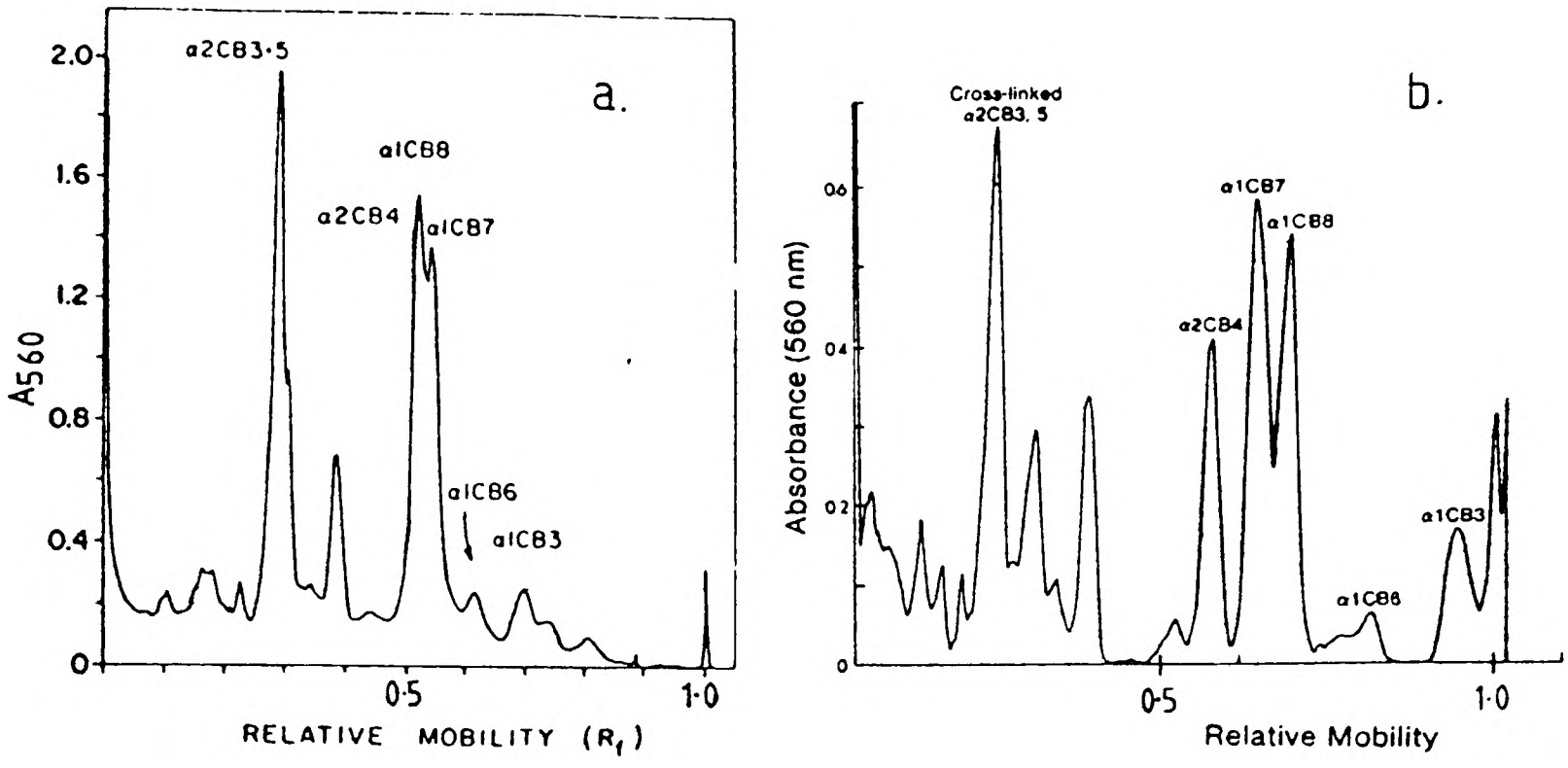


Figure 3.13 Optical densitometric scans of SDS−gel electrophoretic patterns of CNBr digests of collagen (mainly type I) showing the relative electrophoretic mobilities of peptide bands. a) from *Scott and Veis (1976b);* b) from *Scott and Edwards (1981).* Note the changes for 'α1−CB7' and 'α1−CB8' relative to one another.

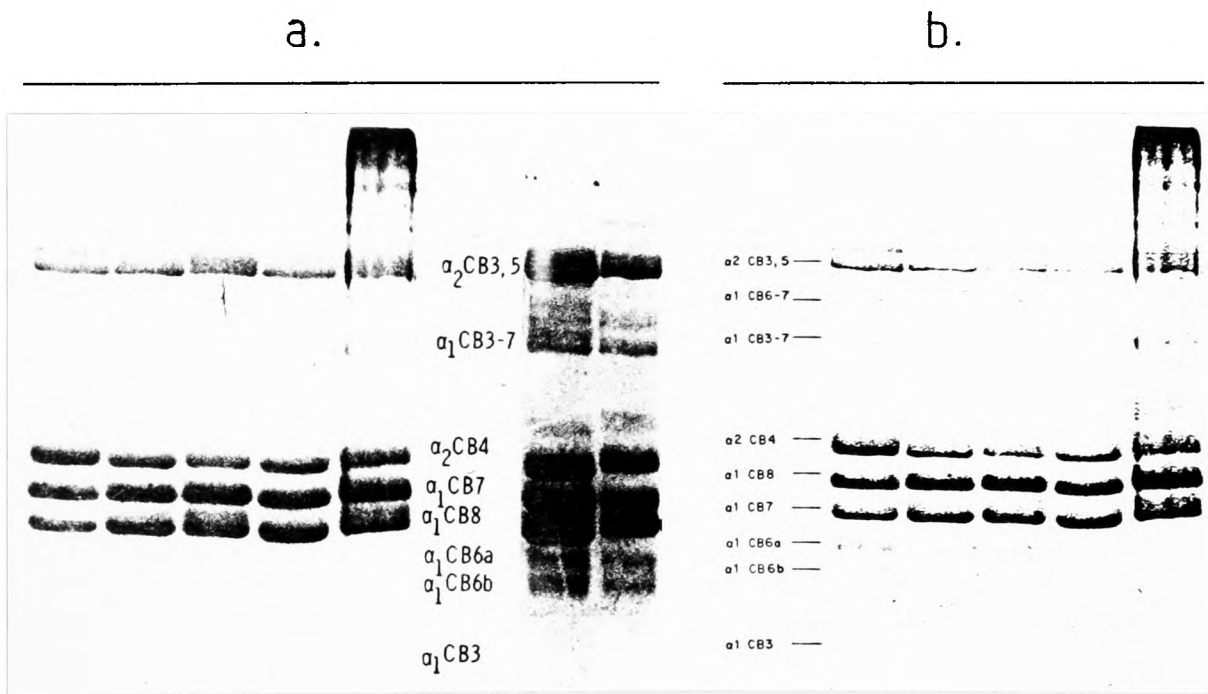


Figure 3.14 SDS−PAGE of CNBr peptides of type I collagen showing the relative positions quoted for peptide bands. a) from *Light and Bailey, 1979*; b) from *Light and Bailey, 1980).* Note the conflicting assignments for peptides α1−CB7 and α1−CB8 between a) and b).

conflicting with the labelling on the gel presented. Although these conflicting assignments may have been caused by genuine errors during the preparation of manuscripts for publication, this seems highly unlikely as these same authors presented photographs of different slab gels in 1985, indicating a faster−migrating peptide α1−CB7 (Kent, Light and Bailey, 1985). Needless to say, it is of crucial importance to know the correct identities of peptide bands on SDS−gels for example for the purpose of those who routinely use optical−scanning of SDS−gels for semi−quantitative determination of cross−linking between CNBr peptides (e.g. see Scott and Veis, 1976a, 1976b; Light and Bailey, 1979; Kent, Light and Bailey, 1985).

The nature of the current project required knowledge of the correct relative mobilities of peptides α1−CB7 and α1−CB8 on SDS−gels. In this work, crude fractions were putatively assigned their CNBr−peptide identities according to their expected elution positions during ion−exchange chromatography on CM 52 at pH 3.6 (see figs. 3.2, 3.4). Further purification of putative peptides α1−CB7 and α1−CB8 enabled amino acid analysis and N−terminal determination (of hydroxyproline in peptide α1−CB7) to be carried out. These modes of characterisation were coupled with the amino acid composition or sequence data available for peptides α1−CB7 and α1−CB8 from calf skin collagen (refer to sections IIB and IIC in this chapter) to facilitate positive identification of these peptides from corneal collagen. Electrophoresis of purified peptides α1−CB7 and α1−CB8 as well as a fraction enriched in peptide α2−CB4 on SDS−gels showed peptide α1−CB8 to be the fastest−migrating peptide of the three, followed by peptide α1−CB7 and then peptide α2−CB4 (fig. 3.15; see also figs. 3.6, 3.9, 3.10, 3.12). This therefore confirms the assignments used for peptides α1−CB7 and α1−CB8 by all the authors who agreed with Herbage et al. (1977) and Burke et al.(1977) and refutes the assignments by those who did otherwise (see Table 3.7), for example Kent, Light and Bailey (1985).

It is likely therefore, that the authors who indicated a faster−migrating peptide α1−CB7 relative to peptide α1−CB8 on SDS−gels would discover their mistake if they were to perform SDS−PAGE of the *purified* peptides simultaneously, and in adjoining

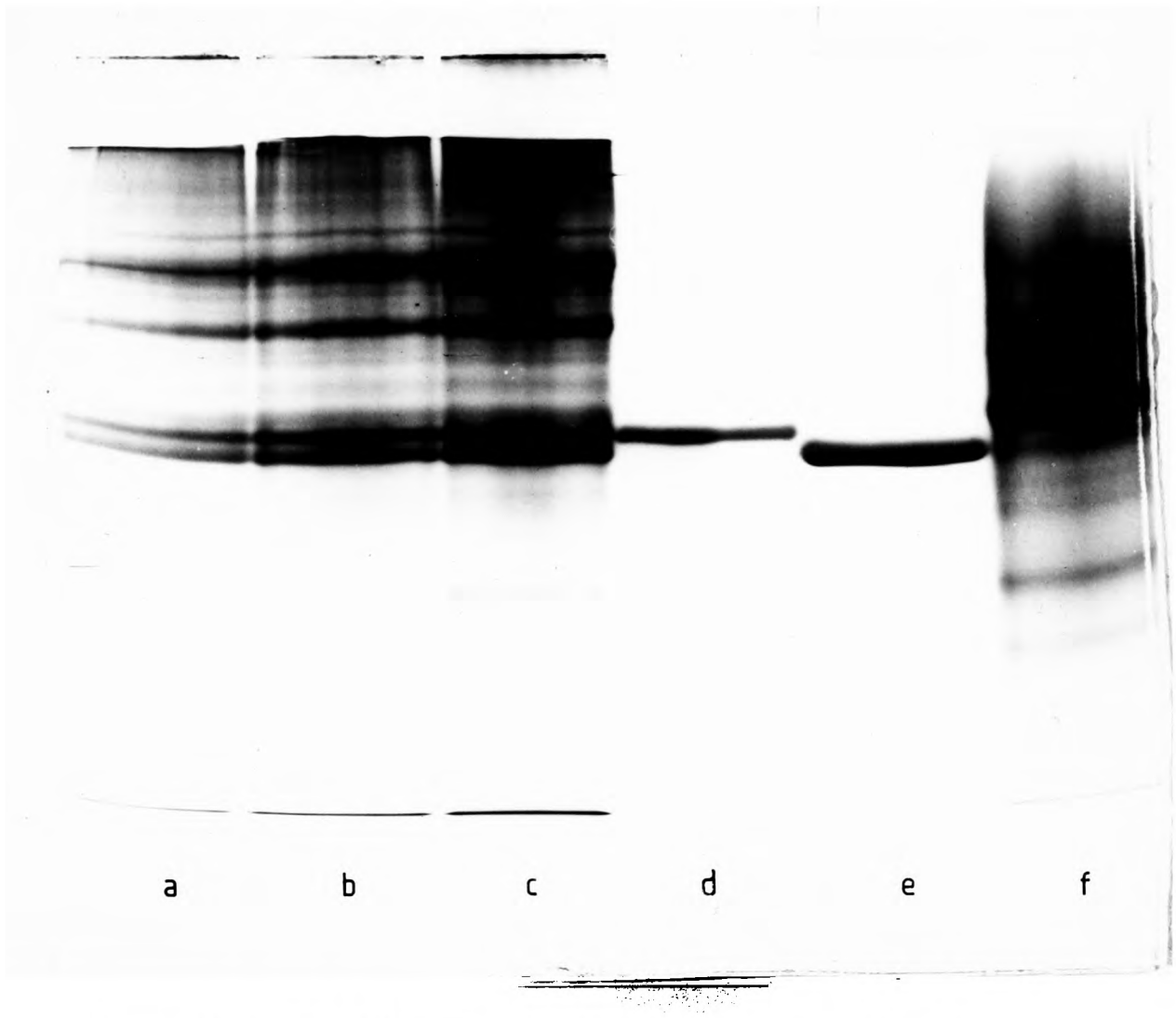


Figure 3.15 Clarification of the relative electrophoretic mobilities of peptides $\alpha 1-CB7$ and $\alpha 1-CB8$ on SDS−gels. Tracks *a*−*c*: total CNBr digests of bovine corneal collagen.
Track *d*: pure peptide $\alpha 1-CB7$ of bovine corneal collagen.
Track *e*: pure peptide $\alpha 1-CB8$ of bovine corneal collagen.
Track *f*; a fraction enriched in peptide $\alpha 2-CB4$ of bovine corneal collagen.

tracks of the same slab gel. It is highly unlikely that the use of type I collagen from different tissues or electrophoresis conditions with minor differences would produce results that would run counter to those presented in this work.

The assignment of identities to CNBr−peptide bands on SDS−gels presented in this work relied initially on the elution positions of the crude fractions in the 'pH 3.6' fractionations (see figs. 3.2−3.6). In addition to the confirmation of peptide α1−CB7 and peptide α1−CB8 band identities, amino acid analysis also confirmed the assignment for peptide α1−CB3 (see Table 3.5). The assignment for peptide α2−CB3,5 was supported by the amino acid analysis data of Panjwani and Harding (1978a) while the assignment for the cross−linked peptide "α1−CB6 dimer" was supported by amino acid analysis as well as partial amino acid sequence data (Harding and Crabbe, 1979; J.J. Harding, *personal communication*). The relative position of peptide α1−CB6 on the SDS gels in this work was not proven but is in line with the assignments used by most authors (see Bornstein and Sage, 1980; Laurent et al., 1981). This was also the case for peptide α2−CB4.

From their known amino acid sequences, peptides α1−CB7, α1−CB8 and α2−CB4 have imino acid contents comprising 24%, 22% and 21% of their total residues respectively. These differences may explain the lower mobility of the 271−residue peptide α1−CB7 relative to the 279−residue peptide α1−CB8 and also why peptide α2−CB4, a 321−residue peptide, is only just resolved from peptide α1−CB7 on SDS−gels (see fig. 3.15) due to the postulated 'retarding' effect of imino acid residues on the electrophoretic mobility of 'collagen' when compared to 'globular' proteins (Furthmayr and Timpl, 1971). It should be realised however, that the lower electrophoretic mobilities of 'collagenous' proteins relative to 'globular' proteins could also be simply due in part, to the lower mean residue weight for collagen of 91.6 daltons when compared to that for 'globular' proteins of 115.0 daltons (see Butkowski et al., 1982). This is due to the relative abundance of the small amino acids such as glycine, proline, hydroxyproline and alanine in collagen.

CHAPTER 4

PINPOINTING THE SITES OF HYDROXYLYSINE GLYCOSIDES IN PEPTIDE α1-CB7 OF BOVINE CORNEAL COLLAGEN

The amino acid sequence of the 271-residue peptide α1-CB7 of calf skin collagen (fig. 4.1) indicates that this peptide contains eleven lysyl residues occurring at positions 564, 573, 581, 603, 648, 657, 684, 725, 729, 756 and 806 in the helical region of the α1(I) chain. Of these, those at positions 581, 725 and 806 occur in 'post-Gly' positions (position 'X' in the repeating 'Gly-X-Y' triplet sequence). Amino acid sequence studies of vertebrate collagens have so far indicated that 'post-Gly' lysyl residues are never hydroxylated, and thus never occur as hydroxylysine or hydroxylysine glycoside whereas 'pre-Gly' (those in the 'Y' positions) lysyl residues have been found to occur as hydroxylysine and hydroxylysine glycoside (see Fietzek and Kühn, 1976; Butler et al., 1977; Francis et al., 1978; Babel and Glanville, 1984; Glanville et al., 1985). The eight 'pre-Gly' lysyl residues at positions 564, 573, 603, 648, 657, 684, 729 and 756 are therefore the 'possible' sites of hydroxylysine glycosides in peptide α1-CB7.

This chapter describes the characterisation of all the 'pre-Gly' and 'post-Gly' lysyl positions of peptide α1-CB7. This was achieved by initially cleaving the purified corneal collagen peptide with the proteolytic enzymes trypsin and V8-protease whose peptide-bond cleavage specificities are known. Using the known amino acid sequence of peptide α1-CB7 from calf skin, isolated tryptic and V8-protease peptides of the corneal peptide were then identified by their amino acid compositions, simultaneously revealing the positions of their lysyl residues. Identified peptides whose amino acid compositions contained hydroxylysine were subsequently subjected to alkaline hydrolysis and analysed for hydroxylysine glycoside (see 'Methods', Chapter 2.IIH), thus revealing the presence or absence of hydroxylysine glycoside at the deduced lysyl positions.

I. TRYPTIC PEPTIDES OF CITRACONYLATED PEPTIDE α1-CB7

Trypsin catalyses the hydrolysis of peptide bonds on the carboxyl side of lysine and arginine residues (Bergmann et al., 1939). The cleavage is slow when the lysine or arginine residues are adjacent to acidic residues or cystine (Hirs et al., 1956) and does not occur when lysine or arginine is followed by proline (Margoliash et al., 1961) or hydroxyproline (Rexrodt et al., 1973).

Although cleavage by trypsin is widely used in studies of protein primary structure due to its reliable specificity towards lysine and arginine residues, it does have a serious disadvantage with respect to the study of hydroxylysine glycosides. Where lysyl residues occur as hydroxylysine glycoside, it has been found that peptide bonds on the carboxyl side of these residues resist cleavage by trypsin (Francis et al., 1978; Babel and Glanville, 1984). As post-translational modifications of lysyl residues during collagen biosynthesis do not necessarily go to completion (see Chapter 1.III), the calculated 'percentage' glycosylation at the glycoside sites are not representative of the true situation unless it can be shown that all the peptides resulting from cleavage and non-cleavage of lysyl bonds have been recovered, analysed and totally accounted for in the calculations. This complication is of course, in addition to that already presented in the task of isolating pure peptides from such a heterogeneous mixture of a large number of peptides. For these reasons, it is difficult for example, to assess the degree of validity of the values obtained for glycoside sites that have been determined in the bovine α1(II) chain (see Butler et al., 1977; Francis et al., 1978).

These complications can however be avoided by ensuring that all 'Lys—' bonds are left uncleaved, hence limiting tryptic cleavage solely to 'Arg—' bonds. 'Masking' of lysyl residues for this purpose can be achieved by 'trifluoroacetylation' (Goldberger and Anfinsen, 1962), 'maleylation' (Butler, P.J.G. et al., 1967) or 'citraconylation' (Dixon and Perham, 1968) of the substrate prior to tryptic cleavage.

Tryptic cleavage of citraconylated peptide α1-CB7 theoretically yields thirteen peptides referred to as peptides Tc1 to Tc13 in the present work (fig. 4.1). Looking at the amino acid sequence of peptide α1-CB7 (of calf skin collagen), peptides Tc2,

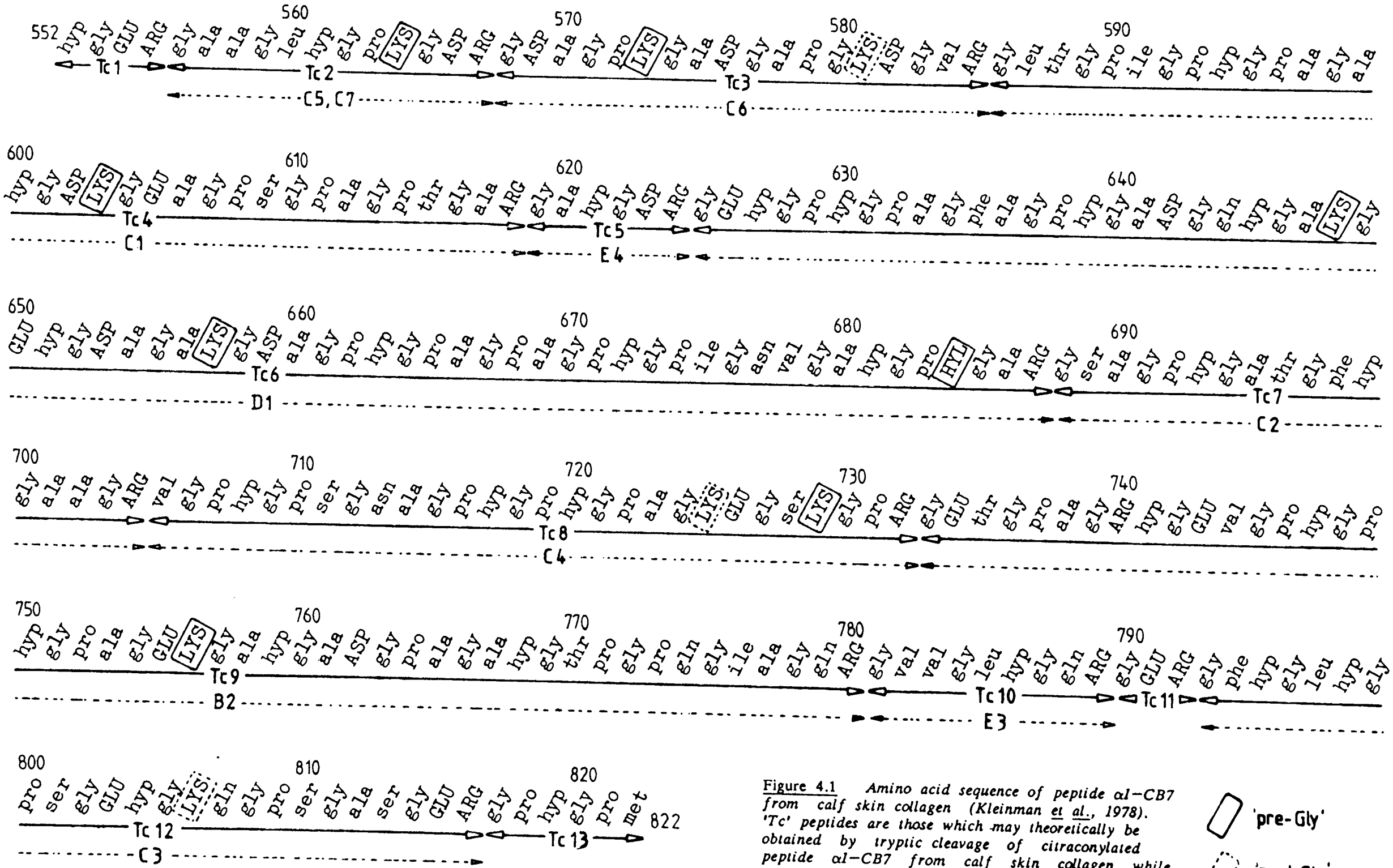


Figure 4.1 *Amino acid sequence of peptide α1–CB7 from calf skin collagen (Kleinman et al., 1978). 'Tc' peptides are those which may theoretically be obtained by tryptic cleavage of citraconylated peptide α1–CB7 from calf skin collagen while 'B', 'C', 'D' and 'E' peptides are those isolated in this work from tryptic digests of citraconylated peptide α1–CB7 from bovine corneal collagen.*

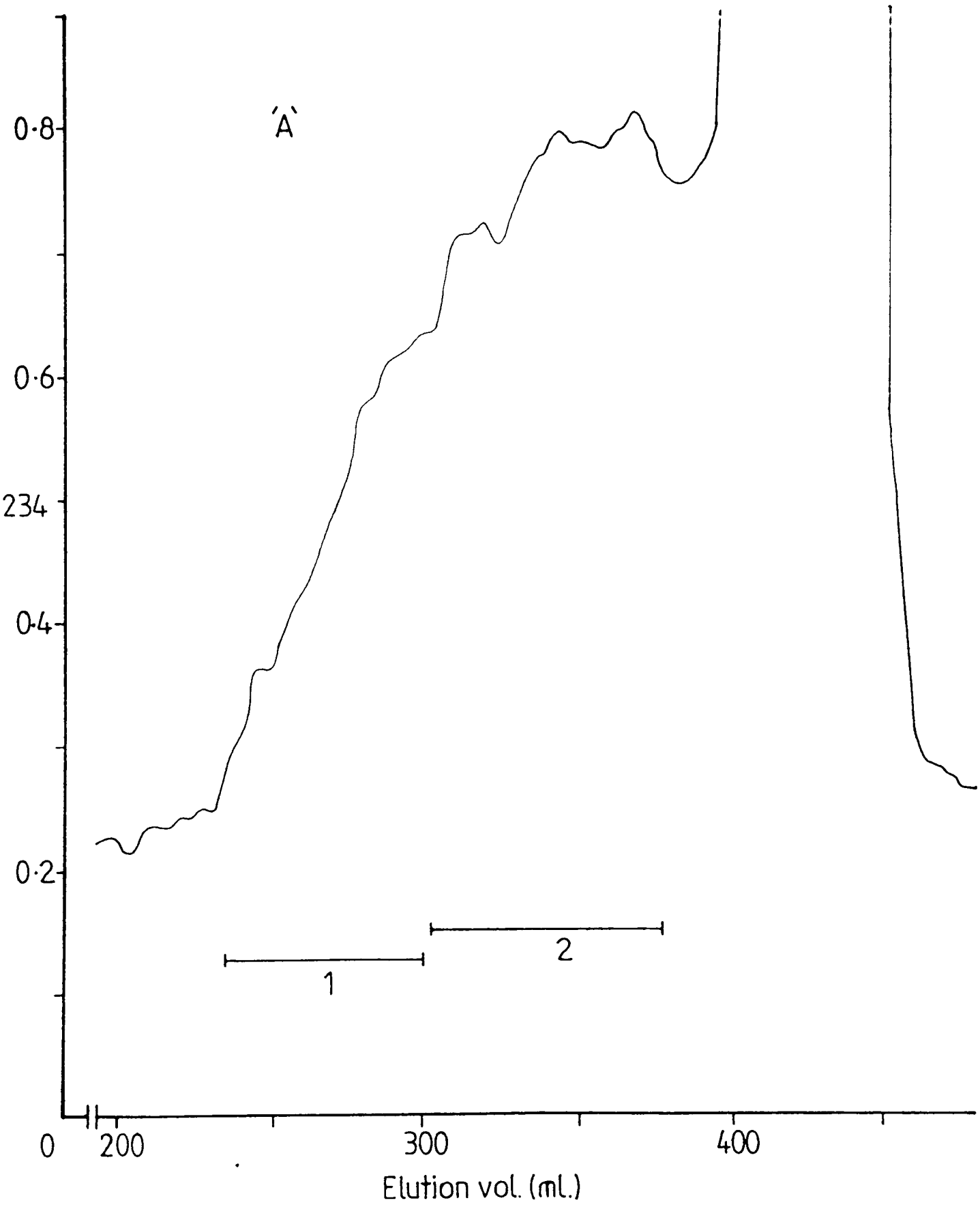


Figure 4.2 Fractionation 'A': Gel chromatography of the first batch of tryptic peptides of citraconylated peptide α1−CB7 from bovine corneal collagen on Sephadex G−50, superfine (105 × 2.3 cm) at 37°C. The column was eluted with 100mM acetic acid at a flow−rate of 19ml/hr. 4ml fractions were collected.

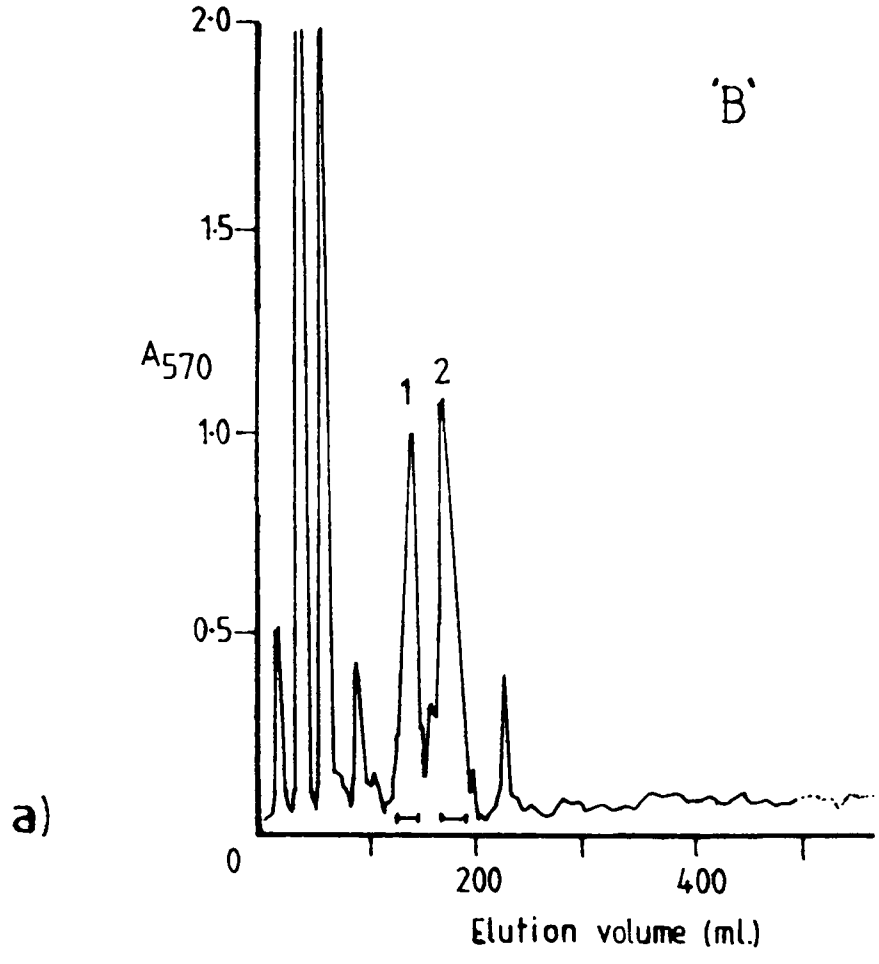


Figure 4.3 Fractionations 'B' and 'C': Ion−exchange chromatography on AG50W−X4 resin (100 × 1 cm) of fractions containing tryptic peptides of citraconylated peptide α1−CB7 from bovine corneal collagen. The column was eluted with sequential pyridine acetate buffers. For full details of chromatography, see 'Methods' (Ch2.IID1). Bars indicate some of the fractions pooled. a) Fractionation 'B': chromatography of fraction *A1* (see fig. 4.2). b) Fractionation 'C': chromatography of fraction *A2* (see fig. 4.2)

Tc3, Tc4, Tc8 and Tc9 have one 'pre−Gly' lysyl residue each. The longest Tc peptide, the 63−residue peptide Tc6 contains three 'pre−Gly' lysyl residues and thus is the only Tc peptide with more than one possible site of hydroxylysine glycoside. One of these, position 684, was found to be 65%−70% hydroxylated in calf skin collagen and is represented as 'Hyl' (Fietzek et al., 1973; Kleinman et al., 1978).

Peptides Tc3 and Tc8 also have one 'post−Gly' lysyl residue each. Peptide Tc12 has one lysyl residue which occurs in a 'post−Gly' position. The other Tc peptides Tc1, Tc5, Tc7, Tc10, Tc11 and Tc13 do not contain any lysyl positions.

I A. ISOLATION OF TRYPTIC PEPTIDES OF CITRACONYLATED PEPTIDE α1−CB7

Two batches of peptide α1−CB7 were citraconylated and cleaved with trypsin on two separate occasions. Following tryptic digestion, citraconyl groups were removed from the ε−amino groups of lysyl and hydroxylysyl residues. After freeze−drying, the digests were initially fractionated, on two separate occasions, on a Sephadex G−50_S column, 100 × 2.3 cm at 37°C. The results of these fractionations are now discussed in turn.

1. First batch of tryptic peptides

The first tryptic digest (60mg) was dissolved in 2.5ml 100mM acetic acid pH 3.0 and fractionated on Sephadex G−50_S in the same buffer, eluting at a flow−rate of 19ml per hour. The chromatogram is shown in figure 4.2. The largest, 63−residue theoretical peptide Tc6 was expected to elute first, resolved from the other Tc peptides of lengths 48 residues down to 3 residues (see fig. 4.1). The chromatogram indicated that this was not achieved. The fractions were pooled as two pools, A1 and A2 (fig. 4.2), and freeze−dried for further fractionation by ion−exchange chromatography on the high−resolution AG50W−X4 column.

Fractionation of pool A1 (35mg) on AG50W−X4 resin gave the chromatogram shown in figure 4.3(a) while fractionation of pool A2 (20mg) gave the chromatogram shown in figure 4.3(b). The test−tube fractions obtained from these two fractionations

'B' and 'C' were pooled as shown and freeze−dried. Portions of 'B' and 'C' peptides were hydrolysed for amino acid analysis. The amino acid compositions of these fractions were then compared with those of expected Tc peptides. The amino acid compositions of 'B' and 'C' peptides which matched those of the expected Tc peptides are shown in Table 4.1. The other 'B' and 'C' fractions gave amino acid compositions (not shown) which, by their high glycine and imino acid contents, indicated that they were mixtures of collagen peptides.

Of the peptide fractions obtained from fractionation 'B', fraction B2 gave an amino acid composition matching that of (theoretical) peptide Tc9, a 48−residue peptide which contains one 'pre−Gly' lysyl residue at position 756 (Table 4.1; fig. 4.1). The composition of fraction B2 showed that it contained 0.11 'total' hydroxylysine residues per peptide. The single isoleucine residue found per peptide B2 (Tc9) presumably occurs at position 776, where it participates in the single 'Gly−Ile' bond of the α1(I) chain which is cleaved by animal collagenase. Peptide Tc9 also contains the sequence 766−786 which is thought to be the binding site for fibronectin (Kleinman et al., 1981a). The overall yield (calculated from the quantity of peptide α1−CB7 digested) for peptide B2 was 16%.

Fractionation 'B', the fractionation of sample A1, expected to contain the large 63−residue peptide Tc6, did not yield any peptide fractions which matched that of Tc6. Fraction B1, a major peak in fractionation 'B' (fig. 4.3 [a]), did however give an amino acid composition (not shown) which indicated that it was a mixture of peptide Tc6 and other tryptic peptides of peptide α1−CB7. Further fractionation of fraction B1 would have been necessary if peptide Tc6 was to be purified from it. This was not done but a second attempt was made to isolate peptide Tc6 from a subsequent tryptic digest (described later).

Fractionation 'C', the separation of the smaller tryptic peptides contained in fraction A2, gave a number of large and small peaks (fig. 4.3[b]).

The amino acid composition of fraction C1 matched that of the 33−residue peptide Tc4, the third largest theoretical tryptic peptide (Table 4.1; fig. 4.1). Peptide

Table 4.1

Amino acid compositions of tryptic peptides of citraconylated peptide α1-CB7 from bovine corneal collagen isolated from fractionations 'B' and 'C' alongside those of theoretical 'Tc' peptides predicted from the amino acid sequence of calf skin peptide α1-CB7 *(fig. 4.1)*. Values are expressed as residues per peptide. '0' indicates values less than 0.2. For 'hydroxylysine' the lower limit is set at 0.1.

	C5	C7	*Tc2*	C6	*Tc3*	C1	*Tc4*	C2	*Tc7*	C4	*Tc8*	B2	*Tc9*	C3	*Tc12*
Hydroxyproline	1.1	1.3	1	0	0	2.4	2	2.2	2	2.8	3	4.3	5	2.3	3
Aspartic acid	1.1	1.1	1	2.8	3	1.2	1	0	0	1.2	1	1.2	1	0	0
Threonine	0	0	0	0	0	1.7	2	1.1	1	0	0	2.1	2	0	0
Serine	0	0	0	0	0	0.8	1	1.0	1	1.8	2	0	0	3.1	3
Glutamic acid	0	0	0	0.2	0	1.2	1	0	0	1.2	1	5.0	5	3.0	3
Proline	1.2	1.2	1	1.7	2	6.1	6	1.2	1	5.7	6	7.3	7	3.0	2
Glycine	3.7	3.5	4	6.4	6	10.7	11	5.8	6	8.8	9	15.8	16	7.6	8
Alanine	2.0	1.9	2	2.6	3	4.8	5	3.9	4	2.4	2	7.0	7	1.0	1
Valine	0	0	0	1.0	1	0	0	0	0	1.0	1	1.2	1	0	0
Isoleucine	0	0	0	0.2	0	1.0	1	0	0	0	0	1.0	1	0	0
Leucine	0.8	1.1	1	0	0	1.1	1	0	0	0	0	0	0	1.1	1
Phenylalanine	0	0	0	0	0	0	0	0.9	1	0	0	0	0	1.0	1
Hydroxylysine	0.8	0	0	0.18	0	0.17	0	0	0	0	0	0.11	0	0	0
Lysine	0.4	1.0	1	1.86	2	0.93	1	0	0	2.1	2	1.1	1	1.0	1
Arginine	1.0	1.0	1	1.1	1	0.9	1	0.9	1	1.0	1	1.9	2	0.9	1
total	12.1	12.1	12	18.04	18	33.0	33	17.0	17	28.0	28	48.01	48	24.0	24

Tc4 contains one 'pre−Gly' lysyl residue at position 603. Analysis of peptide C1 (Tc4) indicated the occurrence of 0.17 hydroxylysine residues at this position per peptide. The overall yield of peptide C1 was 42%.

The amino acid composition of fraction C2 matched that of the 17−residue peptide Tc7 which does not contain lysyl residues. The overall yield of peptide C2 was 43%.

Analysis of fraction C3 showed it to be the 24−residue peptide Tc12 which contains a 'post−Gly' lysyl residue at position 806 (Table 4.1; fig. 4.1). The absence of hydroxylysine from its amino acid composition was thus in keeping with the 'post−Gly' position of this lysine. The overall yield of peptide C3 was 28%.

Amino acid analysis of fraction C4 gave a composition matching that of peptide Tc8, a 28−residue peptide containing one 'post−Gly' lysyl residue at position 725 and one 'pre−Gly' lysyl residue at position 729 (Table 4.1; fig. 4.1). The amino acid composition of peptide C4 showed an absence of hydroxylysine, indicating that neither of these two lysyl residues were hydroxylated. The overall yield of peptide C4 was 22%.

Fractions C5 and C7 both gave amino acid compositions which matched that of peptide Tc2, a 12−residue peptide which contains a 'pre−Gly' lysyl residue at position 564 (Table 4.1; fig. 4.1). Fraction C5 contained 0.8 hydroxylysine residues per peptide while fraction C7 contained 1.0 residues of lysine and no hydroxylysine. Separation of pairs of peptides with and without hydroxylated lysine has been reported before (Rexrodt et al., 1973; Dixit et al., 1975; Panjwani and Harding, 1978b). From the overall yields of fractions C5 and C7 which were 7% and 14% respectively, peptide C5/C7 (Tc2) was determined as having 0.27 residues of hydroxylysine at position 564. The total overall yield for this peptide was therefore 21%.

Amino acid analysis of fraction C6 showed it to be the 18−residue peptide Tc3 which contains two lysine residues, one 'pre−' and one 'post−' glycine, at positions 573 and 581 respectively (Table 4.1; fig. 4.1). The 0.18 hydroxylysine residues per peptide found in the amino acid composition of peptide C6 therefore most probably

resides at position 573. The overall yield of peptide C6 was 15%. The 0.2 glutamate residues per peptide (Table 4.1) probably arose from a minor contaminant.

Those identified peptides isolated in fractionations 'B' and 'C', and which contained hydroxylysine, were stored at –20°C until they were subjected to alkaline hydrolysis for hydroxylysine glycoside analysis. Analysis of their alkaline hydrolysates will be described later (section IB).

The theoretical tryptic peptides Tc1, Tc5, Tc10, Tc11 and Tc13, whose lengths are 4, 6, 9, 3 and 6 residues respectively, were not recovered in either fractionations 'B' or 'C'. It was thus likely that they were present in the test–tube fractions following the pooled fraction A2 (see fig. 4.2). Although these fractions were pooled and freeze–dried, the products were unfortunately discarded as this fraction was thought to consist mainly of *salts* derived from buffers. However, with respect to the location of glycoside sites in peptide α1–CB7, these peptides were not important as the peptide sequence (fig. 4.1) shows that they do not contain any lysyl positions. Attempts were made to isolate these peptides from a subsequent tryptic digest. This will be described in the next section.

Of the total 1.8 hydroxylysine residues per peptide α1–CB7 of bovine corneal collagen (see Table 3.3), only 0.73 residues have so far been accounted for by peptides B2 (Tc9), C1 (Tc4), C5/C7 (Tc2) and C6 (Tc3). By deduction, 1.07 hydroxylysine residues therefore occur in peptide Tc6 (which was still to be isolated at this stage), spread between the three 'pre–Gly' lysyl positions 648, 657 and 684 (see fig. 4.1).

A second batch of peptide α1–CB7 was thus citraconylated and cleaved with trypsin, the prime aim being to isolate peptide Tc6 containing positions 648, 657 and 684.

2. Second batch of tryptic peptides

A second tryptic digest of (citraconylated) peptide α1–CB7 (51mg) was dissolved in 2.5ml of 30mM ammonium acetate buffer pH 4.8 and fractionated in this buffer on

Sephadex G−50_S. In an attempt to improve on the resolution of peptide peaks obtained in fractionation 'A' of the first tryptic digest, elution was carried out at a slower flow−rate of 10ml per hour. The resulting chromatogram shown in figure 4.4 indicates a slight improvement in the resolution of peptide peaks in this fractionation, 'D', over that in fractionation 'A' (fig. 4.2). Fractions were pooled as fractions D1 to D7 as indicated and freeze−dried.

Amino acid analysis of fraction D1 gave an amino acid composition, shown in Table 4.2, which matched that of peptide Tc6, the largest, 63−residue theoretical tryptic peptide of peptide α1−CB7 (fig. 4.1) which was not isolated from the first tryptic digest. Peptide Tc6 contains 'pre−Gly' lysyl residues at positions 648 and 657, and a 65%−70% hydroxylated 'pre−Gly' lysyl residue, indicated as "Hyl"$_{\wedge}^{(684)}$ in the sequence of peptide α1−CB7 from calf skin. Peptide D1 (Tc6) contained 1.1 hydroxylysine residues out of a total of 3.0 residues of 'Lys + Hyl' per peptide. The overall yield of peptide D1 was 41%.

The quantity of hydroxylysine found in peptide D1 plus that found in peptides B2, C1, C5/C7 and C6 together amount to 1.83 residues of hydroxylysine which is in good agreement with the total of 1.8 hydroxylysine residues found per peptide α1−CB7 (Table 3.3).

Fractions D2 to D6 (fig. 4.4), on amino acid analysis, gave compositions (not shown) with characteristically high glycine and imino acid contents, indicating them to be mixtures of collagen peptides. Amino acid analysis carried out on fraction D7 indicated it to have very little amino acid content, indicating that this fraction consisted mainly of non−proteins, most probably 'salt'.

Fractions D5 (3.5mg) and D6 (4.5mg) were pooled together and fractionated on the AG50W−X4 ion−exchange column in an attempt to isolate the small Tc peptides not isolated from the first tryptic digest, which were peptides Tc1, Tc5, Tc10, Tc11 and Tc13, consisting of 4, 6, 9, 3 and 6 residues respectively (fig. 4.1). The chromatogram for this fractionation, 'E', is shown in figure 4.5. The fractions were pooled as indicated, freeze−dried and subjected to amino acid analysis. Amino acid

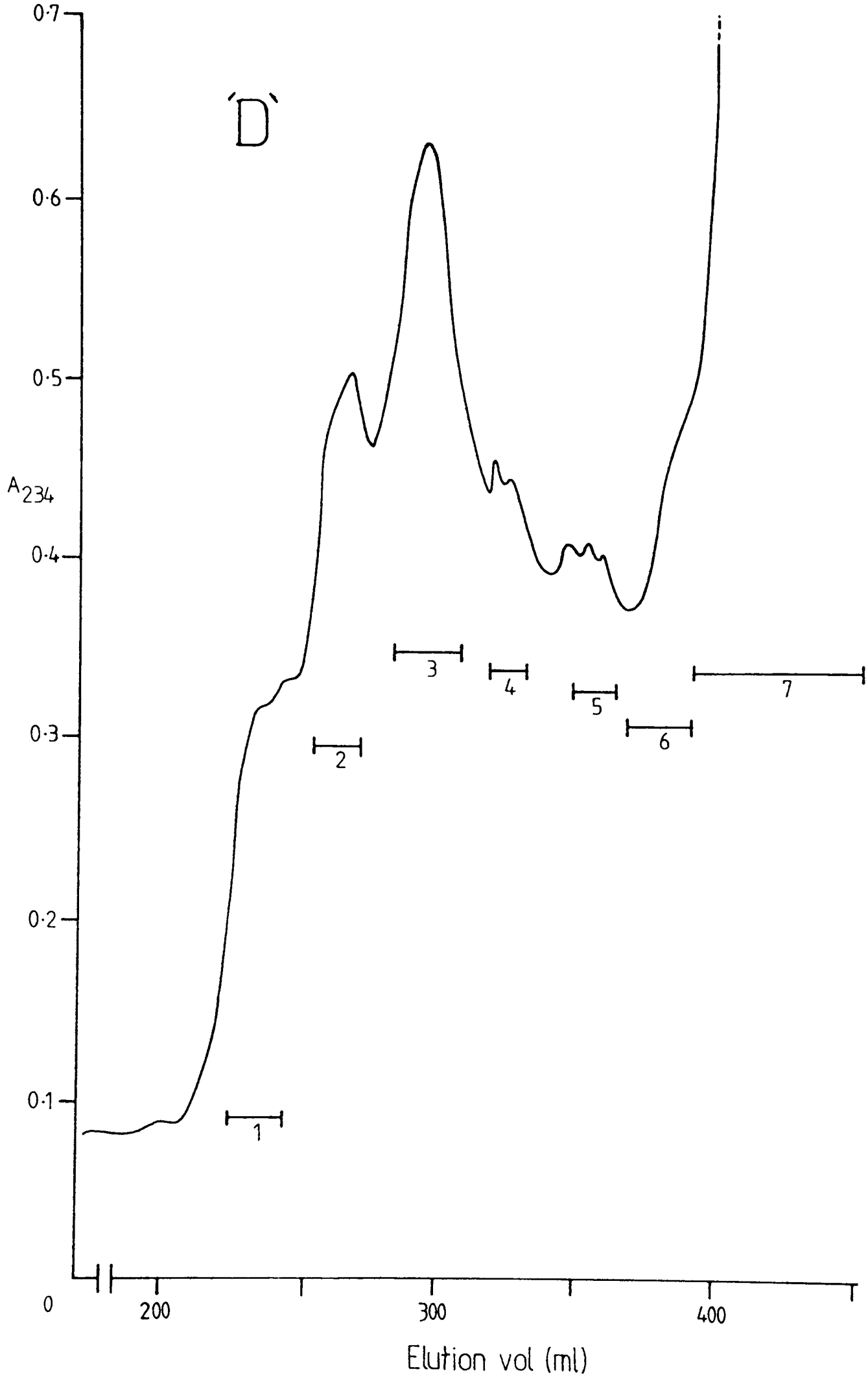


Figure 4.4 Fractionation 'D': Gel chromatography of the second batch of tryptic peptides of citraconylated peptide α1−CB7 from bovine corneal collagen on Sephadex G−50$_S$ (105 × 2.3 cm) at 37°C. The column was eluted with 30mM ammonium acetate pH 4.8 at a flow−rate of 10ml/hr. Bars indicate the fractions pooled.

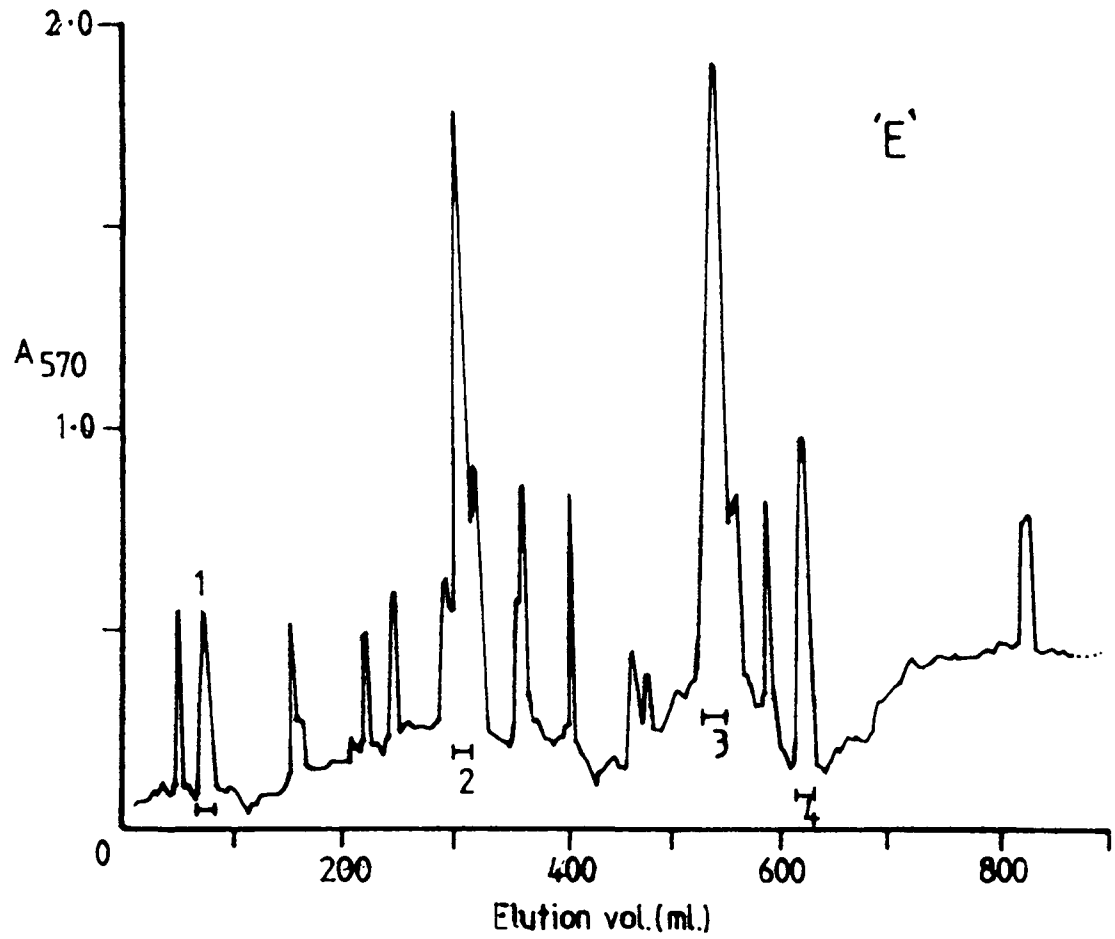


Figure 4.5 Fractionation 'E': Ion−exchange chromatography on AG50W−X4 resin of combined fractions *D5* and *D6* (see fig. 4.4). See 'Methods' (Ch 2.IID1) for full details of chromatography. Bars indicate some of the fractions pooled.

Table 4.2

Amino acid compositions of tryptic peptides of citraconylated peptide α1−CB7 from bovine corneal collagen isolated from fractionations 'D' and 'E' alongside those of theoretical Tc peptides as deduced from the amino acid sequence of peptide α1−CB7 from calf skin (see fig. 4.1). '0' signifies a value of less than 0.2. For hydroxylysine, the lower limit is set at 0.1.

	D1	*Tc6*	E4	*Tc5*	E2	*Tc7*	E3	*Tc10*
Hydroxyproline	7.0	8	0.9	1	2.1	2	1.0	1
Aspartic acid	4.3	4	1.2	1	0	0	0	0
Threonine	0	0	0	0	0.9	1	0	0
Serine	0.2	0	0	0	1.2	1	0	0
Glutamic acid	3.3	3	0	0	0	0	1.1	1
Proline	10.1	9	0.1	0	1.1	1	0	0
Glycine	20.5	21	1.8	2	5.7	6	3.0.	3
Alanine	10.6	11	1.1	1	4.2	4	0	0
Valine	1.1	1	0	0	0	0	1.8	2
Isoleucine	1.0	1	0	0	0	0	0	0
Leucine	0	0	0	0	0	0	1.2	1
Phenylalanine	1.0	1	0	0	0.8	1	0	0
Hydroxylysine	1.1	1	0	0	0	0	0	0
Lysine	1.9	2	0	0	0	0	0	0
Arginine	1.0	1	0.9	1	1.0	1	0.9	1
total	63.1	63	6.0	6	17.0	17	9.0	9

compositions of 'E' fractions which matched those of theoretical Tc peptides are presented in Table 4.2.

Amino acid analysis of fraction E2 showed it to be the 17−residue peptide Tc7 (Table 4.2; fig. 4.1). Peptide Tc7 was also isolated previously as fraction C2 from the first tryptic digest (Table 4.1), and does not contain any lysyl positions. The overall yield of peptide E2 was 34%.

The amino acid composition of fraction E3 matched that of peptide Tc10 (Table 4.2; fig. 4.1), the 9−residue peptide which was not isolated from the first tryptic digest. Peptide Tc10 does not contain any lysyl positions. The total of only one imino acid residue (hydroxyproline + proline) per peptide in peptide E3 presumably represents the residue at position 786 in the 9−residue peptide Tc10. The relatively low imino acid content in the immediate sequences surrounding the 'Gly−Ile' bond 775—776 is thought to result in the relative flexibility in this region of the α1(I) chain, rendering the 'Gly−Ile' bond susceptible to cleavage by animal collagenase (see Kleinman et al., 1981a). The overall yield of peptide E3 was 77%.

Amino acid analysis of fraction E4 showed it to be the 6−residue peptide Tc5 which was also not isolated from the first tryptic digest (Table 4.2). This peptide does not contain any lysyl positions (fig. 4.1). The yield of peptide E4 was 63%.

The amino acid composition of fraction E1 (not shown) did not correspond entirely to any expected Tc peptide but homoserine (CNBr cleavage product of methionine) was found to comprise about 4% of the total residues. Correcting this figure for losses due to the formation of 'homoserine−lactone', about 8% of the total composition can thus be attributed to homoserine and hence methionine. The only expected methionine residue in peptide α1−CB7 (or any CNBr peptide) occurs at the C−terminus and represents 16.7% of the 6−residue peptide Tc13 (see fig. 4.1). Fraction E1 was therefore a mixture which contained a significant amount of peptide Tc13. No further attempt was made to purify peptide Tc13 from fraction E1. Peptide Tc13 does not contain any lysyl positions.

The tryptic peptides isolated from this second tryptic digest plus those isolated

from the first tryptic digest have thus been identified as ten out of the thirteen theoretical Tc peptides and account for 258 out of the total 271 amino acid residue positions in peptide α1−CB7 according to the sequence from calf skin (see fig. 4.1). Peptides Tc1, Tc11 and Tc13 which were not isolated, account for the remaining 13 amino acid residue positions of peptide α1−CB7.

With respect to the eleven lysyl positions in peptide α1−CB7, the isolation of peptide D1 from this second tryptic digest means that all the lysine and hydroxylysine residues of peptide α1−CB7 from bovine corneal collagen had now been allocated to their respective tryptic peptides.

(Refer to fig. 4.1 and Tables 4.1, 4.2 for the following discussion).

In keeping with the common observation that 'post−Gly' lysyl residues have never been found to occur as hydroxylysine, it was shown that peptide C3 (Tc12), which contains Lys−806 occurring in a 'post−Gly' position, did not contain hydroxylysine. The 'post−Gly' lysyl residue at position 725 is also not hydroxylated, as shown by the absence of hydroxy lysine in the amino acid composition of peptide C4 (Tc8). Bearing this in mind, it can be assumed that the 'post−Gly' Lys−581 is also not hydroxylated.

Thus, the results presented so far have shown that 0.27, 0.18, 0.17 and 0.11 residues of hydroxylysine occur at the 'pre−Gly' lysyl positions 564, 573, 603 and 756 respectively in peptide α1−CB7. The 'pre−Gly' lysyl residue at position 729 is not hydroxylated as no hydroxylysine was found in peptide C4.

Thus, with the assumption that the 'post−Gly' lysyl residue at position 581 is not hydroxylated, all the lysyl residue positions in bovine corneal collagen peptide α1−CB7 have been individually characterised with respect to hydroxylation except for those at positions 648, 657 and 684. These lysyl residues, occurring together in peptide D1 (Tc6) are all in 'pre−Gly' positions and between them, account for 1.1 hydroxylysine residues out of the total of 1.8 residues found per peptide α1−CB7 of bovine corneal collagen (Table 3.3).

All the identified tryptic peptides of peptide α1−CB7 which were found to contain hydroxylysine were subjected to hydroxylysine glycoside analysis.

I B. QUANTITATIVE AND QUALITATIVE DETERMINATION OF HYDROXYLYSINE GLYCOSIDES IN THE TRYPTIC PEPTIDES

Tryptic peptides C5, C6, C1, D1 and B2, identified as peptides Tc2, Tc3, Tc4, Tc6 and Tc9 respectively, were hydrolysed in 2M sodium hydroxide in preparation for the determination of the diglycoside glucosylgalactosylhydroxylysine (GGH), the monoglycoside galactosylhydroxylysine (GH) and 'free', unglycosylated hydroxylysine (FH) as described in Chapter 2 (section IIH). Peptides C3 (Tc12) , C4 (Tc8), and C7(≡2) although containing lysine, were not prepared for glycoside analysis as they did not contain hydroxylysine (Table 4.1).

The results of these analyses are summarised in Table 4.3.

(Refer to fig. 4.1 for the lysyl positions quoted in the following discussion).

Glycoside analysis carried out on the alkaline hydrolysate of peptide C5 (Tc2) showed it to contain 0.21 GGH, 0.3 GH and 0.18 FH residues per peptide (Table 4.3). Peptide Tc2 was also isolated as fraction C7 which did not contain hydroxylysine (Table 4.1). It can thus be calculated from the relative yields of fractions C5 and C7 (a ratio of 1:2) that peptide Tc2 contains 0.07 GGH, 0.1 GH and 0.06 FH residues per peptide. The sum of these values, at 0.23 residues, is close to the value of 0.27 (total) hydroxylysine residues per peptide C5/C7 (Tc2) as determined by amino acid analysis of the corresponding acid hydrolysates. It was concluded therefore, that 0.07 GGH and 0.1 GH residues occur at position 564 in peptide α1−CB7.

The alkaline hydrolysate of peptide C6 (Tc3) contained 0.06 GGH, 0.06 GH and 0.06 FH residues per peptide (Table 4.3). The sum of these values, at 0.18 residues is equal to the value of 0.18 hydroxylysine residues per peptide as determined from its acid hydrolysate. Although peptide C6 (Tc3) contains two lysyl positions, it was concluded that 0.06 GGH and 0.06 GH occur at position 573 per peptide as the other lysyl residue (Lys−581) occurs in a 'post−Gly' position.

The alkaline hydrolysate of peptide C1 (Tc4) contained 0.04 GGH, 0.06 GH and

Table 4.3

Hydroxylysine glycoside and hydroxylysine contents of tryptic peptides from bovine corneal peptide α1−CB7 isolated in this work. The values determined for 'whole' corneal peptide α1−CB7 are displayed for reference as are the values for *'total−hydroxylysine'* as determined previously in the respective acid hydrolysates (see * below). Values are expressed as residues per peptide. The columns specific for glycoside contents are headed in **bold**.

peptide	**GGH**	**GH**	*FH*	*'hyl'**
C5†	0.21	0.3	0.18	*0.8*
C7†	–	–	–	*0*
C5/C7‡	0.07	0.1	0.06	*0.27*
C6	0.06	0.06	0.06	*0.18*
C1	0.04	0.06	0.05	*0.17*
D1	0.27	0.40	0.48	*1.1*
C4	–	–	–	*0*
B2	0	0.08	0.04	*0.11*
C3	–	–	–	*0*
α1-CB7	0.40	0.57	0.78	*1.8*

GGH, glucosylgalactosylhydroxylysine; GH, galactosylhydroxylysine; *FH*, free (unglycosylated) hydroxylysine

† Identical peptides recovered from two separate peaks of a single fractionation.

‡ Weighted mean values, the yield ratio C5:C7 being 1:2

– Not analysed for hydroxylysine glycoside as the peptide did not contain hydroxylysine.

Note: Identified peptides which did not contain lysine are automatically excluded from this table.

0.05 FH residues per peptide (Table 4.3). The sum of these values, at 0.15 residues, is close to the value of 0.17 hydroxylysine residues per peptide as determined from the composition of its acid hydrolysate. It was concluded therefore, that 0.04 GGH and 0.06 GH residues occur at position 603 in peptide α1−CB7.

Analysis of the alkaline hydrolysate of peptide D1 (Tc6) showed that it contained 0.27 GGH, 0.40 GH and 0.48 FH residues per peptide (Table 4.3). The sum of these values, at 1.15 residues is in good agreement with the value of 1.1 hydroxylysine residues per peptide as determined from the composition of its acid hydrolysate. Peptide D1 (Tc6) contains three 'pre−Gly' lysyl residues at positions 648, 657 and 684 (fig. 4.1). Position 684, represented as "Hyl" in the amino acid sequence of calf skin peptide α1−CB7, is 65%−70% hydroxylated in the calf skin peptide. The identical position in the α1(I) chain of chick skin is 70% hydroxylated (Highberger et al., 1975). The lysyl residues at positions 648 and 657 in the calf and chick skin peptides were not reported as being hydroxylated. It may be expected therefore, that a major portion of the hydroxylysine glycoside found in peptide D1 (Tc6) in the present work, is at position 684. However, it was still not possible to rule out Lys−648 and Lys−657 as possible glycoside sites as two particular lysyl positions have been shown to be unmodified in calf skin peptide α1−CB3 but significantly modified in bovine corneal peptide α1−CB3 (Panjwani and Harding, 1978b).

With respect to the last 'pre−Gly' lysyl residue at position 756 in peptide α1−CB7, glycoside analysis of tryptic peptide B2 (Tc9) showed that it contained 0.08 GH and 0.04 FH residues per peptide (Table 4.3). No GGH was detected. The sum of these values, at 0.12 residues, is close to the value of 0.11 hydroxylysine residues per peptide as determined from the acid hydrolysate of peptide B2 (Tc9). It was concluded therefore that 0.08 GH residues occur at position 756 in peptide α1−CB7.

The sum totals of these values obtained from analysis of tryptic peptides, of 0.44 GGH, 0.70 GH and 0.69 FH residues per peptide α1−CB7 are close to the values obtained from analysis of the whole, uncleaved peptide α1−CB7, which were 0.40 GGH, 0.57 GH and 0.78 FH residues per peptide (Table 3.6).

In a summary at the end of this chapter, data regarding hydroxylation and glycosylation of lysyl positions in peptide α1−CB7 as obtained by analysis of these tryptic peptides will be compiled together with the forthcoming data obtained from analysis of V8−protease peptides.

II. V8−PROTEASE PEPTIDES OF PEPTIDE α1−CB7

In the last section, we saw how all but three of the possible sites of hydroxylysine glycosides in peptide α1−CB7 were characterised separately within their respective tryptic peptides of citraconylated α1−CB7.

Although analysis of peptide D1 (Tc6) revealed the total amount of hydroxylysine glycoside distributed between the remaining three possible sites at positions 648, 657 and 684, the amounts of glycosides occurring at these individual sites were still unknown. Thus, it was not known whether the 0.4 GH and 0.27 GGH residues found per peptide D1 (Tc6) occurred at a single site or spread between two or all three sites.

In order to characterise these possible sites individually, it was necessary to cleave peptide Tc6 or peptide α1−CB7 in a way which would isolate positions 648, 657 and 684 individually in three different peptides. Tryptic cleavage at these lysyl residues was ruled out for the same reason which necessitated the inhibition of cleavage at lysyl residues in the first place, which was that any cleavage at lysyl residues would inevitably complicate peptide separation and interpretation of results (see start of section I of this chapter). Chymotryptic cleavage at Phe−635 would have been of little use as this does not separate any of the three lysyl residues in question (see fig. 4.1).

A solution to the problem at hand was the use of 'staphylococcal protease' otherwise known as V8−protease, a protease from *Staphylococcus aureus*, strain V8. This enzyme catalyses the hydrolysis of peptide bonds on the carboxyl side of glutamic acid and aspartic acid residues (Drapeau et al., 1972; Houmard and Drapeau, 1972) except where the bonds involve proline (Sutton et al., 1977). Peptide bonds involving hydroxyproline are also resistant to cleavage, as has been shown during the use of

V8−protease to cleave the α1(III) chain (Fietzek et al., 1979) and peptide α1(I)−CB8 (Glanville et al., 1983) from calf skin for primary sequence analysis. During the sequencing of part of the human α1(IV) chain, it was indicated that a Glu−Hyl bond between positions *431* and *432* (in the 95−kdalton pepsin fragment) was cleaved by V8−protease although the hydroxylysine residue was fully glycosylated (Babel and Glanville, 1984). However, it was not clear whether any such cleavage occurred at a similar bond involving a hydroxylysine glycoside, between positions *631* and *632* of the same α1(IV) fragment.

The proteolytic activity of V8−protease also involves a dual specificity. When the reaction is carried out in ammonium bicarbonate buffer, pH 7.8, cleavage is achieved only at glutamic acid residues while cleavage in sodium or potassium phosphate buffer, pH 7.8, facilitates the cleavage at both glutamic acid and aspartic acid residues (Drapeau, 1977).

Of the ten aspartic acid residues contained in peptide α1−CB7, those occurring at positions 653 and 659 conveniently interrupt the sequences between the lysyl residues at positions 648, 657 and 684, the three possible sites of hydroxylysine glycosides yet to be characterised separately (fig. 4.1). A strategy was developed by which the possible glycoside sites at positions 648, 657 and 684 can be isolated and characterised separately using the specifities of V8−protease and the favourable positions of the aspartic acid residues at positions 653 and 659, and of the eleven glutamic acid residues in peptide α1−CB7.

II A. STRATEGY FOR THE CLEAVAGE OF PEPTIDE α1−CB7 WITH V8−PROTEASE

Peptide α1−CB7 would first be cleaved with V8−protease in 100mM ammonium bicarbonate/2mM EDTA pH 7.8 (see Chapter 2.IIC) in order to achieve cleavage solely at glutamic acid residues occurring at positions 554, 605, 726, 734, 743, 755, 791 and 815. In theory this would liberate nine V8−protease peptides of α1−CB7, assigned 'P1' to 'P9', as predicted from the sequence of peptide α1−CB7 from calf skin (fig.

4.6). Cleavages were not expected at glutamic acid residues 626, 650 and 803 as these residues are followed by hydroxyproline in the sequence of peptide α1−CB7. The largest theoretical peptide, P3, containing the possible glycoside sites at positions 648, 657 and 684 plus the 'post−Gly' lysyl residue at position 725, would be separated out from the digest by gel chromatography on Sephadex G−50_S because, being 121 residues long, it is more than twice as large as the next largest theoretical peptide, P2, which is 51 residues long (fig. 4.6).

Peptide P3 would then be digested further with V8−protease, this time in 50mM potassium phosphate buffer/2mM EDTA, pH 7.8 (see Chapter 2.IIC) in order to achieve cleavages at aspartic acid residues at positions 623, 642, 653 and 659. In theory, this would liberate five V8−protease peptides of P3, assigned 'P3a' to 'P3e' (fig. 4.12). Isolation of these peptides and subsequent characterisation of peptides P3c, P3d and P3e would ultimately ascertain the distribution of hydroxylysine glycosides between positions 648, 657 and 684.

Meanwhile, peptides P1, P2 and P4 to P9 (see fig. 4.6) would also be isolated and characterised in order to verify the results obtained from the analysis of tryptic peptides previously described.

II B. ISOLATION OF V8−PROTEASE PEPTIDES OF α1−CB7

Peptide α1−CB7 of bovine corneal collagen was prepared as described in chapters 2 and 3. V8−protease cleavage of peptide α1−CB7 was carried out in 100mM ammonium bicarbonate buffer (2mM EDTA) pH 7.8. After isolation from the V8−protease digest, putative peptide P3 was cleaved with V8−protease in 50mM potassium phosphate buffer (2mM EDTA) pH 7.8. These digestions with V8−protease were first performed starting with a small batch of peptide α1−CB7 essentially to assess the extent of cleavage achieved and the capabilities of the fractionation procedures in separating the resultant peptides before using a larger, main batch of peptide α1−CB7. The exploratory digestions will be described first.

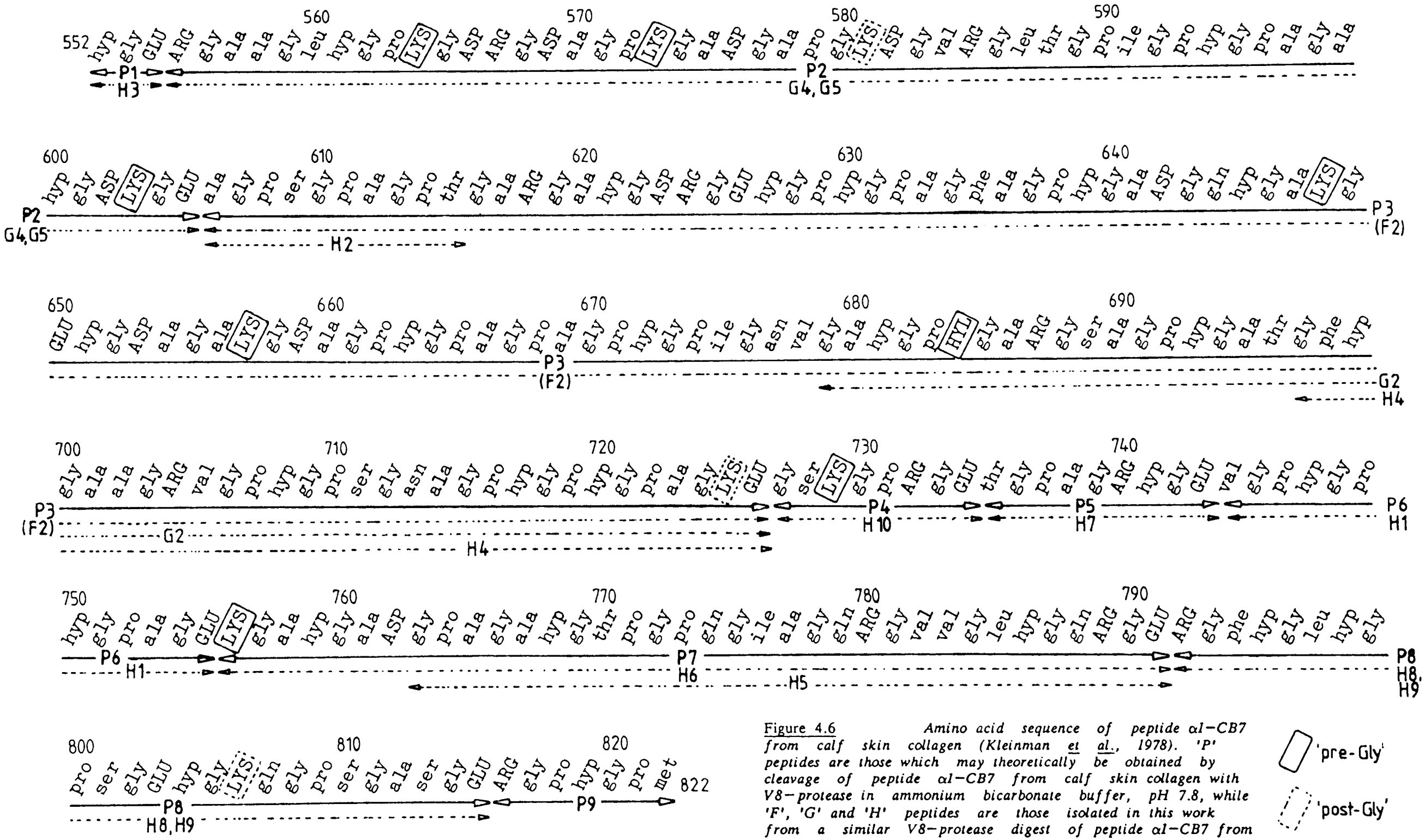


Figure 4.6 *Amino acid sequence of peptide α1−CB7 from calf skin collagen (Kleinman et al., 1978). 'P' peptides are those which may theoretically be obtained by cleavage of peptide α1−CB7 from calf skin collagen with V8−protease in ammonium bicarbonate buffer, pH 7.8, while 'F', 'G' and 'H' peptides are those isolated in this work from a similar V8−protease digest of peptide α1−CB7 from bovine corneal collagen. 'F2' is shown in brackets as the amino acid composition of fraction F2 indicated that it was not entirely homogeneous. It was treated as 'putative' peptide P3 which was subsequently redigested with V8−protease (see fig. 4.12).*

1. Exploratory digestions with V8−protease

After digestion of a small batch of peptide α1−CB7 (28mg) with V8−protease in ammonium bicarbonate buffer, the digest was fractionated on Sephadex G−50_S (100 × 2.3 cm) in 30mM ammonium acetate buffer, pH 4.8, at 37°C, and eluted at a flow rate of 10 ml per hour. The chromatogram obtained for this fractionation is shown in figure 4.7.

Fraction 'a', expected to consist mainly of the largest, 121−residue peptide P3, was subjected to amino acid analysis. The amino acid composition of fraction 'a' was similar but not identical to that of theoretical peptide P3 (Table 4.4). The differences lie, for example, in the high glutamate and aspartate contents, plus the presence of leucine in fraction 'a' when compared with the composition of P3. A small quantity of V8−protease was hydrolysed in 6M HCl and subjected to amino acid analysis, giving the composition shown in Table 4.4. The amino acid composition of V8−protease showed high contents of glutamate and aspartate, implying therefore that theoretical peptide P3 in fraction 'a' was contaminated with V8−protease. Bearing in mind its estimated length of 115 residues and an estimated molecular weight of 11.4 to 12 kilodaltons (Drapeau et al., 1972), it was most likely that V8−protease, having undergone little or no autodigestion, had eluted from the column in fraction 'a' together with peptide P3 which is 121 residues long.

Analysis of the other pooled fractions gave compositions (not shown) which did not correspond entirely to those of the expected peptides P1 to P9, but were typical of collagen peptides, with about 33% glycine and high imino acid contents. Fraction 'b', expected to contain small peptides due to its elution position, was the only fraction whose amino acid composition contained homoserine, the CNBr digestion product of methionine which occurs in peptide P9 (see fig. 4.6), the theoretical carboxy−terminal heptapeptide of the V8−protease digest of peptide α1−CB7. It was concluded therefore, that the conditions employed during the incubation of peptide α1−CB7 with V8−protease in ammonium bicarbonate had resulted in cleavages which were most probably at the expected glutamic acid residues. However, it would

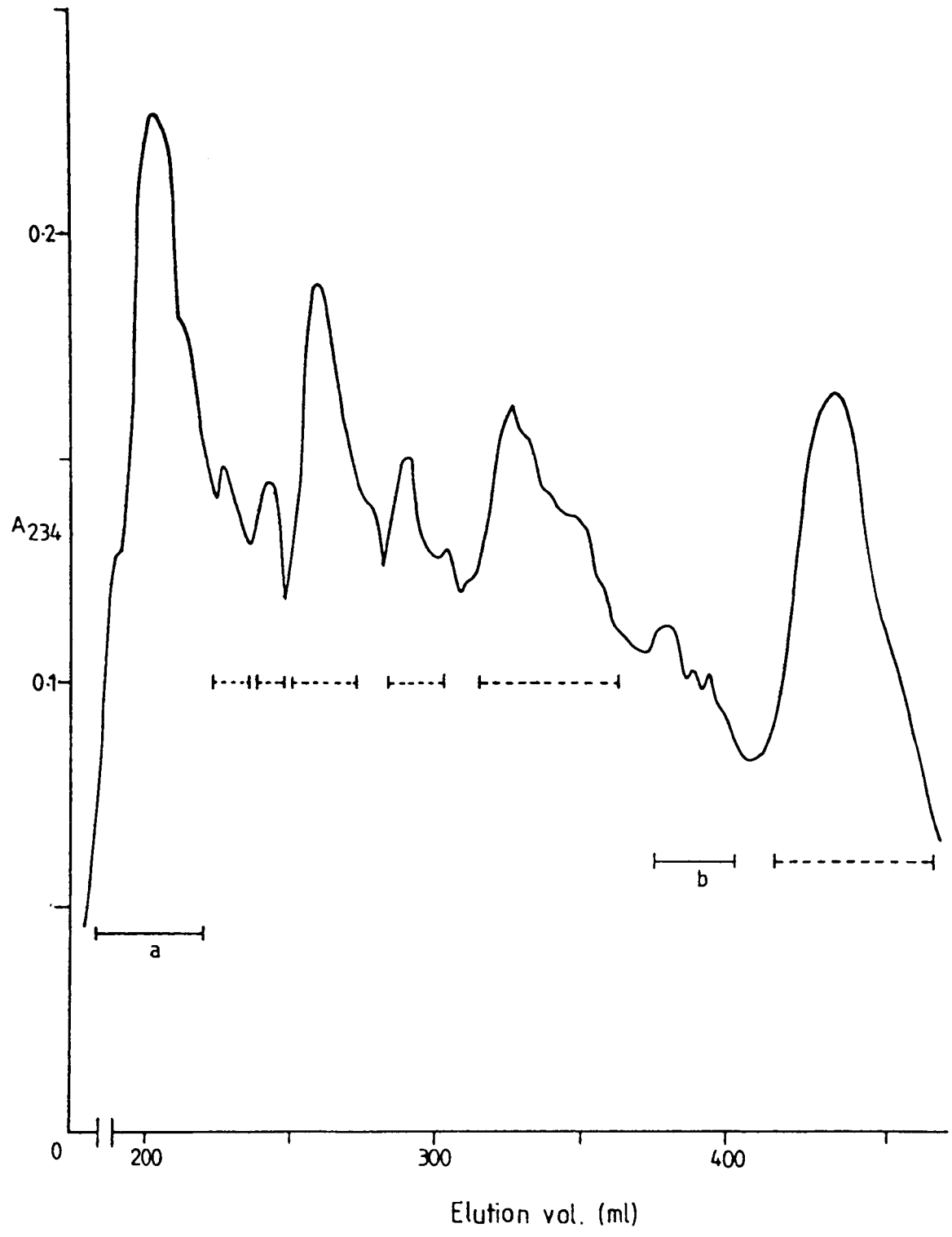


Figure 4.7 Gel chromatography on Sephadex G−50$_S$, of products from an 'exploratory' V8−protease cleavage (see text) of peptide α1−CB7 in ammonium bicarbonate buffer pH 7.8. Details of the chromatography were as indicated for fractionation 'D' (fig. 4.4).Bars indicate the fractions pooled, the broken bars indicating those not mentioned by name in the text.

Table 4.4

Amino acid composition of fraction 'a' (see fig. 4.7) obtained from an exploratory V8−protease digest of peptide α1−CB7 shown alongside that calculated for theoretical peptide P3 from the known sequence and that obtained for V8−protease. Values are expressed as residues per 100 amino acid residues. '0' signifies a value of less than 0.2.

	fraction 'a'	P3	V8-protease
Hydroxyproline	10.4	11.6	0
Aspartic acid	6.4	5.0	21.8
Threonine	1.1	1.7	7.8
Serine	2.2	2.5	4.8
Glutamic acid	5.2	3.3	10.8
Proline	14.1	14.9	9.1
Glycine	32.6	33.1	8.2
Alanine	15.6	17.4	5.9
Valine	2.2	1.7	4.8
Methionine	0	0	1.04
Isoleucine	1.5	0.83	6.4
Leucine	0.4	0	3.7
Tyrosine	0	0	3.5
Phenylalanine	1.1	1.7	3.5
Hydroxylysine	1.2	0.83	0
Histidine	0	0	2.8
Lysine	3.4	2.48	4.8
Arginine	2.7	3.3	1.0

seem necessary to employ fractionation methods of higher resolution such as ion−exchange chromatography on AG50W−X4 resin if the resultant peptides were to be isolated at acceptable levels of purity.

Fraction 'a' (5.2 mg) was then digested further with V8−protease, this time in 50 mM phosphate buffer pH 7.8 in order to facilitate cleavage at aspartic acid residues at positions 623, 642, 653 and 659 of peptide P3 contained in fraction 'a'.

Fractionation of this V8−protease digest on Sephadex $G-50_S$ (100 × 2.3 cm) gave the chromatogram shown in figure 4.8. The occurrence of several peaks at elution volumes different to that of fraction 'a' on the same chromatography column (fig. 4.7) indicated that fraction 'a' had been cleaved by V8−protease under the conditions employed. On freeze−drying and amino acid analysis, fractions 'c' and 'd' gave amino acid compositions (not shown) with high glutamate and aspartate contents, not unlike that of V8−protease (Table 4.4). Fraction 'e' gave an amino acid composition that was not unlike that of the 67−residue long theoretical peptide P3e (Table 4.5), the largest of the theoretical peptides P3a to P3e (see fig. 4.12). The values for glutamate and aspartate were however, too high for fraction 'e' to be considered as peptide P3e in sufficiently pure form for further analysis. The other pooled fractions gave amino acid compositions which indicated that they were mixtures of (collagen) peptides. These compositions are not presented due to their relatively inconclusive nature.

However, it was decided that, as the amino acid composition of fraction 'e' was not unlike that of peptide P3e, it was likely that cleavage had been achieved at least at some of the aspartic residues in the putative peptide P3, and that the remaining fractions were likely to contain the expected peptides P3a to P3d. The possibility that non−specific digestions may have occurred at residues other than aspartic acid could not however be ruled out.

Rather than trying to purify individual V8−protease peptides from mixtures such as fractions 'b' and 'e' for analysis, effort was instead channelled towards preparing a larger batch of peptide α1−CB7 for cleavage with V8−protease. A decision was made

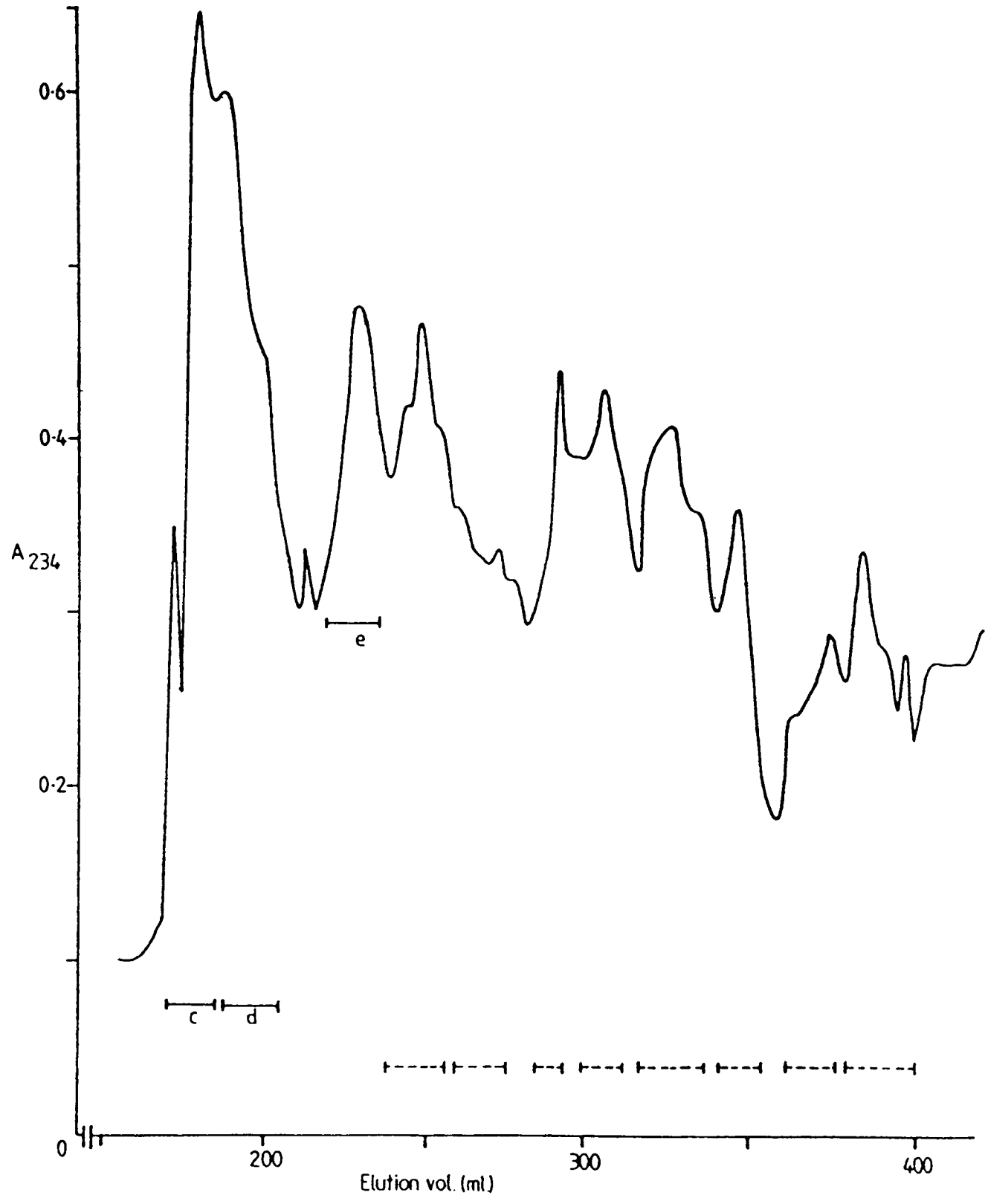


Figure 4.8 Gel chromatography , on Sephadex G−50$_S$, of products from an 'exploratory' V8−protease cleavage (see text) of fraction '*a*' (fig. 4.7) in potassium phosphate buffer pH 7.8. Details of chromatography were as indicated for fractionation 'D' (fig. 4.4) except that the flow−rate was 9ml/hr and the fractions collected were 2.7ml each. Bars indicate the fractions pooled, the broken bars indicating those not mentioned by name in the text.

Table 4.5

Amino acid composition of fraction 'e' (fig. 4.8) obtained from an exploratory V8−protease digest of fraction 'a' (fig. 4.7) alongside that of theoretical peptide P3e (see fig. 4.12). Values are expressed as residues per 100 amino acid residues.

	Fraction 'e'	P3e
Hydroxyproline	13.0	11.9
Aspartic acid	7.4	3.0
Threonine	2.2	1.5
Serine	3.4	3.0
Glutamic acid	4.7	1.5
Proline	13.6	17.9
Glycine	29.2	32.8
Alanine	13.3	16.4
Valine	2.6	3.0
Isoleucine	2.1	1.5
Leucine	0.3	0
Tyrosine	0.3	0
Phenylalanine	2.0	1.5
Hydroxylysine	1.4	1.5
Lysine	2.2	1.5
Arginine	2.3	3.0

to continue using the same methods and conditions as those employed in these exploratory digestions for digestion of the subsequent batch except that the ensuing fractionations were to be improved.

2. V8−protease peptides of peptide α1−CB7 :

Cleavage in ammonium bicarbonate buffer

Peptide α1−CB7 (71mg) was digested with V8−protease in 100mM ammonium bicarbonate buffer (2mM EDTA), pH 7.8 in order to facilitate cleavage of peptide bonds on the carboxyl side of glutamic acid residues. The digest was fractionated on Sephadex G−50_S, 128 × 1.4 cm, at 37°C and eluting at a flow−rate of 3.9 ml per hour. In comparison to the fractionation of the previous, 'exploratory' digest, the column was thus longer and the flow−rate, slower, in order to optimise the separation of peptides. Also, drawing on the lessons learnt earlier, smaller volumes (1.9 ml) were collected per test−tube in the fraction−collector and these were pooled in a stringent manner in order to avoid contamination of the early collagen peptide fractions with V8−protease. The chromatogram for this fractionation, 'F', is shown in figure 4.9.

Amino acid analysis of fraction F1 (2.6 mg) gave a composition (not shown) with high glutamate and aspartate contents, reminiscent of V8−protease (see Table 4.4). As only 2.4 mg of V8−protease was used in the digestion, most, if not all of the protease can be said to have been separated out from the digestion products of peptide α1−CB7 present in fraction F2 onwards (fig. 4.9).

Fraction F2 (11.5 mg) gave an amino acid composition similar but not identical to that of theoretical peptide P3 (Table 4.6), the expected 121−residue peptide containing the lysyl residues of interest at positions 648, 657 and 684 (see fig. 4.6). There was enough similarity between the two compositions to indicate that fraction F2 represented most of the sequence of peptide P3 and thus could be used as starting material for the subsequent digestion by V8−protease in potassium phosphate buffer. The high values for glutamate, aspartate and leucine in fraction F2 (Table 4.6) may indicate slight contamination by the protease. However, as fraction F2 was to be set

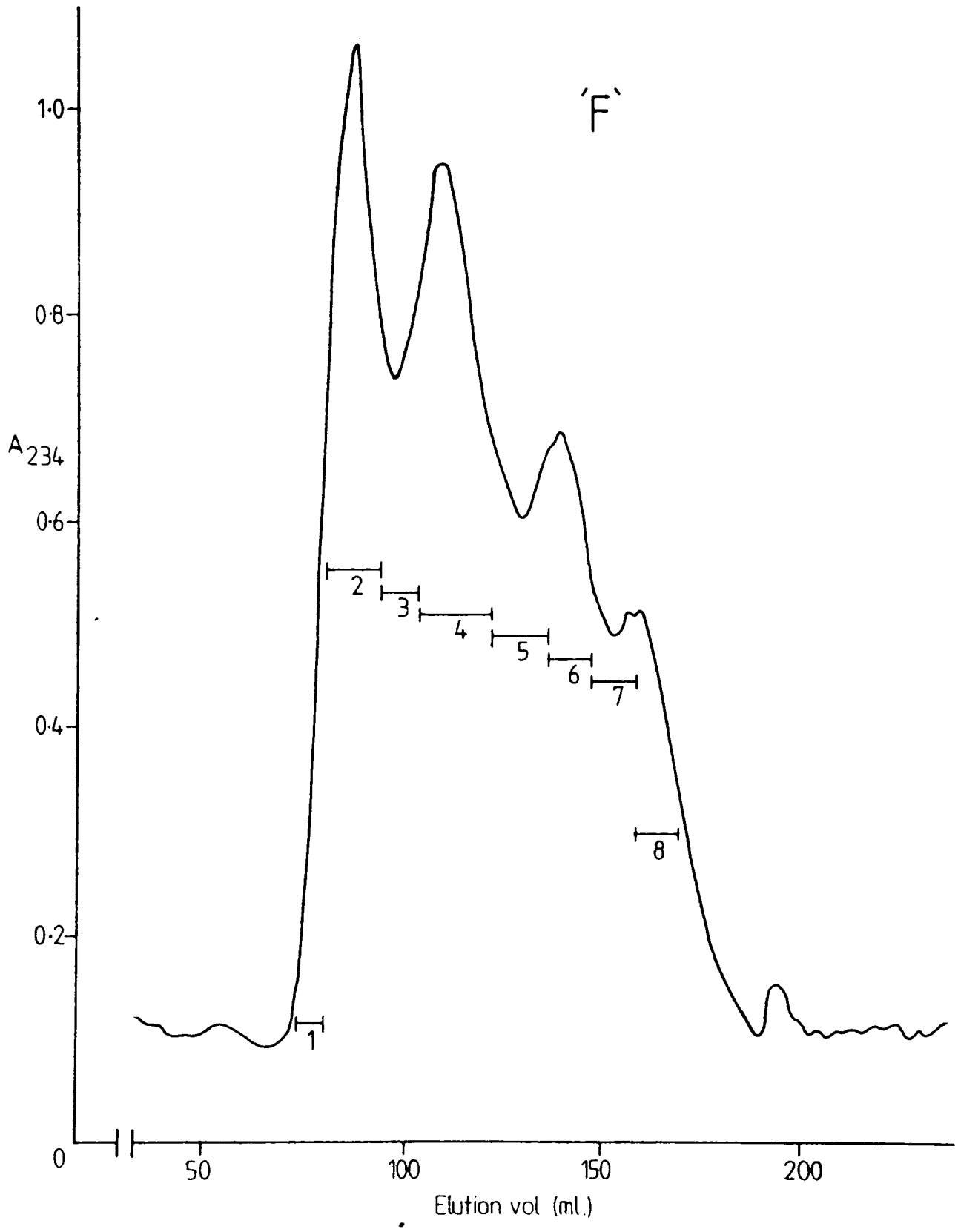


Figure 4.9 Fractionation 'F': Gel chromatography, on Sephadex G−50$_s$, of a V8−protease digest of peptide α1−CB7 (71mg) from bovine corneal collagen. The column (128 × 1.4 cm) was kept at 37°C and eluted with 30mM ammonium acetate pH 4.8 . Flow−rate: 3.9ml/hr. Tube fractions: 1.9ml. Bars indicate the fractions pooled.

Table 4.6

Amino acid composition of fraction F2 (fig. 4.9) alongside that of theoretical V8−protease peptide P3 (fig. 4.6). Values are expressed as residues per peptide.

	fraction F2	P3
Hydroxyproline	16.2	14
Aspartic acid	8.1	6
Threonine	2.2	2
Serine	2.9	3
Glutamic acid	5.4	4
Proline	17.1	18
Glycine	34.1	40
Alanine	19.8	21
Valine	2.3	2
Isoleucine	1.5	1
Leucine	0.3	0
Phenylalanine	2.2	2
Hydroxylysine	1.4	1
Lysine	3.2	3
Arginine	4.3	4
total	121.0	121

aside as putative peptide P3 for a subsequent digestion with the same protease, there was little point in trying to remove this contamination. The overall yield of fraction F2 (putative peptide P3) was 37%.

Amino acid analysis of fractions F3 to F8 (compositions not shown) indicated that they consisted of collagen peptides by their glycine and imino acid contents. The compositions did not match those of the expected peptides P1 to P9, indicating that they were mixtures.

Fractions F3 and F4, expected to contain the larger V8−protease peptides of peptide α1−CB7 due to their early elution positions (fig. 4.9) were combined (total of 34mg) and fractionated on the high resolution AG50W−X4 ion−exchange resin (100 × 1 cm) as described in chapter 2. Similarly, fractions F5 to F8, expected to contain the smaller peptides, were also combined (total of 31mg) and fractionated on AG50W−X4 resin. These fractionations, 'G' and 'H' respectively, are now discussed in succession.

Fractionation of combined F3/F4 gave the chromatogram shown in figure 4.10. The amino acid compositions of fractions G4 and G5 both matched that of peptide P2 (Table 4.7; fig. 4.6), the second largest, 51−residue peptide expected from the V8−protease digest except that G4 and G5 contained 0.41 and 0.62 residues of hydroxylysine respectively per peptide. Reciprocal decreases in their lysine contents account for the presence of hydroxylysine. Peptide P2 contains three 'pre−Gly' lysyl residues at positions 564, 573 and 603 which were determined as being hydroxylated to various extents from the previous analyses of tryptic peptide fractions C5/C7, C6 and C1 respectively. Fractions G4 and G5 most probably represent peptide P2 with different permutations of lysyl hydroxylation which separate during the ion−exchange chromatography on AG50W−X4 resin. This phenomenon was encountered before with respect to tryptic peptide fractions C5 and C7 which turned out to be different only in the extent of lysyl hydroxylation (and glycosylation). The overall yields obtained for fractions G4 and G5 were 16% and 12% respectively, the total overall yield of P2 therefore being 28%. From the yields of G4 and G5, peptide P2 therefore contains 0.5 residues of hydroxylysine distributed between positions 564, 573 and 603. This

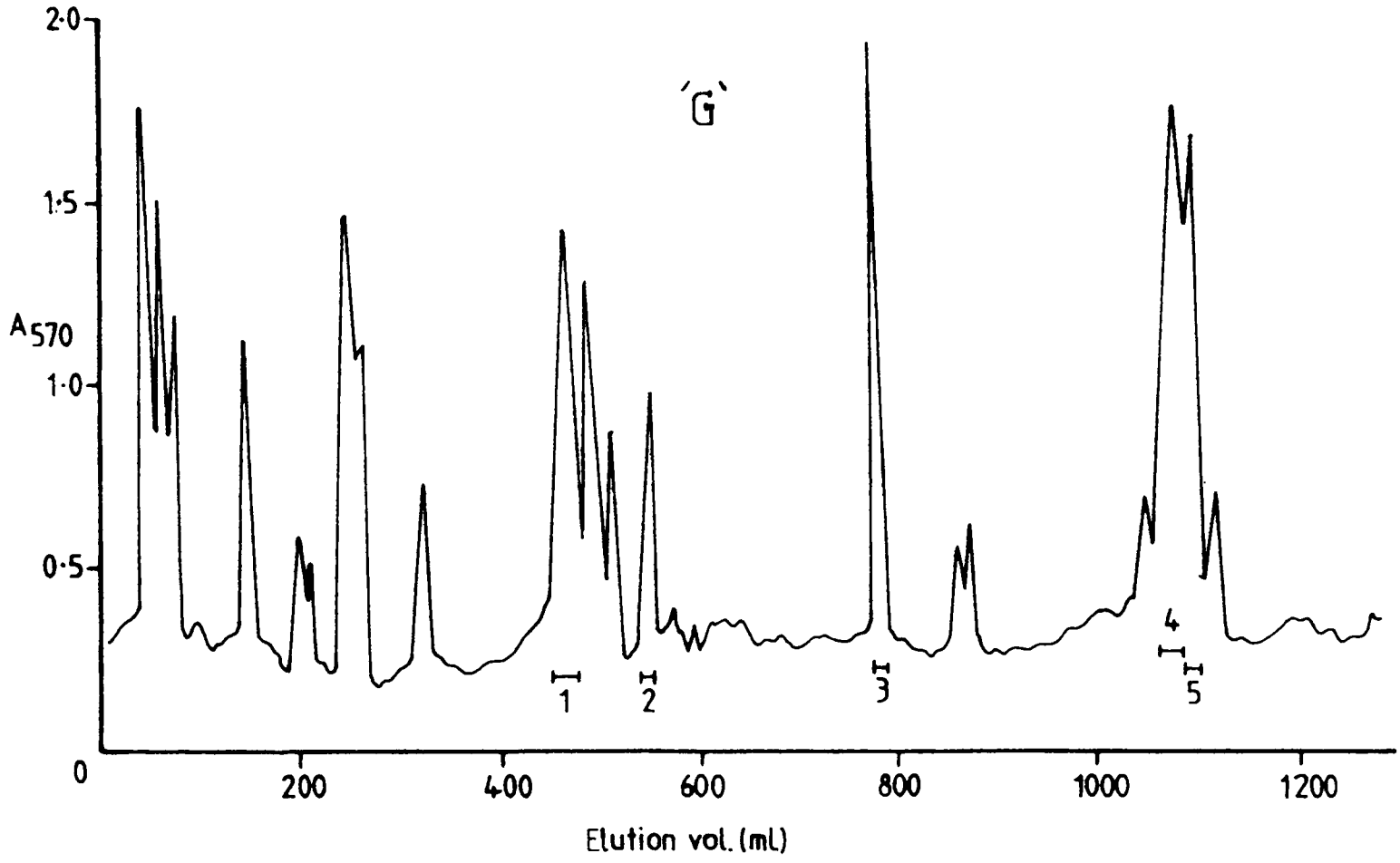


Figure 4.10 Fractionation 'G': Ion−exchange chromatography, on AG50W−X4 resin, of V8−protease peptides of peptide α1−CB7 contained in the combined fractions *F3* and *F4* (see fig. 4.9). For details of chromatography, see 'Methods' (Ch 2.IID1). Bars indicate some of the fractions pooled.

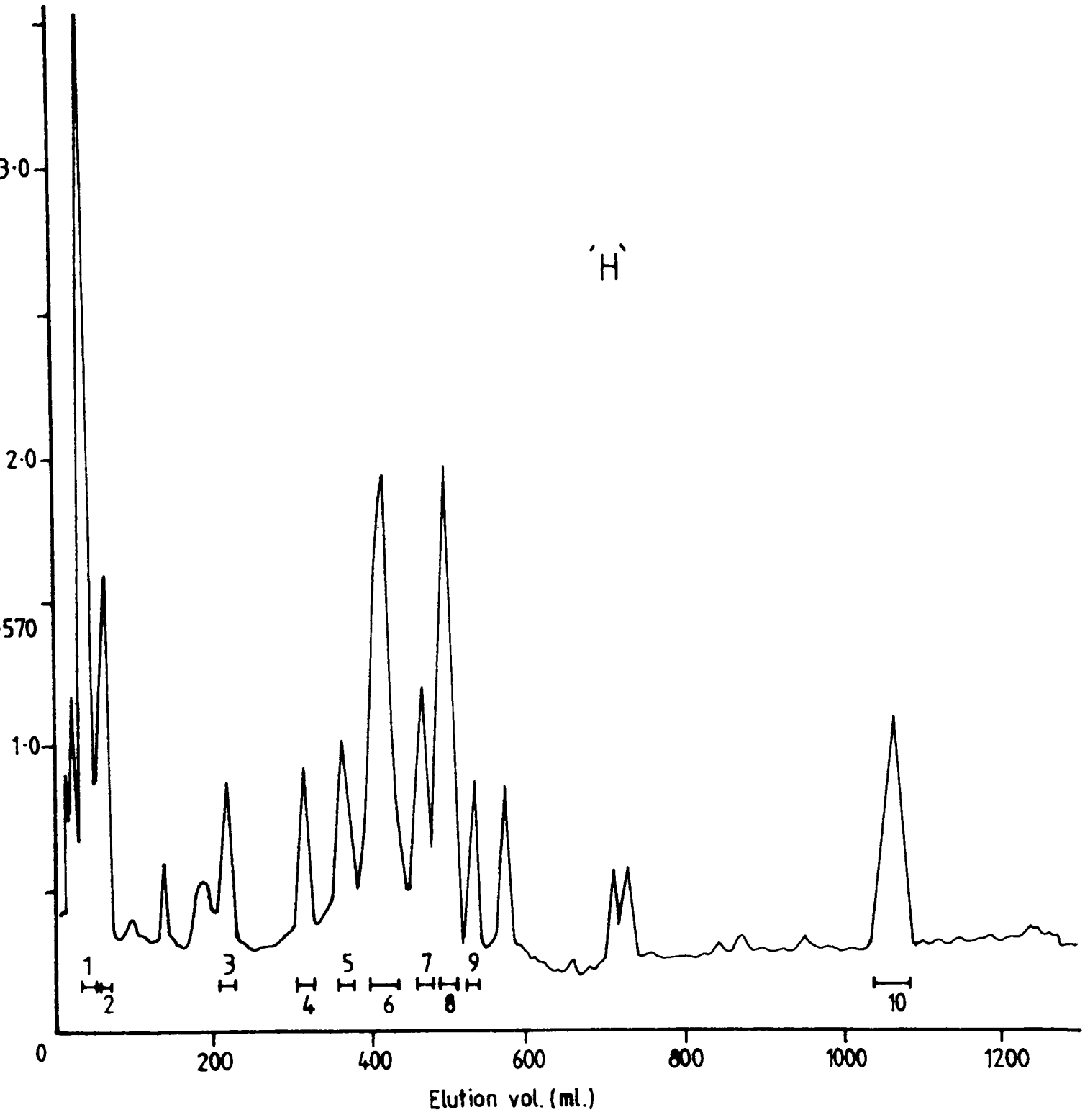


Figure 4.11 Fractionation 'H': Ion−exchange chromatography, on AG50W−X4 resin of V8−protease peptides of peptide α1−CB7 from bovine corneal collagen contained in the combined fractions *F5, F6, F7* and *F8* (see fig. 4.9). The ion−exchange column and chromatography procedures were as indicated in 'Methods' (Ch 2.IID1). Bars indicate some of the fractions pooled.

Table 4.7

Amino acid compositions of V8-protease peptides of peptide α1-CB7 from bovine corneal collagen isolated in fractionations 'G' and 'H' (figs. 4.10, 4.11) alongside those of the theoretical peptides predicted from the known sequence (fig. 4.6). Values are expressed as residues per peptide. '0' signifies a value of less than 0.2. This lower limit is set at 0.1 for hydroxylysine.

	H3	*P1*	C4	C5	*P2*	H2	↓	C2	↓	H4	↓	H10	*P4*	H7	*P5*	H1	*P6*	H6	*P7*	H5	↓	H8	H9	*P8*
Hydroxyproline	0.9	1	2.9	3.4	3	0.4	0	6.0	6	3.9	4	0	0	1.1	1	1.9	2	2.3	3	2.4	2	2.1	1.9	3
Aspartic acid	0	0	4.9	5.3	5	0	0	1.2	1	1.1	1	0	0	0	0	0	0	1.2	1	0.3	0	0	0	0
Threonine	0	0	1.1	0.8	1	0.8	1	1.2	1	0	0	0	0	0.8	1	0	0	0.9	1	0.8	1	0	0	0
Serine	0	0	0	0	0	0.8	1	2.0	2	1.0	1	0.8	1	0	0	0	0	0.2	0	0.3	0	2.9	2.7	3
Glutamic acid	0.8	1	1.2	0.9	1	0	0	1.0	1	1.3	1	1.0	1	0.8	1	0.9	1	3.8	4	3.9	4	2.9	2.8	3
Proline	0.2	0	6.5	6.1	6	2.7	3	6.6	7	4.5	5	1.1	1	1.1	1	2.9	3	3.4	3	3.1	3	2.8	2.8	2
Glycine	1.0	1	15.9	16.4	17	3.2	3	15.8	16	9.7	10	3.0	3	3.1	3	4.1	4	12.2	12	9.8	10	8.2	8.1	8
Alanine	0	0	7.2	6.9	7	2.1	2	7.8	8	4.2	4	0.2	0	1.1	1	1.3	1	5.1	5	3.1	3	1.2	1.2	1
Valine	0	0	1.0	1.2	1	0	0	1.1	1	0.8	1	0	0	0	0	0.9	1	1.8	2	1.8	2	0	0	0
Isoleucine	0	0	1.0	1.0	1	0	0	0	0	0	0	0	0	0	0	0	0	1.0	1	0.8	1	0	0	0
Leucine	0	0	2.1	2.2	2	0	0	0	0	0	0	0	0	0	0	0	0	1.1	1	0.8	1	1.0	0.9	1
Phenylalanine	0	0	0	0	0	0	0	0.9	1	0.8	1	0	0	0	0	0	0	0	0	0	0	1.0	1.1	1
Hydroxylysine	0	0	0.41	0.62	0	0	0	1.1	1	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0
Lysine	0	0	3.6	3.43	4	0	0	1.1	1	1.3	1	0.9	1	0	0	0	0	0.88	1	0	0	0.9	1.2	1
Arginine	0	0	3.1	2.9	3	0	0	1.9	2	1.3	1	1.0	1	1.0	1	0	0	2.0	2	1.8	2	1.0	1.3	1
total	2.9	3	50.91	51.15	51	10.0	10	47.7	48	29.9	30	8.0	8	9.0	9	12.0	12	36.05	36	28.9	29	24.0	24.0	24

value is close to the sum total of 0.62 hydroxylysine residues for these lysyl positions, as determined previously in the tryptic peptides C5/C7 (Tc2), C6 (Tc3) and C1 (Tc4) as indicated in Table 4.3.

Amino acid analysis of fraction G2 gave an amino acid composition (Table 4.7) indicating it to be the peptide sequence 679−726 (see fig. 4.6). Most of this peptide sequence was expected to be in putative peptide P3 (F2) which had been set aside (see beginning of this section) for further digestion with V8−protease. The low overall yield of 8% for this peptide indicated that the unexpected cleavage on the carboxyl side of Val−678 which liberated this peptide was only a partial split. Nevertheless, peptide G2, which contains the 'pre−Gly' lysyl residue Hyl−684 and the 'post−Gly' Lys−725, contained 1.1 residues of hydroxylysine and 1.1 residues of lysine per peptide. It had been shown previously in tryptic peptide C4 (Tc8) that Lys−725 is not hydroxylated. This was expected, as Lys−725 is a 'post−Gly' residue (see fig. 4.6). By deduction, position 684 in peptide α1−CB7 of bovine corneal collagen is occupied by a fully hydroxylated lysine residue. This lysyl residue is only hydroxylated to the extent of 65−70% in calf skin (Fietzek et al., 1973) and 70% in chick skin (Highberger et al., 1975).

Of the other fractions pooled from fractionation 'G' (fig. 4.10), fraction G1 gave an amino acid composition indicating that it was a mixture containing the 36−residue peptide P7 while fraction G3 gave a composition reminiscent of regions from peptide P2. The other 'G' fractions also consisted of mixtures of collagen by nature of their glycine and imino acid contents. These amino acid compositions are not presented due to their inconclusive nature.

Meanwhile, ion−exchange chromatography of the combined fractions F5, F6, F7 and F8 (see fig. 4.9) which were expected to contain the smaller V8−protease peptides, gave the chromatogram, 'H', shown in figure 4.11.

The amino acid composition of fraction H3 matched that of the tripeptide P1 occuring at the N−terminus of peptide α1−CB7 (Table 4.7; fig. 4.6). Peptide H3 (P1), whose overall yield was 44%, did not contain lysine. Of interest is the order in

which the single residues of hydroxyproline, glutamate and glycine occur in bovine corneal collagen. Firstly, the liberation of this tripeptide by V8−protease assigns the glutamate as being a glutamic acid residue (assuming cleavage does not occur at glutamine residues) at the carboxyl−terminal of the peptide. Secondly, since (4)−hydroxyproline never occurs in 'post−Gly' positions, at least in vertebrate collagens, the tripeptide is thus sequenced 'Hyp−Gly−Glu'. The hydroxyproline, occurring at the N−terminal position 552 (fig. 4.6) is therefore presumably that which is detected during N−terminal determination of peptide α1−CB7 by the dansylation method (see Chapter 2.IIF; Chapter 3.IIB).

Amino acid analysis of fraction H10 gave a composition which matched that of the 8−residue peptide P4 (Table 4.7; fig. 4.6), the predominance of basic amino acids causing this peptide to elute late in the chromatography (fig. 4.11). The 0.2 residues of alanine per peptide probably arose from a contaminant. Peptide H10 (P4) contains a 'pre−Gly' lysyl residue at position 729 in the sequence for peptide α1−CB7. However, no hydroxylation of this lysyl residue was detected, agreeing with the results obtained for tryptic peptide C4 (Tc8) shown earlier in Table 4.1. The overall yield of peptide H10 (P4) was 74%.

The amino acid compositions of fractions H7 and H1 were those of the 9− and 12−residue peptides P5 and P6 respectively (Table 4.7; fig. 4.6). The early elution of H1 in the ion−exchange fractionation (fig. 4.11) was probably due to the presence of one glutamic acid residue and no basic amino acid residues in peptide H1 (P6). Neither peptide H7 (P5) nor peptide H1 (P6) contained lysine. Their overall yields were 69% and 74% respectively.

Amino acid analysis of fraction H6 gave a composition which matched that of the 36−residue peptide P7 of peptide α1−CB7 (Table 4.7; fig. 4.6), except that it contained 0.2 residues of serine which probably arose from a minor contaminant. The depressed value obtained for hydroxyproline was compensated by an increased value for proline. In addition, it also contained 0.17 residues of hydroxylysine per peptide, occurring therefore at position 756 in the sequence of peptide α1−CB7 (see fig. 4.6).

Previous analysis of tryptic peptide B2 (Tc9) which also contained position 756, indicated a comparable value of 0.11 residues of hydroxylysine per peptide (Table 4.1). The 1.0 residue of isoleucine per peptide found in peptide H6 (P7) occurs at position 776 in the calf skin sequence, which, together with Gly−775 presumably forms the cleavage site in the α1(I) chain for animal collagenase. From the known sequence of peptide α1−CB7 of calf skin, it can also be seen that peptide H6 (P7) would also contain the proposed fibronectin−binding site encompassed by residues 766−786 (fig. 4.6; see Kleinman et al., 1981a).

The overall yield of peptide H6 (P7), at 33%, was lower than those obtained so far for peptides from this fractionation, probably due to two reasons. The first was implied earlier in that fraction G1 was thought to be a mixture of peptides containing P7. Being the third largest V8−protease peptide at 36−residues, some of it was probably originally present in fraction F4 (fig. 4.9) which was subsequently re−fractionated in fractionation 'G'. The second reason for its relatively low yield compared to the other 'H' peptides was that another fraction, H5, was recovered and whose amino acid composition indicated it to be a partial peptide of peptide H6 (P7).

The amino acid composition of fraction H5 (Table 4.7) corresponds to that of the near complete sequence of peptide P7 (positions 756−791) being that of positions 763 to 791 (see fig. 4.6) except for 0.3 residues each of aspartate and serine per peptide which probably arose from contaminants. The missing seven residues at the N−terminal end of P7 were most probably liberated as an intact 7−residue peptide of positions 756−762 by the slow hydrolysis of the peptide bond formed between Asp−762 and Gly−763. The overall yield of peptide H5 (positions 763−791), a shortened form of peptide H6 (P7), was 12%. Including peptide H6, the total overall yield for positions 763−791 was therefore 45%. Peptide H5 does not contain lysine.

Amino acid analysis of fractions H8 and H9 showed that they corresponded to the 24−residue peptide P8 (Table 4.7; fig. 4.6), eluting as two adjacent peaks during the chromatography (fig. 4.11). It is not known what caused peptide P8 to elute as two peaks although partial deamidation of glutamine at position 807 was suggested as a

possible cause (J.J. Harding, *personal communication*). The low hydroxyproline contents of H8/H9 relative to the calf skin sequence were compensated for by the correspondingly high proline contents. Fractions H8 and H9 contained 0.9 and 1.2 residues of lysine per peptide, with no hydroxylysine detected in either. This was expected, due to the 'post−Gly' position of Lys−806, which was also determined as not being hydroxylated from the analysis of tryptic peptide C3 (Tc12) described earlier (Table 4.1). The overall yields of fractions H8 and H9 were 43% and 14% respectively, giving a total overall yield of 57% for peptide P8 (H8/H9).

V8−protease peptide P8, isolated as the fractions H8 and H9 described above, is of special interest when compared with tryptic peptide Tc12 which was isolated as fraction C3 (fig. 4.3[b]). It seems that totally different proteolytic cleavages of peptide α1−CB7, one with trypsin (after citraconylation) on the carboxyl side of arginine residues Arg−792, Arg−816 and another with V8−protease on the carboxyl side of glutamic acid residues Glu−791, Glu−815 have resulted in two 24−residue peptides with identical amino acid compositions (see Tables 4.1, 4.7). These two peptides, C3 (Tc12) and H8/H9 (P8), overlap over 23 of their 24 residues and differ only with respect to their N− and C− terminal residues with tryptic peptide C3 (Tc 12) having its only arginine residue (Arg−816) at the C−terminus while V8−protease peptide H8/H9 (P8) has its only arginine residue (Arg−792) at its N−terminus (see figs. 4.1, 4.6).

The digestion of peptide α1−CB7 by V8−protease in 100mM ammonium bicarbonate buffer (2mM EDTA) pH 7.8 was expected to cleave only next to glutamic acid and therefore to liberate only the nine peptides P1 to P9. However it has been shown to liberate peptides G2 and H5 which indicated, on inspection of the sequence of peptide α1−CB7, that unexpected partial cleavages had occurred on the carboxyl side of Val−678 and Asp−762 respectively (fig. 4.6). In addition to these partial cleavages at valine and aspartic acid, amino acid analysis of fractions H2 and H4 (fig. 4.11) indicated that partial cleavages had also occurred at two separate Thr−Gly bonds.

Amino acid analysis of fraction H2 gave a composition which indicated it to be the peptide of positions 606−615 in the sequence of peptide a1−CB7 (Table 4.7; fig. 4.6) which had resulted from an unexpected cleavage on the carboxyl side of threonine at position 615. Peptide H2 eluted early in fractionation 'H' presumably due to its lack of the basic amino acids lysine and arginine and its relatively small size. The overall yield of peptide H2 was 38%.

The amino acid composition of fraction H4 matched that of the peptide sequence 697−726 in the sequence of peptide α1−CB7 (Table 4.7; fig. 4.6). The peptide sequence 697−726 arises from an unexpected cleavage on the carboxyl side of the threonine residue at position 696. Fraction H4 was found to contain 1.3 residues of lysine per peptide. Hydroxylysine was not found in H4. This is consistent with the 'post−Gly' position of this lysyl residue at position 725 (fig. 4.6). The finding that peptide H4 did not contain hydroxylysine is consistent with the earlier finding that tryptic peptide C4 (Tc8), which also contains this 'post−Gly' Lys−725, did not contain hydroxylysine. The absence of hydroxylysine at position 725 demonstrates that the 1.1 hydroxylysine residues found in peptide G2 (Table 4.7; fig. 4.6) occurs as a single hydroxylysine residue at the 'pre−Gly' position 684, without relying solely on the common assumption that 'post−Gly' lysyl residues are never hydroxylated. The overall yield of fraction H4 was 14%.

All the peptides isolated from the V8−protease digestion of peptide α1−CB7 in conditions (ammonium bicarbonate buffer) that should promote specific cleavage after glutamic acid residues have now been described. The next section describes the isolation of peptides resulting from the digestion of putative peptide P3 (fraction F2) by V8−protease in potassium phosphate buffer. The peptides expected from this digestion would enable the individual characterisation of the possible sites of hydroxylysine glycosides at positions 648, 657 as well as 684. The determination of hydroxylysine glycosides at all the hydroxylated lysyl positions isolated in the relevant V8−protease peptides will be described after the next section.

3. V8−protease peptides of putative peptide P3:

Cleavage in potassium phosphate buffer.

It was mentioned earlier with respect to tryptic peptide D1 (Tc6) that it was still not possible at the time to ascertain the exact distribution of hydroxylysine glycosides found in peptide D1 between the 'pre−Gly' lysyl residues at positions 648, 657 and 684 (see Table 4.3; fig. 4.1). In order to characterise individually these possible glycoside sites, it was necessary to cleave between these positions. This is possible by cleaving the V8−protease peptide P3 with V8−protease, this time in conditions (potassium phosphate buffer) which would facilitate cleavage on the carboxyl side of aspartic acid residues at positions 653 and 659 (see fig. 4.6).

V8−protease cleavage at these, plus two other aspartic acid residues at positions 623 and 642 in peptide P3 was expected to liberate five theoretical peptides P3a, P3b, P3c, P3d and P3e, the first two with no lysine, the latter three containing the lysyl residues to be characterised, at positions 648, 657 and 684 respectively (see fig. 4.12). Theoretical peptide P3e also contains Lys−725 which however, due to its 'post−Gly' position, was not expected to be hydroxylated.

Fraction F2 (11.5mg), set aside as putative peptide P3 after fractionation 'F', was digested with V8−protease in 50 mM potassium phosphate buffer (2mM EDTA) pH 7.8 as described in chapter 2 (section IIC2) in order to achieve the desired cleavages at the four aspartyl residues. The digest was fractionated on Sephadex $G-50_S$ as performed in fractionation 'F' (fig. 4.9). The chromatogram for this fractionation, 'J', is shown in figure 4.13.

The amino acid compositions of fractions J1 and J2 (not shown) indicated that they contained V8−protease by virtue of their high glutamate and aspartate contents (see Table 4.4). The typically high glycine, proline and hydroxyproline contents in the amino acid compositions (not shown) of fractions J3 to J7 indicated that these peptides were collagen peptides. The amino acid compositions of fractions J3 to J7 did not correspond to any of the theoretical V8−protease peptides. Repeated amino acid analysis of fraction J8 showed very little trace of amino acids, suggesting that fraction

J8 was mainly 'salts' originating from the potassium phosphate buffer employed in the V8−protease digestion.

Of the theoretical peptides expected, the largest, 67−residue peptide P3e was expected to have eluted as a distinct peak in fractionation 'J', resolved from the other peptides P3a, P3b, P3c and P3d. This did not materialise in fractionation 'J', with the implication that, although the range of peaks in fractionation 'J' (fig. 4.13) indicated that V8−protease had cleaved fraction F2 (putative P3), some unexpected cleavages may have occurred, liberating peptides of a different range from those expected.

Fractions J3 to J7 were therefore combined for refractionation on the high resolution ion−exchange resin AG50W−X4 in order to isolate the cleavage products of putative peptide P3 (F2). Chromatography was performed as for fractionations 'G' and 'H' (figs. 4.10, 4.11). The chromatogram for this fractionation, 'K', is shown in figure 4.14.

Amino acid analysis of fraction K3 gave a composition indicating that it consisted of equimolar amounts of peptides P3b and P3c (Table 4.8; fig. 4.12). The 19−residue peptide P3b and the 11−residue peptide P3c each contain one basic amino acid and two acidic amino acid residues and thus may have had nearly identical charge properties. These two peptides may have co−eluted in fractionation 'K' for this reason. However, as peptide P3c immediately follows peptide P3b in the sequence of peptide α1−CB7, it is also possible that fraction K3 represents the uncleaved double−peptide P3b−P3c due to the 'Asp−Gly' bond at position 642−643 not being cleaved by V8−protease (see fig. 4.12). The amino acid composition of fraction K3 matched that of combined P3b and P3c except for a small quantity of serine which probably arose from a contaminant. Fraction K3 (P3b−c) contained 0.15 residues of hydroxylysine per peptide which, according to the sequence of calf skin peptide α1−CB7, resides at position 648. This was the first time in this work that this 'pre−Gly' lysyl position had been individually characterised. The yield of fraction K3 (P3b−c) with respect to the combined peptides P3b and P3c was 14% as calculated from the 71 mg of original starting material (peptide α1−CB7) or 39% as calculated

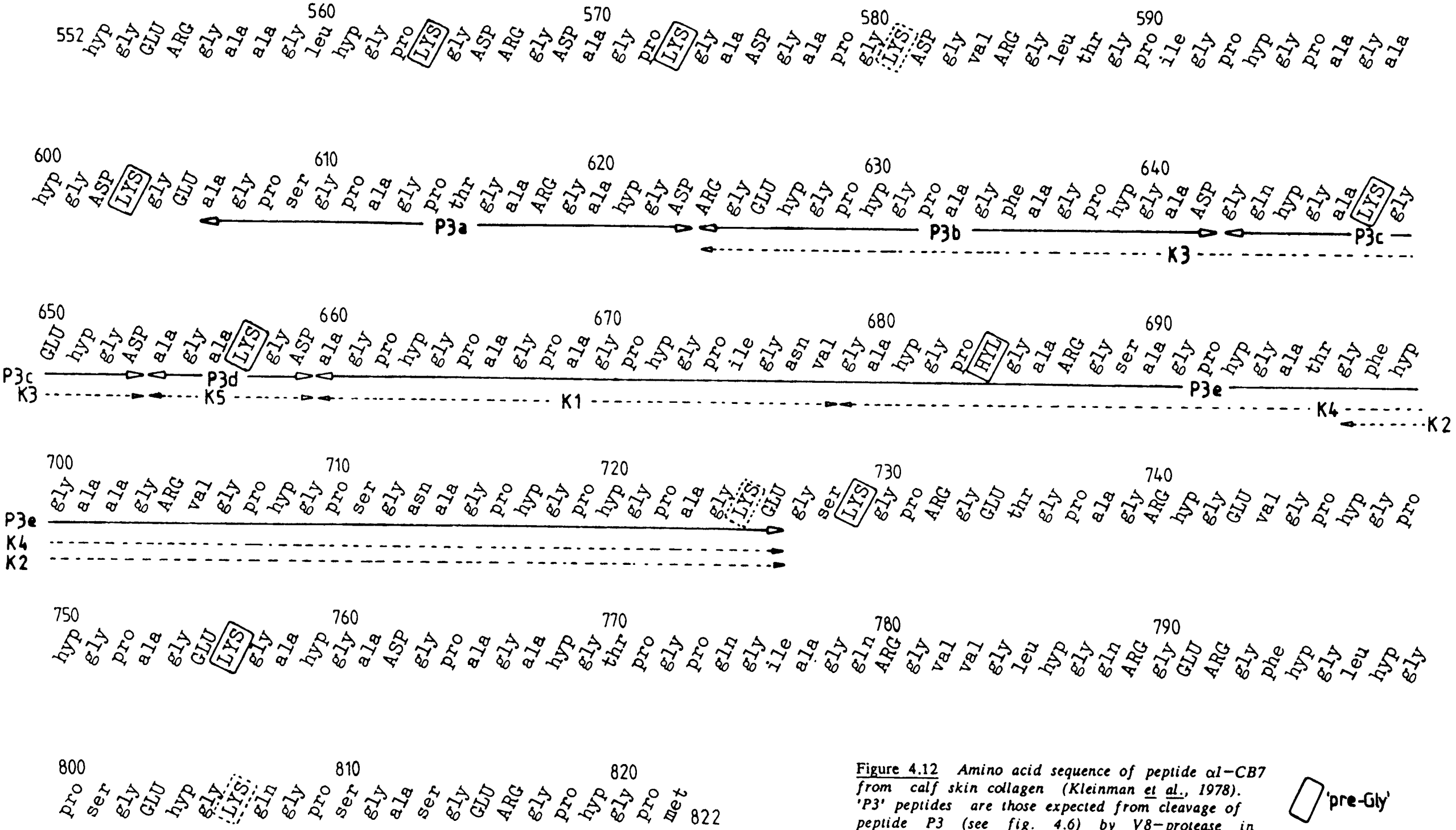


Figure 4.12 *Amino acid sequence of peptide α1–CB7 from calf skin collagen (Kleinman et al., 1978). 'P3' peptides are those expected from cleavage of peptide P3 (see fig. 4.6) by V8–protease in potassium phosphate buffer pH 7.8. 'K' peptides are those isolated in the present work from the V8–protease digest of fraction F2 (see fig. 4.9) which was set aside as putative peptide P3.*

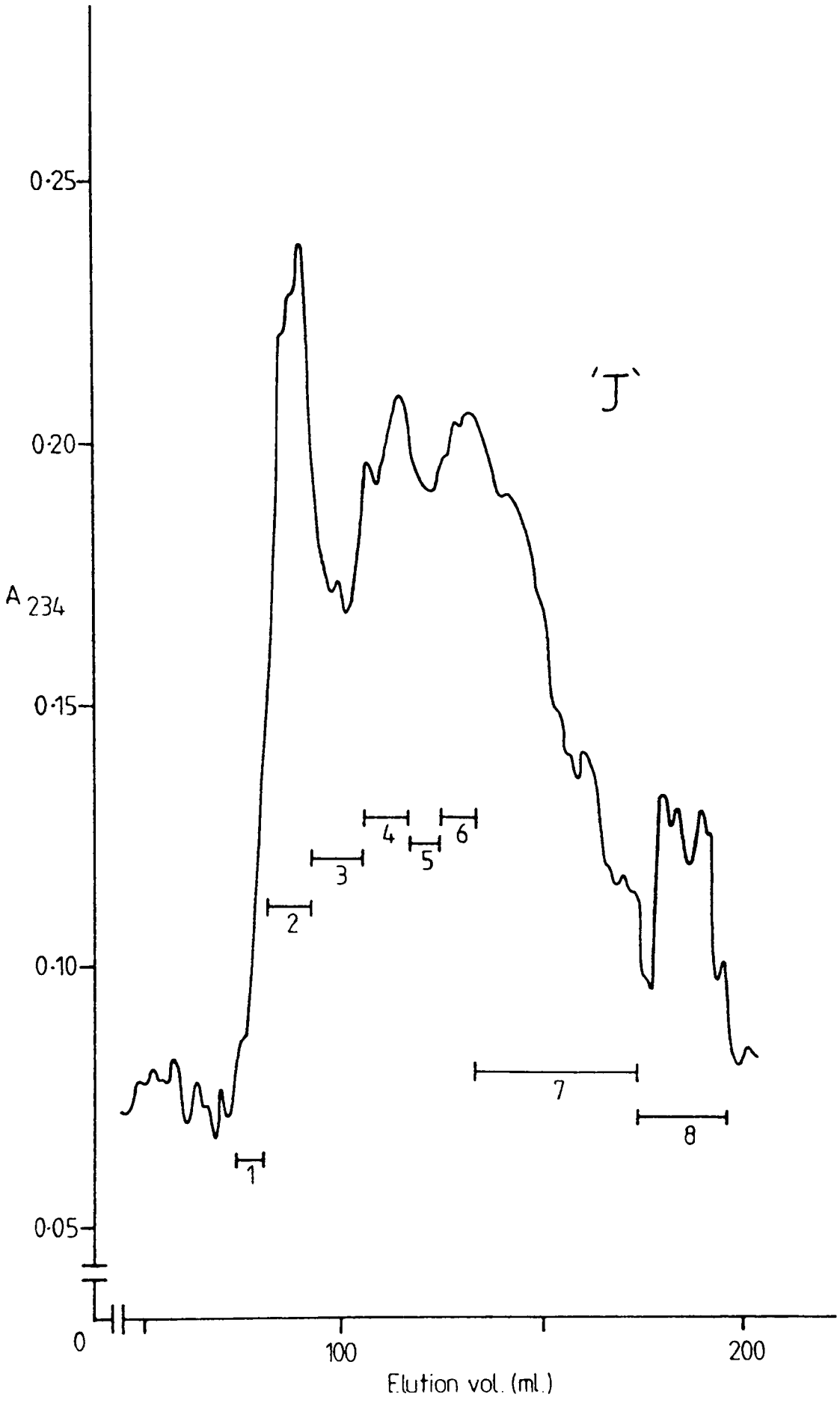


Figure 4.13 Fractionation 'J': Gel chromatography, on Sephadex G−50$_s$, of products from cleavage of putative peptide *P3* (fraction *F2*) with V8−protease in potassium phosphate buffer (Methods: Ch 2.IIC2). Chromatography was carried out as indicated for fractionation 'F' (fig. 4.9). Bars indicate the fractions pooled.

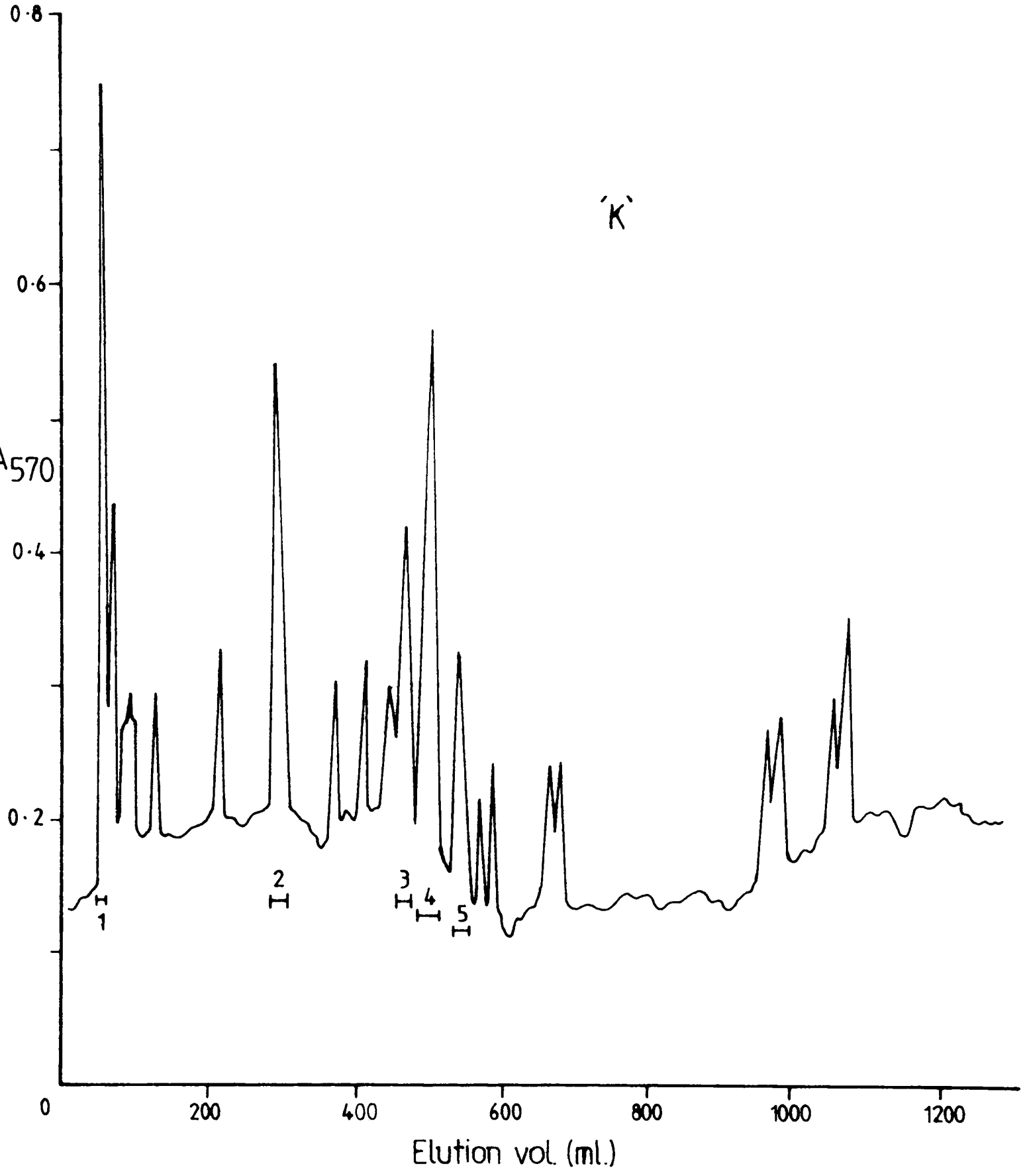


Figure 4.14 Fractionation 'K': Ion−exchange chromagraphy on AG50W−X4 resin, of V8−protease peptides of putative peptide *P3* (fraction *F2*) contained in the combined fractions *J3, J4, J5, J6* and *J7*. Chromatography was carried out as for all the other fractionations on this resin (see Ch 2.IID1). Bars indicate some of the fractions pooled.

Table 4.8

Amino acid compositions of V8−protease peptides of peptide α1−CB7 from bovine corneal collagen isolated in fractionation 'K' (fig. 4.14) alongside those of the theoretical peptides predicted from the known sequence (fig. 4.12). The peptides resulted from digestion of fraction F2 (fig. 4.9) with V8−protease in potassium phosphate buffer. Values are expressed as residues per peptide. '0' signifies a value less than 0.2.
For 'hydroxylysine' the lower limit is set at 0.1.

	K3	*P3b-c*	K5	*P3d*	K1	660 ↓ 678	K4	679 ↓ 726	K2	697 ↓ 726
Hydroxyproline	3.6	5	0	0	2.2	2	5.8	6	3.9	4
Aspartic acid	1.8	2	1.0	1	1.3	1	0.9	1	1.5	1
Threonine	0	0	0	0	0	0	1.3	1	0	0
Serine	0.2	0	0	0	0	0	2.1	2	1.0	1
Glutamic acid	2.8	3	0	0	0	0	1.2	1	1.3	1
Proline	4.2	3	0	0	5.2	5	6.5	7	5.3	5
Glycine	9.7	10	2.2	2	5.6	6	16.1	16	9.8	10
Alanine	4.3	4	1.9	2	2.9	3	8.3	8	4.5	4
Valine	0	0	0	0	0.8	1	0.8	1	0.8	1
Isoleucine	0	0	0	0	0.9	1	0	0	0	0
Leucine	0	0	0	0	0	0	0	0	0	0
Phenylalanine	0.9	1	0	0	0	0	1.0	1	0.8	1
Hydroxylysine	0.15	0	0.1	0	0	0	1.1	0.7*	0	0
Lysine	0.91	1	0.75	1	0	0	0.8	1.3*	1.0	1
Arginine	1.3	1	0	0	0	0	2.1	2	0.9	1
total	29.86	30	5.95	6	18.9	19	48.0	48	30.8	31

**Lys-684 is ~70% hydroxylated and Lys-725 is not hydroxylated in peptide α1-CB7 from calf skin (Fietzek et al., 1973; Kleinman et al., 1978)*

from fraction F2 (11.5 mg), the putative peptide P3 used as starting material for the V8−protease cleavage in phosphate buffer.

The amino acid composition of fraction K5 indicated it to be the hexapeptide P3d which contains the 'pre−Gly' lysyl position 657 in the sequence of peptide α1−CB7 (Table 4.8; fig. 4.12). Peptide K5 (P3d) contained 0.1 residues of hydroxylysine per peptide, thus indicating 10% hydroxylation of Lys−657 in peptide α1−CB7 of bovine corneal collagen. This was the first time that this lysyl position had been individually characterised in the present work. The yield of peptide K5 (P3d) was 23% relative to the original peptide α1−CB7 or 63% relative to fraction F2 (putative peptide P3).

Amino acid analysis of fraction K4 indicated it to be the peptide sequence 679−726 arising from an unexpected cleavage on the carboxyl side of valine at position 678 (Table 4.8; fig. 4.12). The unexpected cleavage (only Asp− bonds were expected to be cleaved) at Val−678 had also occurred before during the V8−protease cleavage of whole peptide α1−CB7 in ammonium bicarbonate buffer which liberated peptide G2, identical to peptide K4 (compare 'G2' and 'K4', Tables 4.7 and 4.8). It was however unclear as to why a similar 'Val−Gly' bond after Val−705 did not seem to be cleaved. Peptide K4 contains the third 'pre−Gly' lysyl residue in peptide P3 occurring at position 684 (see fig. 4.12). It also contains the 'post−Gly' lysyl residue at position 725. Peptide K4 contained 1.1 residues of hydroxylysine per peptide, occurring almost certainly at position 684 and 0.8 residues of lysine representing the single lysine residue at position 725 in peptide α1−CB7 (Table 4.8, fig. 4.12). These results agree with those obtained for the previously isolated identical peptide G2 which indicated the occurrence of 1.1 hydroxylysine and 1.1 lysine residues per peptide (Table 4.7). The absence of hydroxylysine at the 'post−Gly' lysyl position 725 had been previously proven by characterisation of tryptic peptide C4 (Tc8) and V8−protease peptide H4 (positions 697−726) which showed the occurrence of lysine but not hydroxylysine (Tables 4.1, 4.7). The overall yield of peptide K4 was 10% relative to peptide α1−CB7, or 28% relative to fraction F2 (putative peptide P3).

The amino acid composition of fraction K1 indicated that it was the 'other'

peptide, of sequence positions 660−678, resulting from the unexpected cleavage on the carboxyl side of Val−678 which also liberated its 'sister' (unexpected) peptide K4 (Table 4.8; fig. 4.12). Peptide K1 (660−678) did not contain any of the basic amino acid residues lysine and arginine as can be seen from its amino acid composition (Table 4.8). This explains its early elution in the fractionation on AG50W−X4 resin (fig. 4.14). The overall yield of peptide K1 was 16% of peptide α1−CB7, or 46% of fraction F2 (putative peptide P3).

Amino acid analysis of fraction K2 indicated it to be the peptide sequence 697−726 (Table 4.8; fig. 4.12), containing the fourth and last lysyl residue of theoretical peptide P3 at position 725 in the sequence of peptide α1−CB7. Peptide K2 did not contain hydroxylysine indicating that this lysine residue in bovine corneal collagen is not hydroxylated. This was expected as it is in a 'post−Gly' position. This peptide, K2, the result of an unexpected cleavage on the carboxyl side of threonine at position 696, is identical to peptide H4 which was isolated earlier from the V8−protease digest of peptide α1−CB7 in ammonium bicarbonate buffer (Table 4.7; fig. 4.6). Peptide H4, whose elution position on AG50W−X4 resin (fig. 4.11) and amino acid composition were identical to those of peptide K2, had already been shown to lack hydroxylysine (Table 4.7).

Peptide P3a was not identifiable among all the 'K' fractions analysed on the amino acid analyser. However, this peptide sequence, of positions 606 to 623, does not contain any lysyl positions (fig. 4.12) and was thus relatively unimportant for the purposes of this project. Residue positions 606−623 are within the theoretical tryptic peptides Tc4 and Tc5 which had been isolated earlier as fractions C1 and E4 respectively (see fig. 4.1).

With the isolation and amino acid analysis of peptides K3, K5, K4 and K2, the distribution of hydroxylysine between the lysyl positions 648, 657, 684, and 725 respectively in peptide P3 (of peptide α1−CB7) has now been determined. All the eleven lysyl residues of peptide α1−CB7, occurring at positions 564, 573, 581, 603, 648, 657, 684, 725, 729, 756 and 806 in the α1(I) chain of bovine corneal collagen

have been characterised with respect to the extent of lysyl hydroxylation within their respective tryptic and V8−protease peptides.

The next section describes the characterisation of hydroxylysine glycosides in the identified V8−protease peptides which contained hydroxylysine.

II C. QUANTITATIVE AND QUALITATIVE DETERMINATION OF HYDROXYLYSINE GLYCOSIDES IN THE V8−PROTEASE PEPTIDES.

The isolated V8−protease peptides described in the last two sections correspond to 256 out of the total of 271 amino acid residue positions present in peptide α1−CB7 of bovine corneal collagen (see figs. 4.6, 4.12). From the known sequence of peptide α1−CB7, the 'missing' fifteen residues are represented by the 8−residue sequence 616−623 and the 7−residue long theoretical peptide P9, the C−terminal V8−protease peptide of peptide α1−CB7. These positions however, do not contain any lysyl positions. The distribution of hydroxylysine glycosides among the isolated V8−protease peptides will now be discussed starting from the N−terminal end of peptide α1−CB7.

(Refer to Table 4.9 for the glycoside values and figures 4.6, 4.12 for positions in the amino acid sequence to aid the following discussion)

Peptide H3 (P1), the N−terminal tripeptide of peptide α1−CB7 was not analysed for hydroxylysine glycoside as it lacks lysine and hydroxylysine. Peptide H3 is followed in the sequence by the theoretical peptide P2 which was isolated as the two peptide fractions G4 and G5 (Table 4.7; fig. 4.6). This peptide contains the four lysyl positions 564, 573, 581 and 603. Both peptide fractions G4 and G5 contained hydroxylysine. Hydroxylysine glycoside analysis of fraction G4 (16% of the maximum yield for peptide P2) indicated the occurrence of 0.13 residues of the diglycoside, glucosylgalactosylhydroxylysine GGH), 0.14 residues of the monoglycoside, galactosylhydroxylysine (GH) and 0.16 residues of 'free', unglycosylated hydroxylysine per peptide (Table 4.9). Analysis of fraction G5 (12.6% of the maximum yield for peptide P2) showed the occurrence of 0.17 GGH, 0.20 GH and 0.17 FH residues per peptide. From these results and the relative yields of fractions G4 and G5, it can be

Table 4.9

Hydroxylysine glycoside and hydroxylysine contents of V8−protease peptides from bovine corneal peptide α1−CB7 isolated in this work. The values determined for 'whole' corneal peptide α1−CB7 are displayed for reference as are the values for *'total−hydroxylysine'* as determined previously in the respective acid hydrolysates (see * below). Values are expressed as residues per peptide. The columns specific for glycoside contents are headed in **bold.**

peptide	**GGH**	**GH**	***FH***	*'hyl'**
G4†	0.13	0.14	0.16	*0.41*
G5†	0.17	0.20	0.17	*0.62*
G4/G5‡	0.15	0.17	0.16	*0.50*
K3	0.06	0.03	0.06	*0.15*
K5	0.03	0.02	0.04	*0.10*
K4	0.12	0.34	0.54	*1.10*
H4, K2	–	–	–	*0*
H10	–	–	–	*0*
H6	0	0.07	0.06	*0.17*
H8, H9	–	–	–	*0*
α1-CB7	0.40	0.57	0.78	*1.8*

GGH, glucosylgalactosylhydroxylysine; GH, galactosylhydroxylysine; *FH*, free (unglycosylated) hydroxylysine

† Identical peptides recovered from two separate peaks of a single fractionation.

‡ weighted mean values, the yield ratio, G4:G5 being 1.27:1

– not analysed for hydroxylysine glycoside as peptides did not contain hydroxylysine

Note: Identified peptides not containing lysyl residues are automatically excluded from this table.

calculated that peptide P2 contained 0.15 GGH, 0.17 GH and 0.16 FH residues per peptide. As position 581 in peptide α1−CB7 is occupied by a 'post−Gly' lysyl residue, these quoted amounts of hydroxylysyl residues of peptide P2 are therefore distributed between the 'pre−Gly' positions 564, 573 and 603. The proportions distributed between these positions can be deduced from the data obtained from the relevant tryptic peptides (see fig. 4.1).

Peptide H2, representing the sequence 606−613 immediately following peptide P2 (G4/G5), did not contain lysine or hydroxylysine and therefore was not analysed for glycosides.

(Refer to Table 4.9 for the glycoside values and figs. 4.6, 4.12 for positions in the amino acid sequence to aid the current discussion)

Peptide K3 (P3b−P3c) corresponding to positions 624−653 (Table 4.8) contains the 'pre−Gly' lysyl position 648. Glycoside analysis of peptide K3 showed the occurrence of 0.06 GGH, 0.03 GH and 0.06 FH residues per peptide which therefore reside at position 648 in the sequence of peptide α1−CB7. Hence there is relatively little hydroxylation and even less glycosylation of this lysyl position.

Glycoside analysis of peptide K5 (P3d) showed it to contain 0.03 GGH, 0.02 GH and 0.04 FH residues per peptide. These trace amounts of hydroxylysine and hydroxylysine glycoside therefore reside at position 657.

Peptide K1 which immediately follows peptide K5 (P3d) in the sequence of peptide α1−CB7 does not contain lysine or hydroxylysine and therefore was not subjected to glycoside analysis.

Peptide K4, representing the sequence 679−726 which follows peptide K1 in the sequence of peptide α1−CB7 contains the 'pre−Gly' lysyl position 684 and the 'post−Gly' lysyl position 725 in the known sequence. The amino acid composition of peptide K4 showed the presence of one hydroxylysine and one lysine residue per peptide (Table 4.8). Glycoside analysis carried out on peptide K4 showed it to contain 0.12 GGH, 0.34 GH and 0.54 FH residues per peptide confirming the presence of a full hydroxylysine in this region, almost half of it being glycosylated. These values, of

the total values for peptide α1−CB7 of 0.40 GGH, 0.57 GH and 0.78 FH residues per peptide indicate therefore, that position 684 is the major site of hydroxylysine glycoside in peptide α1−CB7 of bovine corneal collagen.

The designation of all the hydroxylysine and hydroxylysine glycoside of peptide K4 to position 684 and none to position 725 was justified by the amino acid compositions of peptide H4/K2 (peptide sequence 697−726) containing position 725, which indicated that lysyl position 725 only contained lysine and not hydroxylysine in accordance with the 'post−Gly' location of position 725 (Tables 4.7, 4.8).

Peptide H10 (P4), next in the amino acid sequence of peptide α1−CB7, containing the lysyl residue at the 'pre−Gly' position 729, did not contain any hydroxylysine (Table 4.7) and thus was not analysed for hydroxylysine glycosides. The lack of hydroxylation at this lysyl residue deviates from the general observation in this work that all the other 'pre−Gly' lysyl positions in peptide α1−CB7 of bovine corneal collagen are hydroxylated to some extent. The lack of hydroxylation at this position was also shown in the analysis of tryptic peptide C4 (fig. 4.1; Table 4.1).

(Refer to Table 4.9 for the glycoside values and figs. 4.6, 4.12 for positions in the amino acid sequence to aid the current discussion)

Peptides H7 (P5) and H1 (P6) which successively follow peptide H10 (P4) in the amino acid sequence, did not contain lysine or hydroxylysine and therefore were not analysed for glycosides.

Peptide H6 (P7), next in the sequence of peptide α1−CB7, contains the 'pre−Gly' lysyl position 756. Analysis of its alkaline hydrolysate showed that this peptide contained 0.07 GH and 0.06 FH residues per peptide. None of the diglycoside GGH was detected. The sum total of the GH and FH residues was close to the 0.17 residues of total hydroxylysine per peptide detected in its acid hydrolysate (Table 4.7). There was the possibility that V8−protease cleavage on the carboxyl side of glutamic acid residues would be inhibited if the glutamate residue is followed by a glycosylated hydroxylysine residue, the parallel situation being the inhibition of cleavage by trypsin on the carboxyl side of glycosylated hydroxylysine residues (see the start to sections I

and II of this chapter). The fact that peptide H6 (P7) was the result of cleavage at the Glu−Lys bond between positions 755 and 756 (fig. 4.6) and the finding for peptide H6 (P7) that 0.07 GH residues occurs at position 756 indicate that V8−protease cleavage of Glu−Lys bonds is not necessarily inhibited when the lysyl residue involved occurs as galactosylhydroxylysine.

(Refer to Table 4.9 for the glycoside values and figs. 4.6, 4.12 for positions in the amino acid sequence to aid the current discussion)

The last lysyl position in peptide α1−CB7 occurs at position 806 which was isolated within V8−protease peptide H8/H9 (P8). Fractions H8 and H9 which both had amino acid compositions identical to theoretical peptide P8, did not contain hydroxylysine (Table 4.7). This is consistent with the 'post−Gly' position of Lys−806. Fractions H8 and H9 were thus not analysed for glycosides.

The theoretical V8−protease peptide, P9, representing the last seven amino acid positions at the C−terminal end of peptide α1−CB7 was not isolated during the course of this project. This peptide does not contain any lysine or hydroxylysine in the sequence of peptide α1−CB7 and is therefore not important for the purpose of identifying glycoside sites.

All the isolated V8−protease peptides of peptide α1−CB7 from bovine corneal collagen have now been described. The data obtained from analysis of these peptides will now be compiled with the data obtained previously from the tryptic peptides, especially with respect to hydroxylation of lysine and glycosylation of hydroxylysine at the eleven lysyl positions in peptide α1−CB7.

III. SITES OF HYDROXYLYSINE AND HYDROXYLYSINE GLYCOSIDES IN PEPTIDE α1−CB7 OF BOVINE CORNEAL COLLAGEN : A SUMMARY.

Before summarising the findings reported in this chapter, a reminder of some aspects of the work might be useful.

Firstly, collagen of the bovine corneal stroma is mainly type I collagen (see Chapter 1.IIG). The results presented in Chapter 3 are consistent with this view. The amino acid sequence for peptide α1−CB7 of calf skin type I collagen was thus used as the basis and guideline for the work carried out. Twenty−four different peptides with twenty−four different amino acid compositions were isolated from digestions of bovine corneal collagen peptide α1−CB7 with trypsin and V8−protease. These peptides had amino acid compositions consistent with those of theoretical tryptic and V8−protease peptides. All together, they account for 265 out of the total 271 amino acid residue positions in the primary sequence of peptide α1−CB7 from calf skin.

Secondly, the main aim of the present study was to pinpoint the location of galactose and glucosyl−galactose attachment in peptide α1−CB7. These saccharides are covalently−attached only via the hydroxyl groups of hydroxylysine residues. Needless to say, enzymatic hydroxylation of lysine residues is a necessary prerequisite for the glycosylation of the resulting hydroxylysine residues during biosynthesis of the (pro)collagen chains. Neither hydroxylation nor glycosylation necessarily proceed to completion at any particular lysyl position in a population of collagen (pro−) α chains As there are eleven lysyl positions in peptide α1−CB7, it was necessary to fragment the peptide with proteolytic enzymes to facilitate individual characterisation of all eleven positions.

Peptide α1−CB7, and peptides derived from it were first subjected to acid hydrolysis. Under these conditions, all peptide bonds as well as glycosidic linkages are cleaved. All hydroxylysyl residues, whether they were glucosylgalactosylhydroxylysine (GGH), galactosylhydroxylysine (GH) or 'free' (unglycosylated) hydroxylysine (FH) were thus converted to unglycosylated hydroxylysine during acid hydrolysis in 6M HCl and therefore appeared as 'total' hydroxylysine in amino acid compositions of the acid

Table 4.10

Summary of the distribution of hydroxylysine and hydroxylysine glycosides between the eleven lysyl positions of peptide α1-CB7 from bovine corneal collagen as indicated by analysis of peptides isolated from proteolytic digests. *Asterisks denote 'post-Gly' lysyl positions. Rows 'a' show the values obtained for tryptic peptides while rows 'b' and 'c' show those obtained for V8-protease peptides, the relevant peptides being indicated below in the '*KEY*'. GGH, GH and FH values were obtained by analysis of alkaline hydrolysates, and Hyl† values by analysis of acid hydrolysates. GGH, glucosylgalactosylhydroxylysine; GH, galactosylhydroxylysine; FH free (unglycosylated) hydroxylysine; Hyl†, 'total' hydroxylysine. Values are expressed as residues per peptide.

		position number	564	573	* 581	603	648	657	684	* 725	729	756	* 806
GGH	a		.07	.06	.04		.27			–		0	–
	b			*.15*			*.06*	*.03*		–	–	*0*	–
	c								*.12*				
GH	a		.10	.06	.06		.40			–		.08	–
	b			*.17*			*.03*	*.02*		–	–	*.07*	–
	c								*.34*				
FH	a		.06	.06	.05		.48			–		.04	–
	b			*.16*			*.06*	*.04*		–	–	*.06*	–
	c								*.54*				
Hyl†	a		.27	.18	.17		1.1			0		.11	0
	b			*.50*			*.15*	*.10*		*0*	*0*	*.17*	*0*
	c								*1.1*				

KEY

a	C_5, C_7	C6	C1	D1		C4		B2	C3
b		*G4/G5*		*K3*	*K5*	H_4, *K2*	*H10*	*H6*	H_8, H_9
c					*K4*				

Note: values for uncleaved peptide α1-CB7 were 0.4 GGH, 0.57 GH, 0.78 FH and 1.8 Hyl† residues per peptide.

hydrolysates. Thus, as well as facilitating the identification of particular peptides, amino acid analysis of acid hydrolysates also simultaneously revealed the *extent of lysyl hydroxylation* within identified peptides which contained lysyl residues. Particular peptides whose acid hydrolysates contained hydroxylysine were then subjected to alkaline hydrolysis (in 2M NaOH), a procedure which hydrolyses peptide bonds but not glycosidic linkages. It was then possible to determine the *nature and extent of hydroxylysyl glycosylation* by subjecting the alkaline hydrolysates to amino acid analysis, which simply reveals the proportions of the hydroxylysyl residues occurring as the diglycoside GGH, the monoglycoside GH and the unglycosylated FH. (It is not possible, by carrying out amino acid analyses on sample alkaline hydrolysates, to identify peptides while simultaneously determining the relative proportions of GGH, GH and FH, as alkaline hydrolysis destroys some amino acids, accompanied by the generation of amino acid breakdown products [Hill, 1965]. These undesirable aspects of alkaline hydrolysis make it impossible to calculate the amino acid compositions necessary for the identification of peptides).

The extents of hydroxylation and glycosylation at the eleven lysyl positions in peptide α1−CB7 will now be discussed, starting with that nearest the N−terminal end, lysyl position 564. When two identical peptides were isolated from a single proteolytic digest of peptide α1−CB7, 'weighted' mean values for GGH, GH and FH are quoted, having taken into account the relative yields of the respective peptide fractions. Percentage GGH, GH, FH and lysine contents were calculated assuming an integral total of 1.0 lysyl residue at each of the positions. Throughout this work, peptides and fractions prefixed by capital letters 'A' to 'E' represent those derived from tryptic digestions while those prefixed with capital letters 'F' to 'K' (omitting letter 'I') were those derived from the V8−protease digestions. Peptides prefixed with different capital letters were recovered from different fractionations although they may originate from the same tryptic or V8−protease digest.

(Refer to Table 4.10 for a summary of glycoside contents and figs. 4.1, 4.12 for approximate and detailed positions of lysyl residues)

Position 564

The results obtained for tryptic peptide fractions C5 and C7, identical peptides containing lysyl position 564, indicated the occurrence of 0.07 GGH, 0.1 GH and 0.06 FH residues per theoretical peptide Tc2. Therefore, of position 564, 7% is GGH, 10% is GH, 6% is FH and 77% is (unmodified) lysine.

Positions 573 and 581

Lysyl positions 573 and 581 were characterised in tryptic peptide C6 and V8−protease peptide G4/G5 which contained both these positions, except that peptide G4/G5 also contained a further two lysyl positions (564 and 603). Positions 573 and 581 were not separated from each other into different peptides in any of the isolated peptides but as the lysyl residue at position 581 is in a 'post−Gly' position and thus not expected to be hydroxylated, the hydroxylysyl residues (partly glycosylated) found in peptide C6 can be assigned to position 573. This position therefore contains 0.06 GGH, 0.06 GH and 0.06 FH residues per peptide. This means that of position 573, 6% is GGH, 6% is GH, 6% is FH and 82% is lysine.

(Refer to Table 4.10 for a summary of glycoside contents and figs. 4.1, 4.12 for approximate and detailed positions of lysyl residues)

Position 603

Position 603 was isolated and characterised in tryptic peptide C1, which contained 0.04 GGH, 0.06 GH and 0.05 FH residues per peptide (Table 4.10). Thus, of position 603, 4% is GGH, 6% is GH, 5% is FH and 85% is lysine.

The combined totals of 0.17 GGH, 0.22 GH and 0.17 FH residues for positions 564, 573, 581 and 603 determined from the tryptic peptides are close to the values of 0.15 GGH, 0.17 GH and 0.19 FH residues found per V8−protease peptide G4/G5 which contains the same four lysyl positions.

Position 648

Characterisation of V8−protease peptide K3 showed that position 648 contained 0.06 GGH, 0.03 GH and 0.06 FH residues per peptide. Therefore, of position 648, 6% is GGH, 3% is GH, 6% is FH and 85% is lysine.

Position 657

This position, isolated in V8−protease peptide K5, contains 0.03 GGH, 0.02 GH and 0.04 FH residues per peptide. Thus, of lysyl position 657, 3% is GGH, 2% is GH, 4% is FH and 91% is lysine.

(Refer to Table 4.10 for a summary of glycoside contents and figs. 4.1, 4.12 for approximate and detailed positions of lysyl residues)

Position 684

V8−protease peptide K4 contains lysyl positions 684 and 725, the latter being a 'post−Gly' lysyl position. The 0.12 GGH, 0.34 GH and 0.54 FH residues per peptide K4 therefore reside at position 684, making this the ***major site*** of hydroxylysine glycoside (see Table 4.10) in peptide α1−CB7 of bovine corneal collagen. This lysyl position is fully hydroxylated, of which 12% is GGH, 34% is GH and 54% is FH.

Verification of the absence of hydroxylysine glycoside at position 725 is discussed below.

The combined totals of 0.21 GGH and 0.39 GH residues found in V8−protease peptides K3, K5 and K4 are close to the values obtained for tryptic peptide D1 which indicated the occurrence of 0.27 GGH and 0.40 GH residues per peptide. The total of 0.64 FH residues found distributed between peptides K3, K5 and K4 was comparable to, but slightly higher than, the total of 0.48 FH residues found in tryptic peptide D1. Peptide D1 contains the three 'pre−Gly' lysyl positions 648, 657 and 684 present in peptides K3, K5 and K4 respectively but not the 'post−Gly' lysyl position 725 occurring in peptide K4. This 'post−Gly' lysyl position 725 has been shown to be unhydroxylated as expected (see below).

(Refer to Table 4.10 for a summary of glycoside contents and figs. 4.1, 4.12 for approximate and detailed positions of lysyl residues)

Positions 725 and 729

Position 725, occurring as a 'post−Gly' lysine residue, was not found to be hydroxylated by amino acid analysis of the identical V8−protease peptides H4 and K2. The lack of lysyl hydroxylation in tryptic peptide C4 shows that neither Lys−725 nor Lys−729 is hydroxylated. It is surprising that no hydroxylation of the lysyl residue at the 'pre−Gly' position 729 was found as it has been observed that all the other 'pre−Gly' lysyl positions in peptide α1−CB7 are hydroxylated to some extent. The lack of lysyl hydroxylation at position 729 was confirmed by the absence of hydroxylysine in V8−protease peptide H10.

Position 756

Position 756 was isolated and characterised in tryptic peptide B2 and V8−protease peptide H6. Peptide B2 contained 0.08 GH and 0.04 FH residues per peptide but no GGH. Peptide H6 also had no GGH and contained 0.07 GH and 0.06 FH residues per peptide. Position 756 thus contained an average of 0.075 GH and 0.05 FH residues. Therefore of this lysyl position, about 8% is GH, 5% is FH and 87% is lysine.

Position 806

The final lysyl residue in the amino acid sequence of peptide α1−CB7 at position 806 occurs in a 'post−Gly' position. Analysis of tryptic peptide C3 and V8−protease peptide H8/H9 show that the lysine at position 806 is not hydroxylated and hence does not occur as hydroxylysine glycoside.

(Refer to Table 4.10 for a summary of glycoside contents and figs. 4.1, 4.12 for approximate and detailed positions of lysyl residues).

Combined totals

The combined total of 0.44 GGH residues distributed between the eleven lysyl positions as characterised in the tryptic peptides is close to the sum of 0.36 GGH residues as characterised in the V8−protease peptides and is also close to the value of 0.40 GGH residues determined in whole peptide α1−CB7 (Table 4.10). There is also reasonable agreement for the combined values with respect to GH, FH and 'total' hydroxylysine as determined in the tryptic peptides of peptide α1−CB7, V8−protease peptides of peptide α1−CB7 when compared to the values obtained from the 'whole' peptide α1−CB7 (Table 4.10). The largest difference is that of 0.20 residues between the sum of 0.69 FH among the tryptic peptides and the sum of 0.89 FH residues among the V8−protease peptides.

In the next, final chapter, the positions and amounts of hydroxylysine glycoside discovered in peptide α1−CB7 of bovine corneal collagen will be discussed in relation to the 'D−stagger' arrangement of type I collagen molecules especially with respect to the possible role of hydroxylysine glycosides in limiting collagen fibril diameter.

CHAPTER 5

DISCUSSION

The results presented in the earlier part of this thesis (Chapter 3) indicate that collagen of the bovine corneal stroma gives an amino acid composition and a cyanogen bromide (CNBr) peptide pattern similar to those obtained from type I collagen of other tissues. This is in accord with the findings of other investigators, who, by the use of immunohistological and biochemical techniques, have shown that the collagen of bovine as well as of chick, rabbit, and human corneal stroma is largely type I collagen (see Chapter 1.IIG). During the present work, no obvious marker CNBr peptides of the collagen types II, III, IV, V and VI were seen among the peptides detected by sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS−PAGE) of bovine corneal collagen CNBr digests. The absence of type IV collagen was further indicated by the absence of cystine from the amino acid composition of bovine corneal (stromal) collagen. Further evidence that bovine corneal collagen is indeed mainly type I collagen arose from the isolation of particular CNBr peptides. These peptides, corresponding to three major bands of the SDS−PAGE patterns for the corneal collagen CNBr digests, gave amino acid compositions similar to those of peptides α1−CB3, α1−CB7 and α1−CB8 of type I collagen from bovine (calf) skin. Together, these peptides represent 66% of the length of the α1(I) chain.

The 271−residue peptide α1−CB7, representing approximately 26% of the α1(I) chain was purified in quantity for more detailed analysis. Appropriate use of the proteolytic enzymes trypsin and V8−protease to cleave peptide α1−CB7 led to the isolation of over twenty different peptides. Every peptide produced by tryptic and V8−protease cleavage that had an integral amino acid composition was consistent with the amino acid sequence of peptide α1−CB7 from calf skin. Ten different tryptic peptides with ten different amino acid compositions were isolated and were of the size range 6 to 63 residues in length. These overlapped with fourteen different V8−protease peptides with fourteen different amino acid compositions and which were of the size range 3 to 51 residues in length. These results therefore give no cause to

suspect that the amino acid sequence of bovine corneal peptide α1−CB7 may originate from a collagen gene different to that which codes for the α1(I) chain of bovine skin. A similar conclusion can be derived from the work carried out previously on the 149−residue peptide α1−CB3 of bovine corneal collagen (Panjwani and Harding, 1978b).

During the preparation and characterisation of CNBr peptides of bovine corneal collagen, it was essential to know the identities of particular peptides before carrying out the more detailed aspects of this project. However, on scanning the literature for the relevant information, a distinct inconsistency was found concerning the relative electrophoretic mobilities of peptides α1−CB7 and α1−CB8 on SDS−gels. Thus, although most authors quoted a faster−migrating peptide α1−CB8, equally credible authors were quoting a faster−migrating peptide α1−CB7 even up till the year 1985. As SDS−PAGE was to be frequently used in primary assessments of CNBr−peptide identity and purity throughout the early stages of this project, it was thus necessary to determine beyond doubt the relative electrophoretic mobilities of peptides α1−CB7 and α1−CB8 on SDS−gels. Purified peptide fractions obtained from fractionations of CNBr digests on carboxymethyl cellulose were thus subjected to SDS−PAGE and amino acid analysis. One of the peptide fractions was also subjected to N−terminal determination by the 'dansylation' method. In combination with the available primary sequence data for peptides α1−CB7 and α1−CB8, it was thus shown beyond doubt that the 279−residue peptide α1−CB8 has a greater electrophoretic mobility than the 271−residue peptide α1−CB7 on SDS−gels. This anomalous behaviour of these two peptides relative to one another should therefore be given serious consideration by researchers who use SDS−PAGE to examine either the extent of CNBr cleavage at particular methionyl bonds or more commonly, the degree of cross−linking involving peptides α1−CB7 and α1−CB8. This applies especially to those who have indicated incorrect relative mobilities for these two peptides (see Chapter 3.III). The position of bands on SDS−gels has been used too often as the sole method in the identification of collagen fragments with identities assigned solely on the basis of relative molecular

weights. The complex mixture of fragments present in a CNBr digest of collagen and the anomalous migration of some of these fragments makes reliance on this single technique unwise.

The central issue of this current work is that the collagenous corneal stroma, in order to be transparent, contains collagen in the form of thin, uniform fibrils of constant diameter. Many authors have noted an inverse relationship between collagen fibril diameter and the hydroxylysine glycoside contents of various collagens (see Chapter 1.IIE2c.i). It has also been noted that in some collagen diseases such as osteogenesis imperfecta, overmodification of lysyl residues is accompanied by the occurrence of abnormally thin collagen fibrils (see Chapter 1.IIF).

My results confirm that the hydroxylysine glycoside content of corneal collagen is higher than those reported for type I collagen from other tissues such as tendon, skin and sclera (Table 1.1). At a more detailed level, the three CNBr peptides α1−CB3, α1−CB7 and α1−CB8 of bovine corneal collagen isolated in the present work were shown to contain more hydroxylysine glycoside than the corresponding peptides of bovine skin type I collagen (Table 5.1).

The characterisation of these three CNBr peptides in the present work coupled with that of other CNBr peptides reported by Panjwani and Harding (1978a) means that of the total eight CNBr peptides of the bovine corneal α1(I) chain (see fig. 1.9) six have been characterised, namely peptides α1−CB2, α1−CB4, α1−CB8, α1−CB3, α1−CB7 and α1−CB6 which together comprise approximately 95% of the α1(I) chain. Peptide α1−CB2 does not contain any lysyl residues while peptide α1−CB4 was found not to contain hydroxylysine or hydroxylysine glycoside (Panjwani and Harding, 1978a), consistent with its amino acid sequence which indicates two lysine residues occurring in 'post−Gly' positions. Peptides α1−CB8, α1−CB3, α1−CB7 and α1−CB6 of bovine corneal collagen have all been found to contain more hydroxylysine glycoside than the corresponding peptides from bovine skin (Table 5.1). It can be assumed that peptide α1−CB5 of bovine corneal collagen contains one residue of the diglycoside GGH as it contains the major site of glycosylation found in many collagens (see Panjwani and

Harding, 1978a), and that the single lysyl residue in peptide α1−CB0,1 occurring in the 'non−helical' N−telopeptide does not occur as hydroxylysine glycoside. Values from the proven data plus the assumed values for the two uncharacterised peptides amount to 3.34 GGH residues and 2.40 GH residues per thousand amino acid residues or, approximately per α1(I) chain. Doubling these values to account for the occurrence of two α1(I) chains per molecule, and adding the resulting value to the values obtained for peptide α2−CB3,5 (see Table 5.1) results in the values of 7.80 GGH and 6.04 GH residues that can be accounted for per molecule (two α1 plus one α2 chains or 3000 amino acid residues). Each molecule (about 3000 amino acid residues) has been ascertained in this work as containing 9.6 GGH and 6.0 GH residues (three times the 'residues per 1000' values). Having accounted for all the GH residues, it can be deduced that the 1.8 GGH residues still to be accounted for, most probably resides in the 321−residue long peptide α2−CB4 as peptides α2−CB0 and α2−CB2 do not contain any lysyl positions and assuming that the single lysyl residue occurring in the 'non−helical' region of the 12−residue long peptide α2−CB1 does not occur as hydroxylysine glycoside (refer to figs. 1.9 and 1.10 to gain access to the sequence data).

In addition to the characterisation of whole CNBr peptides, the pinpointing of exact locations of glycoside in peptide α1−CB7 makes it possible to discuss the possible effects of particular glycoside sites on collagen fibril structure. The quasi−hexagonal model for collagen molecular−packing, as interpreted from equatorial X−ray diffraction patterns of collagen fibrils (see Miller, 1982; Chapter 1.IIE1), indicates close−packing of collagen molecules. Morgan et al. (1970) postulated that due to a 'steric−hindrance' effect, any glycoside present in the 'overlap' region of the collagen fibril D−stagger arrangement may limit fibril diameter or completely prevent the formation of fibrils. Glycoside residues present in the 'overlap zones' (see figs. 1.8, 1.13) of the collagen molecule would thus be expected to be accomodated only on the fibril perimeter but not inside the fibril.

Panjwani and Harding (1978b) showed that the 149−residue peptide α1−CB3 of

Table 5.1

Hydroxylysine and hydroxylysine glycoside contents of cyanogen bromide peptides from bovine skin and cornea. Values are expressed as residues per peptide.

	Hydroxylysine		Hydroxylysine glycoside	
Peptide	skin	cornea	skin*	cornea
α1-CB3	0.3	0.9	0	0.9
α1-CB6	1.1	2.3	0.3	1.7
α1-CB7	0.9	1.8	0	0.97
α1-CB8	0.8	2.5	0.1	1.21
α2-CB3,5†	6.6	6.9	0.6	2.36

Note: Hydroxylysine ('total') and glycoside contents were determined in acid and alkaline hydrolysates respectively

* difference between total hydroxylysine and that released by alkaline hydrolysis

† preparation not entirely homogeneous for 'cornea'

Values for 'skin' from Volpin and Veis (1973)

Values for corneal α1-CB3 α1-CB6 and α2-CB3,5 *from Panjwani and Harding (1978a) and for α1-CB7 and α1-CB8 from the present work.*

bovine corneal collagen contains hydroxylysine glycoside in an overlap zone at position 531 in the helical region of the α1(I) chain, a position which is not glycosylated in calf skin type I collagen. Before the present work, this was the only glycoside found in an overlap zone of a type I collagen. The ultimate aim of the present work was to identify and characterise glycoside sites in the 271−residue peptide α1−CB7.

It has been a common observation in amino acid sequence studies of collagens that only those lysyl residues occurring in the 'Y' position of the repeating "–Gly–X–Y–" triplet sequence ('pre−Gly' residues) are hydroxylated and subsequently glycosylated while 'post−Gly' lysyl residues (those in the 'X' position) are never modified. The known amino acid sequence of peptide α1−CB7 from calf skin indicates eight 'pre−Gly' lysyl residues out of a total of eleven lysyl residues. This has made the identification of glycoside sites in this peptide somewhat more difficult than the previous characterisation of sites in peptide α1−CB3 (Panjwani and Harding, 1978b) where there is a total of only five lysyl residues, of which only two are in 'pre−Gly' positions.

The results of the present work have shown that out of the three 'post−Gly' lysyl positions, two were occupied by unmodified lysine as expected. The remaining 'post−Gly' lysyl residue at position 581 was not individually characterised but can be assumed to be unmodified. Of the eight 'pre−Gly' lysyl positions in peptide α1−CB7, seven were shown to be hydroxylated and glycosylated to varying extents. The proportions of these lysyl positions occurring as glucosylgalactosylhydroxylysine (GGH), galactosylhydroxylysine (GH), 'free' unglycosylated hydroxylysine (FH) and unmodified lysine (Lys) are presented in figure 5.1.

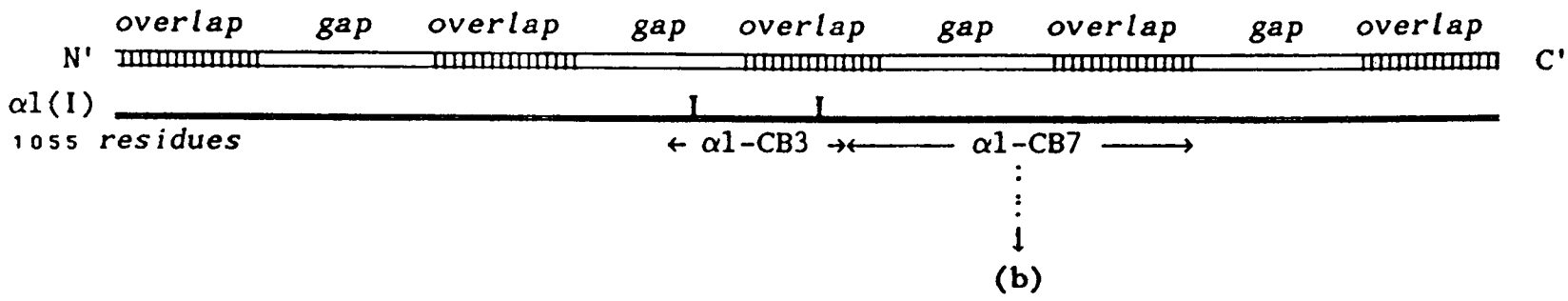
It can be seen in figures 1.8 and 1.13 (see Chapter 1) that the alternating gap and overlap 'regions' of the periodic collagen fibril superimposes four gap 'zones' and five overlap 'zones' onto each collagen molecule, the N−terminus of the molecule marking the beginning of an overlap zone. The traditionally accepted length for the gap region is 0.6D and that for the overlap region is 0.4D where the fibril 'D'period corresponds to the axial repeat length of 67nm or 234 amino acid residues. X−ray

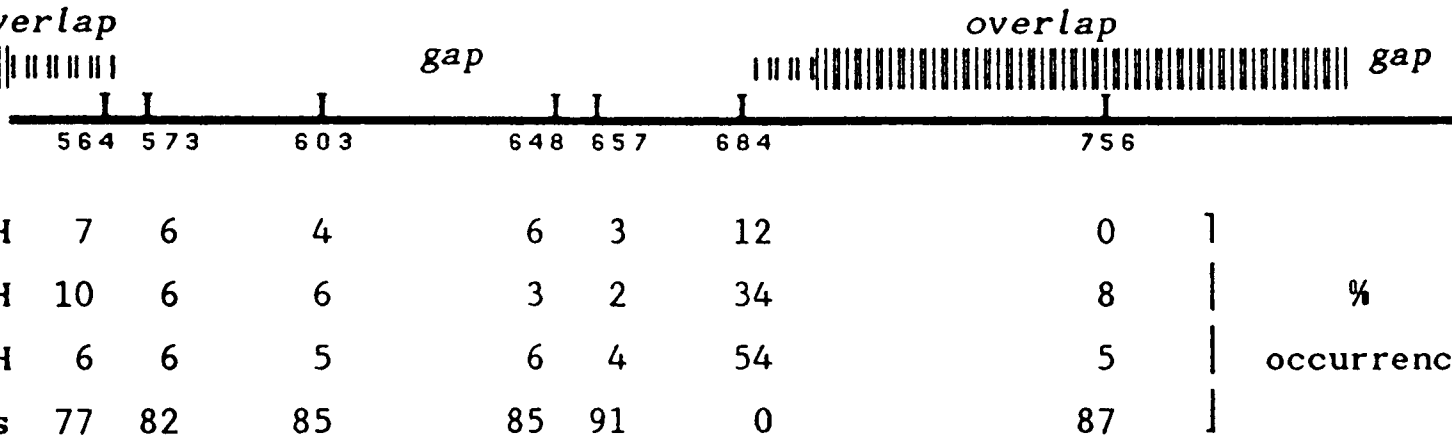
diffraction data plus theoretical calculations for the telopetide regions have resulted in the more recent estimate for the gap:overlap ratio of 0.53D:0.47D (Hulmes et al., 1980; Brodsky and Eikenberry; 1982). Using this updated value, the gap and overlap zones in relation to peptide α1−CB7 are indicated in figure 5.1, in which it can be seen that peptide α1−CB7 spans just over one gap zone and one overlap zone. It should be borne in mind however, that there may be a little flexibility in the exact position of the gap/overlap boundaries due to some flexibility in the 'non−helical' telopeptide domains which mark these points. The seven glycoside sites found in the present work (figure 5.1) will now be discussed in relation to the gap and overlap regions of the fibril and hence their significance in limiting fibril diameter, starting with the site nearest the N−terminal end of peptide α1−CB7.

(All the residue positions in the following discussion refer to positions in the α1(I) chain unless stated otherwise. Refer to fig. 5.1 to aid the discussion)

The site at position 564, of which 7% is GGH and 10% is GH, occurs just inside an overlap zone. In the fibril D−stagger arrangement, Lys−564 occurs almost directly opposite Lys−17^{C} in the C−telopeptide region of a molecule staggered by two D−periods (Fietzek and Kühn, 1975; see fig. 5.1). Lys−17^{C} is an aldehyde−forming lysine which has been found to be one of the major cross−link sites in type I collagen. It is usually cross−linked to Lys−87 of a molecule staggered by 4D (see fig. 1.8 in Chapter 1; Eyre et al., 1984). If however it does form cross−links with Lys−564, the occurrence of 17% of Lys−564 in corneal collagen as hydroxylysine glycoside could play a role in limiting collagen fibril diameter by altering cross−linking properties as well as posing as a simple steric hindrance in the overlap regions of the fibril.

The site at position 573, of which 6% is GGH and another 6% is GH occurs just inside a gap zone. The sites at positions 603 (4% GGH, 6% GH), 648 (6% GGH, 3% GH) and 657 (3% GGH, 2% GH) all occur in the central part of a gap zone. These glycosides at these sites can therefore fit into the axial 'holes' present in





GGH	7	6	4	6	3	12	0	]
GH	10	6	6	3	2	34	8	%
FH	6	6	5	6	4	54	5	occurrence
Lys	77	82	85	85	91	0	87	]

Figure 5.1 Hydroxylysine glycoside sites (I) which have so far been located in the α1(I) chain of bovine corneal collagen in relation to the gap and overlap regions of the collagen fibril: (a) Glycoside sites in peptide α1−CB3 (from *Panjwani and Harding, 1978b*) at positions *408* and *531*. (b) Glycoside sites and their extent of glycosylation in peptide α1−CB7 of bovine corneal collagen. The diagram shows the entire 271−residue peptide α1−CB7 in relation to the gap/overlap structure of the type I collagen fibril. Numbers printed in small numerals beneath the locations of hydroxylysine glycoside refer to amino acid positions in the 'helical' region of the α1(I) chain. Beneath the positions are the respective percentage values for the glycosides, glucosylgalactosylhydroxylysine (GGH) and galactosyl−hydroxylysine (GH), as well as free, unglycosylated hydroxylysine (FH) and unmodified lysine (Lys) which occur at these positions. *Note: lengths for the gap (124 residues) and overlap (110 residues) regions were taken from Fraser et al. (1983).* '|||' are the ends of the overlap regions occupied by the telopeptide domains of collagen molecules.

the fibril and are therefore not expected to limit lateral fibril growth unless they can inhibit fibril growth by mechanisms not purely involving steric effects (see later part of discussion).

This is the first time that any of these five lysyl positions 564, 573, 603, 648 and 657 have been identified and characterised as glycoside sites in any type I collagen. In most cases, glycoside sites identified during sequence analysis are only identified as such if they are almost fully glycosylated.

The next glycoside site at position 684 was found in this work to be the major glycoside site of peptide α1−CB7. Of this position, 12% is GGH and 34% is GH. The remaining proportion of this position is occupied by unglycosylated hydroxylysine. This is the first time that this lysyl position has been identified and characterised as a glycoside site in corneal type I collagen. The identical lysyl position in type I collagen of calf skin is only about 70% hydroxylated, and of which only 30% occurs as the monoglycoside GH (Fietzek and Kühn, 1976). This position occurs just on the inside edge at the C−terminal end of the gap zone which also contained positions 573, 603, 648 and 657. The glycoside occurring at this position could therefore also theoretically fit into the axial 'holes' in the gap region of the fibril. It is interesting to note that the glycoside site found in type I collagen from different tissues such as skin and cornea at position 684 occurs near the gap/overlap boundary (fig. 5.1). This also applies to the major glycoside site at position 87 of the α1(I), α2(I) and α1(III) chain (see Bornstein and Traub, 1979; Seyer and Kang, 1977). It may be possible to speculate that particular glycoside sites may have been conserved among the fibrillar collagens, and which serve particular functions during fibril formation, perhaps acting as 'pegs' in aligning collagen molecules with respect to the 'D−stagger'.

(All the residue positions in the following discussion refer to positions in the α1(I) chain unless stated otherwise. Refer to fig. 5.1 to aid the discussion)

The remaining glycoside site, at position 756, of which 8% is GH, occurs approximately in the middle of the main overlap zone spanning peptide α1−CB7. This

glycoside is thus clearly in a position to present steric hindrance to fibril growth according to the postulate of Morgan et al. (1970). This is the first time that this lysyl position has been identified and characterised as a glycoside site in any type I collagen.

The seven glycoside sites which I have pinpointed in peptide α1−CB7 increases the total number of identified glycoside sites in corneal collagen from two (Panjwani and Harding, 1978b) to nine. In addition, it has been assumed but not proven that glycoside also occurs at position 87 in peptide α1−CB5 which is found to be glycosylated in other type I collagens (see Panjwani and Harding, 1978a).

There was insufficient time to identify the glycoside sites in peptide α1−CB8 of bovine corneal collagen which was determined in this work, for the first time, as containing 0.25 GGH, 0.96 GH and 1.2 unglycosylated hydroxylysine residues per peptide. Of the ten lysyl positions in peptide α1−CB8 of calf skin collagen (Glanville et al., 1983), eight are in 'pre−Gly' positions, signifying eight possible glycoside sites. In the peptide from calf skin, hydroxylysine was found only at lysyl positions 174, 219 and 342 which were found to be 40%, 30% and 20% hydroxylated respectively. It is therefore probable that these three positions are the major glycoside sites in corneal peptide α1−CB8. These three positions all occur within a gap zone, and if glycosylated, would not necessarily play a role in limiting fibril growth, at least according to the postulate of Morgan et al., (1970). However, as the corneal peptide has been shown in this work to contain 2.5 'total' hydroxylysine residues per peptide compared to 0.9 'total' hydroxylysine found per peptide in calf skin, it would be premature to rule out the other 'pre−Gly' lysyl positions 252, 264, 270, 327 and 360 from also being likely glycoside sites in peptide α1−CB8 of corneal collagen. Judging from the previously described distibution of glycosides among seven out of the eight 'pre−Gly' lysyl positions in peptide α1−CB7 (this work) and two out of two 'pre−Gly' lysyl positions in peptide α1−CB3 (Panjwani and Harding, 1978b) of bovine corneal collagen, it seems reasonable to predict that most of the eight 'pre−Gly' lysyl positions in peptide α1−CB8 of corneal collagen will also be found to occur as

glycosides to some extent. In relation to the gap/overlap structure of the fibril, lysyl positions 252, 264 and 270 are in an overlap zone. Lys−327 lies just inside the same overlap zone while Lys−360 in a gap zone (Fraser et al. 1983)

Returning to the identified glycoside sites in peptide α1−CB7, it may seem that the 8% glycosylation at Lys−756 within an overlap zone and 17% glycosylation at Lys−564 just inside an overlap zone, could not be very significant in limiting collagen fibril diameter with respect to their possible steric−hindrance effect. However, if the quasi−hexagonal model for collagen molecular packing is idealized to accomodate straight, rigid rod−like molecules in a tightly−packed hexagonal array, for example as represented earlier in figure 1.13, it can be estimated by simple calculation that a fibril 23.8nm in diameter would on its perimeter have about 20% of its total collagen molecules, each of diameter 1.4nm. *(Collagen fibrils of the corneal stroma are consistently about 25nm thick: Craig and Parry, 1981).* Harding et al. (1980) suggested that hydroxylysine glycosides, in limiting lateral fibril growth, would ultimately be located mainly on the periphery and not in the core of the fibril. Thus, if limitation of fibril diameter is determined by any one particular glycoside residue occurring in an overlap zone, the particular site would on average only have to be 10% glycosylated if present in the α1(I) chain or 20% glycosylated if present in the α2(I) chain to fulfil the criterion that all the peripheral molecules should have the diameter−limiting glycoside. Although purely arithmetical, these considerations do show that 8% glycosylation of Lys−756 of the α1(I) chain as found in the present work can lead to the occurrence of glycoside (galactosylhydroxylysine) at this position on one of the two α1(I) chains of 80% of the peripheral molecules of a corneal collagen fibril. A similar arithmetical treatment for the 17% glycosylated Lys−564 (just within an overlap zone) reveals that all the peripheral molecules could contain glycoside at this site with 70% of these (peripheral) molecules containing glycoside at position 564 in both α1(I) chains. It would seem therefore that the low level of glycosylation at particularly crucial sites could still play a significant role in limiting collagen fibril diameter. In considering the packing of 300nm−long, 1.4nm−thick collagen molecules

in a less 'ideal' manner, I would expect however that the limiting of collagen fibril diameter would be due to the combined effect of glycosides occurring in the overlap zones throughout the entire length of the type I collagen molecule.

The previous discovery, in corneal collagen, by Panjwani and Harding (1978b) of 54% glycosylation at Lys−531 in peptide α1−CB3 which is in an overlap zone, raises an important point with regard to the fibrillar location of hydroxylysine glycosides. Using the 'ideal', arithmetical treatment just described, the glycosylation of approximately 50% of positions 531 of collagen α1(I) chains means that even if each of the α1 chains of every peripheral molecule of a 23.8nm fibril (see fig. 1.13) were to contain a hydroxylysine glycoside residue at position 531, 60% of the glycosides attributed to Lys−531 would still have to be accomodated inside the fibril. (Even considering a less ideal and perhaps more realistic view of a fibril which contains more peripheral molecules resulting from an uneven fibril periphery as seen in cross−section unlike that shown in figure 1.13, it still seems that some glycoside may still have to be accomodated inside the fibril). This brings into question the notion that hydroxylysine glycosides limit fibril diameter solely by a steric−hindrance effect, as the hypothetical fibril just discussed should not be able to tolerate such a large excess of of overlap−zone glycosides within it.

With respect to the other ways in which hydroxylysine glycosides may limit collagen fibril diameter apart from its originally postulated steric effect, it is of interest that Eyre (1980) mentions that the more highly glycosylated type II collagen fibrils (compared to most fibrils of type I collagen) have been observed to be more swollen with water. Also, data derived from equatorial X−ray diffraction patterns of type II fibrils in lamprey notochord sheath indicate that the 'unit−cell' for lateral molecular packing is 40% larger than in type I collagen of rat tail tendon (Eikenberry et al., 1984), indicating that the collagen molecules are packed more loosely in lamprey notochord sheath collagen fibrils. The high level of glycosylation in this collagen was suggested as resulting in increased hydration and hence looser−packing of molecules (Brodsky and Eikenberry, 1985). The collagen fibrils were also determined as having

small, uniform diameters of about 17nm, not, therefore, unlike corneal collagen in this respect.

Parallel to these studies of collagen molecular packing and fibril diameters in native tissues, are the in vitro studies of collagen fibrillogenesis which indicate that saccharides inhibit fibril-formation by encouraging 'collagen-water' interactions over 'collagen-collagen' interactions (Hayashi and Nagai, 1972) especially with regard to a hydrophobic cluster in the C-telopeptide which is thought to facilitate lateral aggregation during fibrillogenesis (Lien et al., 1984). The evidence from these two lines of research therefore point to another mode in which hydroxylysine-linked glycosides could limit collagen fibril diameter in addition to a simple steric hindrance effect.

My view therefore, is that some glycoside can be accomodated inside the fibril although they increasingly disrupt collagen-collagen interactions. Thus, although glycosides in the overlap zones would be expected to have the greater limiting effect on collagen fibril diameter, an increased number of glycosides occurring in the gap zones may also increase the limiting effect on fibril diameter due to the hydrophilic nature of glycosides. All glycoside sites, regardless of their positions relative to the gap and overlap zones may facilitate the limitation of fibril diameter although they probably do so with varying effectiveness. The effects would probably be especially significant where they occur within the three-dimensional D-stagger, near the principal cross-linking sites Lys-9^{N} and Lys-17^{C}. Especially with regard to possible hydrophilic effects, this applies to the glycoside sites discovered in the present work at position 564 (17% glycosylated) which occurs in the immediate environment of Lys-17^{C} within the three dimensional stagger (see figs. 5.1 and 1.13). Glycoside occurring at position 564 may be particularly effective by promoting hydrophilic interactions in the region of the hydrophobic cluster of the C-telopeptide (see Lien et al., 1984).

The glycoside site at position 756 (8% glycosylated) occurring in the middle of an overlap zone could also be expected to present a hydrophilic effect as well as a purely steric effect in limiting collagen fibril diameter although the hydrophilic effect would

not be disrupting the environment around any known cross−link sites. The hydrophilic effect of the other glycoside sites at positions 573, 603, 648, 657 and 684 would be expected to be less than those already mentioned as these sites occur in a gap zone and sufficiently distant from the known principal cross−link sites for type I collagen. However, it should be remembered that other cross−link sites have been shown to be present in corneal collagen although their exact locations are not known (see Harding et al., 1980).

With regard to the *extent of lysyl hydroxylation* and the *nature and extent of hydroxylysyl glycosylation* at particular lysyl positions, it has been confirmed in the present work that in collagen sequences consisting of 'Gly−X−Y' triplets, only 'pre−Gly' lysyl positions are modified while none of the 'post−Gly' lysyl residues (three in peptide α1−CB7) are modified. However it was also shown the 'pre−Gly' lysyl residue at position 729 is not modified while the other 'pre−Gly' lysyl residues differed in their extent and nature of modifications (see figure 5.1). A focus on the amino acid sequences immediately surrounding the 'pre−Gly' lysyl residues (fig. 5.2) do not reveal any obvious pattern to indicate that the enzymes *lysyl hydroxylase* and the *hydroxylysyl glycosyltransferases* are more specific towards particular 'pre−Gly' lysyl residues depending on the immediate sequences surrounding the target residues. Previous characterisation of peptide α1−CB3 of bovine corneal collagen and primary sequence studies of the bovine α1(II) chain also did not reveal any obvious pattern surrounding target residues (Panjwani and Harding, 1978b; Francis et al., 1978). The greater susceptibility of some 'pre−Gly' lysyl positions such as position 684 to hydroxylation and glycosylation by the modification enzymes when compared to other positions such as position 729 may thus result from particular sequences occurring away from target positions and which participate in the binding of the enzymes to procollagen chains.

There still remains the question as to why lysyl residues of corneal type I collagen are more highly modified during biosynthesis than lysyl residues of other type I collagens such as tendon, skin and sclera. As lysyl hydroxylase and the

					% occurring as		
					Hyl*	Glycoside†	
						GGH	GH
	-Gly-Leu-Hyp-Gly-	-Pro-Lys-Gly-	-Asp-Arg-Gly-	-Asp-Ala-	23	7	10
	-Asp-Ala-	-Pro-Lys-	-Ala-Asp-	-Ala-Pro-	18	6	6
	-Ala-Hyp-	-Asp-Lys-	-Glu-Ala	-Pro-Ser-	15	4	6
(a)	-Gln-Hyp-	-Ala-Lys-	-Glu-Hyp-	-Asp-Ala-	15	6	3
	-Asp-Ala-	-Ala-Lys-	-Asp-Ala-	-Pro-Hyp-	9	3	2
	-Ala-Hyp-	-Pro-Hyl-	-Ala-Arg-	-Ser-Ala-	100	12	34
	-Pro-Ala-	-Glu-Lys-	-Ala-Hyp-	-Ala-Asp-	13	0	8
(b)	-Phe-Hyp-	-Pro-Lys-	-Ala-Ala-	-Glu-Hyp-	41	10	20
	-Asn-Asp-	-Ala-Lys-	-Asp-Ala-	-Ala-Hyp-	68	41	17
(c)	-Leu-Hyp-	-Met-Hyl-	-His-Arg-	-Phe-Ser-	~70‡	30	30
	-Leu-Hyp-	-Met-Hyl-	-His-Arg-	-His-Asn-	≳60	60	trace
	-Ala-Hyp-	-Pro-Hyl-	-Glu-Leu-	-Pro-Val-	≳60	60	trace
(a)	-Lys-Glu-	-Ser-Lys-	-Pro-Arg-	-Glu-Thr-	0	0	0
(d)	-Asp-Val-	-Glu-Lys-	-Pro-Glu-	-Ala-Pro-	0	0	0

Figure 5.2 Comparison of amino sequences adjacent to some lysyl residues in 'pre-Gly' positions with respect to the varying extents of lysyl modification: (a) in peptide α1-CB7 of bovine corneal collagen as deduced in this work; (b) in peptide α1-CB3 of bovine corneal collagen (Panjwani and Harding, 1978b); (c) Hyl-87, calf skin α1(I) chain, Hyl-87 and Hyl-174, calf skin α2(I) chain respectively (see Bornstein and Traub, 1979). (d) in the α1(II) chain of collagen from bovine nasal cartilage where most other 'pre-Gly' lysyl positions are highly modified (Francis et al., 1978).

**free (unglycosylated) hydroxylysine plus hydroxylysine glycoside*
†GGH, glucosylgalactosylhydroxylysine; GH, galactosylhydroxylysine.
‡Fietzek et al., 1973

glycosyltransferases are active towards procollagen chains before triple-helix formation but not towards the formed collagen triple-helix, it has been suggested that the slower rates of triple-helix formation for the collagen types II and IV compared to that for type I collagen partly explains the higher hydroxylysine and hydroxylysine glycoside contents of these collagens (see Kivirikko and Myllylä, 1982). The only forseeable way by which triple-helix formation could be slower for corneal (pro)collagen relative to other type I collagens such as tendon would be the occurrence of a relatively low prolyl-4-hydroxylase activity in the cornea, as sufficient hydroxylation of proline is necessary for formation of a stable triple-helix to take place. The activity of prolyl hydroxylase in corneal tissue has not been reported but even if the activity is lower, a slower rate of triple-helix formation need not be the sole factor responsible for the greater extent of lysyl modification. It has been shown for example that triple-helix formation in cartilage cells producing type II procollagen is only 30% slower than in tendon cells producing type I procollagen which does not wholly account for the several-fold increase in lysyl modification observed for type II collagen as compared to type I collagen (Keller et al., 1985).

Increased modification of lysyl residues may also be attributed to an increased ratio of lysyl hydroxylase and the galactosyltransferases to the procollagen chain substrate. This has been found to be the case in virus-transformed human and chick-embryo cells (Myllylä et al., 1981) in which a decrease in the rate of collagen synthesis (translation of mRNA as opposed to triple-helix formation) was responsible for the increase in the ratio of enzyme to substrate. It is not known whether this would apply to the biosynthesis of corneal collagen.

The most likely reason for an increase in the ratio of enzyme to substrate would be if there are increased activities of the enzymes. Higher activities of lysyl hydroxylase and galactosyltransferase have been observed in cartilage cells when compared to tendon and bone cells and also in foetal cells compared to adult cells (see Introduction) which is consistent with the higher modification of lysyl residues observed for type II collagen and collagen of young tissues. Of special significance is

the finding that in normal, healthy individuals, corneal fibroblasts (keratocytes) contains approximately twice the lysyl hydroxylase activity found in skin fibroblasts (Risteli et al., 1980). As fibroblasts in both tissues synthesise mainly type I collagen, the higher lysyl hydroxylase activity in corneal fibroblasts is thus a likely cause of the higher content of hydroxylysine and hydroxylysine-linked glycoside in corneal type I collagen when compared to skin type I collagen as has been shown in this work and by other authors. The generally larger differences found for lysyl hydroxylase activity between tissues (of the same age) when compared to the differences found for glycosyltransferase activities may point to lysyl hydroxylation as being a possible controlled step during the formation of hydroxylysine glycoside from lysine. The higher levels of enzyme activity such as found in the cornea is most probably due to a simple increase in the amount of the enzyme, the alternative explanation being the occurrence of an isoenzyme with a higher specific activity. This possibility has been considered in trying to explain the different extents of lysyl modification in collagen derived from different tissues (see Risteli et al., 1980) but evidence for the existence of isoenzymes of lysyl hydroxylase and glycosyltransferases is not forthcoming. However, from observations indicating that highly purified lysyl hydroxylase failed to hydroxylate lysine occurring in the telopeptide regions of chick type I collagen, Royce and Barnes (1985) have suggested the existence of a lysyl hydroxylase specific for lysyl residues in the non-helical telopeptide regions which is distinct from that active towards lysine in the helical regions.

The different lines of research just discussed, covering studies of lysine modifying enzymes, procollagen triple-helix formation, fibril morphology and sites of glycosylation in the primary structure of collagen point to a broad interest in lysyl hydroxylation and hydroxylysyl glycosylation plus the possible role of hydroxylysine glycosides in controlling collagen fibril diameter. These studies are of great significance to enable a better understanding of how type I collagen, a protein which accounts for over 20% of body protein achieves its diversity in functioning as the body's molecular "scaffold" in a wide variety of tissues as well as the changes that occur in collagen molecular and

fibrillar structure during development and in diseased states. For example, the tensile strength of collagen fibrils are highly dependent on collagen fibril diameter (Parry et al., 1979) while the fragility of bones in osteogenesis imperfecta patients is consistent with the observation of abnormally thin collagen fibrils as well as high levels of lysyl modification in collagen from the skin of osteogenesis imperfecta patients (see Chapter 1.IIF).

The special significance of these studies with respect to the cornea is the apparent necessity for corneal transparency, of collagen fibrils which are thin and of uniform diameter within the corneal stroma. It would be interesting therefore to study the cornea with respect to molecular packing within fibrils, rate of triple-helix formation during biosynthesis, rate of collagen synthesis and levels of the enzymes responsible for the post-translational modification of lysine and proline. With regard to studies of glycoside sites in the primary structure of corneal type I collagen, the studies focussing on peptide α1-CB7 as well as the less-detailed characterisation of peptides α1-CB8 and α1-CB3 in this work indicate that the increased lysyl hydroxylation and hydroxylysyl glycosylation of corneal type I collagen in comparison to skin type I collagen seem to occur throughout most of the 'pre-Gly' lysyl positions in the α1(I) chain.

REFERENCES

Aguilar, J.H., Jacobs, H.G., Butler, W.T. *et al.* (1973) J.Biol. Chem. 248, 5106.

Allmann, H., Fietzek, P.P., Glanville, R.W. *et al.* (1979) Hoppe−Seyler's Z. Physiol. Chem. 360, 861.

Anttinen, H., Oikarinen, A. and Kivirikko, K.I. (1977) Clin. Chim. Acta 76, 95.

Arvidson, S. (1973) Biochim. Biophys. Acta 302, 149.

Ayad, S., Chambers, C.A., Shuttleworth, A. *et al.* (1985) Biochem. J. 230, 465.

Ayad, S., Chambers, C.A., Berry, L. *et al.* (1986) Biochem. J. 236, 299.

Ayad, S. and Weiss, J.B. (1984) Biochem J. 218, 835.

Axelsson, I. and Heinegard, D. (1975) Biochem. J. 145, 491.

Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545.

Bachinger, H.P., Fessler, L.I., Timpl, R. *et al.* (1981) J. Biol. Chem. 256, 13193.

Barnes, M.J., Constable, B.J., Morton, L.F. *et al.* (1974) Biochem J. 139, 461.

Barsh, G.S., Roush, C.L., Bonadio, J. *et al.* (1985) Proc. Natl. Acad. Sci. 82, 2870.

Benedek, G.B. (1983) Nature 302, 383.

Benya, P.D. and Padilla, S.R. (1986) J. Biol. Chem 261, 4160.

Berg, R.A. and Prockop, D,J. (1973) Biochem. Biophys. Res. Commun. 52, 115.

Bergmann, M., Fruton, J.S. and Pollok, H. (1939) J. Biol. Chem. 127, 643.

Birk, D.E. and Lande, M.A. (1981) Biochim. Biophys. Acta. 670, 362.

Birk, D.E. and Trelstad, R.L., (1984) J. Cell.Biol. 99, 2024.

Blumenkrantz, N., Assad, R. and Peterkofsky, B. (1984) J. Biol. Chem. 259, 854.

Bocherding, M.S., Blacik, J.L., Sittig, R.A. *et al.* (1975) Exp. Eye. Res. 21, 59.

Bornstein, P. and Piez, K.A. (1966) Biochemistry 5, 3460.

Bornstein, P. and Sage, H. (1980) Annu. Rev. Biochem. 49, 957.

Bornstein, P. and Traub, W. (1979) in "The Proteins" (Neurath, H. and Hill, R.L. eds.) Vol. IV, p. 411, Academic Press, London.

Brodsky, B. and Eikenberry, E.F. (1982) Methods in Enzymology 82, 127.

Brodsky, B. and Eikenberry, E. (1985) Ann. N.Y. Acad. Sci. 460, 73.

Brodsky, B., Eikenberry, E.F., Belbruno, K.S. *et al.* (1982) Biopolymers 21, 935.

Broek, D.L., Madri, J., Eikenberry, E.F. *et al.* (1985) J. Biol. Chem. 260, 555.

Brokaw, J.L., Doillon, C.J., Hahn, R.A. *et al.* (1985) Int. J. Biol. Macromol. 7, 135.

Bruckner, P. and Eikenberry, E.F. (1984) Eur. J. Biochem. 140, 397.

Bruckner, P., Vaughan, L. and Winterhalter, K.H. (1985).
Proc. Natl. Acad. Sci. USA 82, 2608

Burgeson, R.E. and Hollister, D.W. (1979)
Biochem. Biophys. Res. Commun. 67, 1124.

Burke, J.M., Balian, G., Ross, R. *et al.* (1977) Biochemistry 16, 3243.

Butkowski, R.J., Noelken, M.E. and Hudson, B.G. (1982)
Methods in Enzymology 82, 410.

Butler, P.J.G., Harris, J.I., Hartley, B.S. *et al.* (1967) Biochem J. 103, 78P.

Butler, W.T. (1982) Methods in Enzymology 82, 339.

Butler, W.T. and Cunningham, L.W. (1966) J. Biol. Chem. 241, 3882.

Butler, W.T., Piez, K.A. and Bornstein, P. (1967) Biochemistry 6, 3771.

Butler, W.T., Finch, J.E., Jr. and Miller, E.J. (1977) Biochemistry 16, 4981.

Byers, P.H. and Holbrook, K.A. (1985) Ann. N.Y. Acad. Sci. 460, 298.

Cannon, D.J. and Cintron, C. (1975) Biochim. Biophys. Acta 412, 18.

Capaldi, M.J. and Chapman, J.A. (1982) Biopolymers 21, 2291.

Cetta, G., Tenni, R., Zanaboni, G. *et al.* (1982) Biochem. J. 204, 61.

Chandrasekhar, S., Kleinman, H.K., Hassell, J.R. *et al.* (1984)
Collagen Relat. Res. 4, 323.

Chapman, J.A. and Hardcastle, R.A. (1974) Connect. Tissue Res. 2, 151.

Cheah, K.S.E. (1985) Biochem. J. 229, 287.

Chew, M.K. and Squire, J.M. (1986) Int. J. Biol. Macromol. 8, 27.

Chu, M−L., de Wet, W., Bernard, M. *et al.* (1984) Nature 310, 337.

Church, R.L. (1981) Int. Rev. Connect. Tissue Res. 9, 99.

Cintron, C. (1974) Biochem. Biophys. Res. Commun. 60, 288.

Craig, A.S.and Parry, D.A.D. (1981) J. Ultrastruct. Res. 74, 232.

Croft, L.R. (1972) Biochem. J. 128, 961.

Croft, L.R. (1980) "Handbook of Protein Sequence Analysis",
John Wiley & Sons, Chichester.

Cunningham, L.W. and Fredriksen, D.W., eds. (1982) Methods in Enzymology 82.

Dabbous, M.K., Hammouda, O. and Brinkley, Sr. B. (1981)
Mol. Cell. Biochem. 34, 87.

Davison, P.F. (1978) J. Biol. Chem. 253, 5635.

de Luca, G., Tenni, R., Lauria, A. *et al.* (1983) Ital. J. Biochem. 32, 418.

Denholm, L.J., Hall, C.E., Minov, R.R. *et al.* (1985) Ann. N.Y. Acad. Sci. 460, 412.

Dickson, L.A., Pihlajaniemi, T., Deak, S. *et al.* (1984)
Proc. Natl. Acad. Sci. USA 81, 4524.

Dixit, S.N., Seyer, J.M., Oronsky, A.O. *et al.* (1975) Biochemistry 14, 1933.

Dixit, R., Harrison, M.W. and Dixit, S.N. (1985)
Biochem. Biophys. Res. Commun. 130, 1.

Dixon, H.B.F. and Perham, R.N. (1968) Biochem. J. 109, 312.

Donaldson, D.D. (1966) A.M.A. Arch. Ophthal. 76, 25.

Drapeau, G.R. (1977) Methods in Enzymology 47, 189.

Drapeau, G.R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. 247, 6720.

Eikenberry, E.F., Brodsky, B.B., Craig, A.S. *et al.* (1982)
Int. J. Biol. Macromol. 4, 393.

Eikenberry, E.F., Childs, B., Sheren, S.B. *et al.* (1984) J. Mol. Biol. 176, 261.

Eyre, D.R. (1980) Science 207, 1315.

Eyre, D.R., Brickley–Parsons, D.M. and Glimcher, M.J. (1978) FEBS Lett. 85, 259.

Eyre, D.R., Paz, M.A. and Gallop, P.M. (1984) Annu. Rev. Biochem. 53, 717.

Eyre, D.R., Shapiro, F.D. and Aldridge, J.F. (1985) J. Biol. Chem. 260, 11322.

Farrell, R.A., McCally, R.L. and Tatham, P.E.R. (1973) J. Physiol. 233, 589.

Fessler, J.H. and Fessler, L.I. (1978) Annu. Rev. Biochem. 47, 129.

Fessler, J.H., Doege, K.J., Duncan, K.G. *et al.* (1985) J. Cell. Biochem. 28, 31.

Fietzek, P.P. and Kühn, K. (1975) Mol. Cell. Biochem. 8, 141.

Fietzek, P.P.and Kühn, K. (1976) Int. Rev. Connect. Tissue Res. 7, 1.

Fietzek, P.P., Wendt, P., Kell, I. *et al.* (1972) FEBS Lett. 26, 74.

Fietzek, P.P., Rexrodt, F.W., Hopper, K.E. *et al.* (1973) Eur. J. Biochem. 38, 396.

Fitch, J.M., Gross, J., Mayne, R. *et al.* (1984) Proc.Natl. Acad. Sci. USA. 81, 2791.

Fleischmajer, R., Timpl, R., Tuderman, L. *et al.* (1981)
Proc. Natl. Acad. Sci. USA 78, 7360.

Fleischmajer, R., Olsen, B.R., Timpl, R. *et al.* (1983)
Proc. Natl. Acad. Sci. USA 80, 3354.

Fleischmajer, R., Olsen, B.R. and Kühn, K., eds. (1985) Ann. N.Y. Acad. Sci. 460.

Francis, G., Butler, W.T. and Finch, J.E., Jr. (1978) Biochem.J. 175, 921.

Fraser, R.D.B., McRae, T.P., Miller, A. *et al.* (1983) J. Mol. Biol. 167, 497.

Fujimoto, D., Moriguchi, T., Ishida, T. *et al.* (1978)
Biochem. Biophys. Res. Commun. 84, 52.

Furthmayr, H. and Timpl, R. (1971) Anal. Biochem. 41, 510.

Glanville, R.W., Breitkreutz, D., Meitinger, M. *et al.* (1983) Biochem. J. 215, 183.

Glanville, R.W., Qian, R−Q., Siebold, B. *et al.* (1985) Eur. J. Biochem. 152, 213.

Goldman, J.N. and Benedek, G.B. (1967) Invest. Ophthalmol. 6, 574.

Goodfellow, J.M., Elliot, G.F. and Woolgar, A.E. (1978) J.Mol. Biol. 119, 237.

Grant, M.E. and Jackson,D.S. (1976) in "Essays in Biochemistry" (Campbell, P.N. and Aldridge, W.N., eds.) vol. 12, p. 77, Academic Press, London.

Grant, M.E., Heathcote, J.G. and Orkin, R.W. (1981) Bioscience Reports 1, 819.

Gray, W.R. (1972) Methods in Enzymology 25B, 121.

Hale, P.N. and Maurice, D.M. (1969) Exp. Eye Res. 8, 205.

Hamazaki, H. and Hotta, K. (1979) J. Biol. Chem. 254, 9682.

Hames, B.D. (1981) in "Gel Electrophoresis of Proteins: A Practical Approach" (Hames, B.D. and Rickwood, D., eds.) p. 1, IRL Press, Eynsham.

Harding, J.J. (1965) Advan. Prot. Chem. 20, 109.

Harding, J.J. (1985) Advan. Prot. Chem. 37, 247.

Harding, J.J. and Crabbe, M.J.C. (1979) FEBS Lett. 100, 351.

Harding, J.J. and Panjwani, N.A. (1980)
Int. Congr. Symp. Ser. Roy. Soc. of Med. 40, 209.

Harding, J.J. and Wesley, J.M. (1968), Biochem. J. 106, 749.

Harding, J.J., Crabbe, M.J.C. and Panjwani, N.A. (1980)
Colloques Internationaux du CNRS (Paris) 287, 51.

Harkness, R.D. (1961) Biol. Rev. 36, 399.

Harwood, R. (1980) in "The Enzymology of Post−translational Modifications of Proteins" (Freedman, R.B. and Hawkins, H.C., eds.) Vol 1, p. 3, Academic Press, London.

Hascall, V.C. and Kimura, J.H. (1982) Methods in Enzymology 82, 769.

Hassell, J.R., Cintron, C., Kublin, C. *et al.* (1983)
Arch. Biochem. Biophys. 222, 362.

Hayashi, T. and Nagai, Y. (1972) J. Biochem. (Tokyo) 72, 749.

Heathcote, J.G. and Grant, M.E. (1980) in "The Enzymology of Post-translational Modifications of Proteins" (Freedman, R.B. and Hawkins, H.C., eds.) Vol 1 p. 457.

Heidemann, E. and Linnert, N. (1984) Hoppe-Seyler's Z. Physiol. Chem. 365, 781.

Helseth, D. L. and Veis, A. (1981) J. Biol. Chem. 256, 7118.

Henkel, W. and Glanville, R.W. (1982) Eur. J. Biochem. 122, 205.

Hennecke, M. and Plapp, B.V. (1984) Anal. Biochem. 136, 110.

Herbage, D., Bouillet, J. and Bernengo, J.C. (1977) Biochem. J. 161, 303.

Highberger, J.H., Corbett, C., Kang, A.H. *et al.* (1975) Biochemistry 14, 2872.

Highberger, J.H., Corbett, C., Kang, A.H. *et al.* (1978)
Biochem. Biophys. Res. Commun. 83, 43.

Highberger, J.H., Corbett, C., Dixit, S.N. *et al.* (1982) Biochemistry 21, 2048.

Hill, R. (1965) Advan. in Prot. Chem. 20, 37.

Hirs, C.H.W., Moore, S. and Stein, W.H. (1956) J. Biol. Chem. 219, 623.

Hodson, S. (1977) Invest. Ophthal. Vis. Sci. 16, 589.

Hodson, S. and Wigham, C. (1983) J. Physiol. 342, 409.

Hofmann, H., Fietzek, P.P. and Kühn, K. (1978) J. Mol. Biol. 125, 137

Hofmann, H., Fietzek, P.P. and Kühn, K. (1980) J. Mol. Biol. 141, 293.

Hofmann, H., Voss, T. and Kühn, K. (1984) J. Mol. Biol. 172, 325.

Houmard, J. and Drapeau, G.R. (1972) Proc. Natl. Acad. Sci USA 69, 3506.

Hulmes, D.J.S. (1983) Collagen Relat. Res. 3, 317.

Hulmes, D.J.S. and Miller, A. (1979) Nature 282, 878.

Hulmes, D.J.S., Miller, A., White, S.W. *et al.* (1980) Int. J. Biol. Macromol. 2, 338.

Itoi, M. (1961) Exp Eye Res. 1, 92.

Jakus, M.A. (1961) in "The Structure ofthe Eye" (Smelser, G.K., ed.) p. 343, Academic Press, New York.

Jakus, M.A. (1964) "Ocular Fine Structure" Churchill, J.A. Ltd, London.

Kang, A.H., Piez, K.A. and Gross, J. (1969) Biochemistry 8, 1506.

Kang, A.H., Dixit, S.N., Corbett, C. *et al.* (1975) J. Biol. Chem. 250, 7428.

Kao, W. W−Y. (1985) Curr. Eye. Res. 4, 79.

Kao, W. W−Y. and Foreman, C.A. (1980) Eur. J. Biochem. 106, 41.

Keller, H., Eikenberry, E.F., Winterhalter, K.H. *et al.* (1985) Collagen Relat. Res. 5, 245.

Kenney, M.C., Lewis, D.A., Atton, R.J. *et al.* (1982) Connect. Tissue Res. 10, 127.

Kent, M.J.C., Light, N.D. and Bailey, A.J. (1985) Biochem. J. 225, 745.

Kielty, C.M., Kwan, A.P.L., Holmes, D.F. *et al.* (1985) Biochem. J. 27, 545.

Kimura, S. and Karasawa, K. (1985) Comp. Biochem. & Physiol. 81B, 361.

Kinoshita, J.H. (1962) Invest. Ophthalmol. 1, 178.

Kirsch,E., Glanville, R.W., Krieg, T. *et al.* (1983) Biochem. J. 211, 599.

Kivirikko, K.I. and Myllylä, R. (1979) Int. Rev. Connect. Tissue Res. 8, 23.

Kivirikko, K.I. and Myllylä, R. (1980) in "The Enzymology of Post−translational Modification of Proteins" (Freedman, R.B. and Hawkins, H.C., eds.) Vol 1 p. 53, Academic Press, London.

Kivirikko, K.I. and Myllylä, R. (1982) in "Collagen in Health and Disease" (Weiss, J.B. and Jayson, M.I.V., eds.) p. 101, Churchill Livingstone, Edinburgh.

Kivirikko, K.I. and Myllylä, R. (1985) Ann. N.Y. Acad. Sci. 460, 187 .

Kleinman, H.K., McGoodwin, E.B., Martin, G.R. *et al.* (1978) J. Biol. Chem. 253, 5642.

Kleinman, H.K., Klebe, R.J. and Martin, G.R. (1981a) J. Cell. Biol. 88, 473.

Kleinman, H.K., Wilkes, C.M. and Martin, G.R. (1981b) Biochemistry. 20, 2325.

Krause, A.C. (1934) "The Biochemistry of the Eye", John Hopkins Press, Baltimore.

Kuhlman, R.E. and Resnik, R.A. (1959) Arch. Biochem. Biophys. 85, 29.

Kühn, K., Grassman, W. and Hofmann, U. (1960) Naturvissen−Schaften 47, 258.

Kurkinen, M., Bernard, M.P., Barlow, D.P. *et al.* (1985) Nature. 317, 177.

Laemmli, U.K. and Favre, M. (1973) J. Mol. Biol. 80, 575.

Langham, M.E. and Taylor, I.S. (1956) Brit. J. Ophthal. 40, 321.

Lapière, C.M., Nusgens, B. and Pierard, G.E. (1977) Connect. Tissue Res. 5, 21.

Laurent, G.J., Cockerill, P., McAnulty, R.J. *et al.* (1981) Anal. Biochem. 113, 301.

Lee, R.E. and Davison, P.F. (1984) Exp. Eye Res. 39, 639.

Lees, S. (1986) Int. J. Biol. Macromol. 8, 66.

Lenaers, A., Ansay, M., Nusgens, B.V. *et al.* (1971) 23, 533.

Lewis, M.S., Dubin, M.W. and Aandahl, V.A. (1967) Exp. Eye Res. 6, 57.

Lien, Y−H., Stern, R., Fu, J.C.C. *et al.* (1984) Science 225, 1489.

Light, N.D. and Bailey, A.J. (1979) FEBS Lett. 97, 183.

Light, N.D. and Bailey, A.J. (1980) in " Biology of Collagen"
(Viidik, A.,and Vuust, J., eds.) p. 15, Academic Press, London.

Linsenmayer, T.F., Fitch, J.M., Schmid, T.M. *et al.* (1983) J. Cell. Biol. 96, 124.

Lozano, G., Ninomiya, Y.,Thompson, H. *et al.* (1985)
Proc. Natl. Acad. Sci. USA 82, 4050.

Margoliash, E., Smith, E.L., Kreil, G. *et al.* (1961) Nature 192, 1125.

Martin, G.R., Timpl, R., Müller, P.K. *et al.* (1985)
Trends In Biochemical Sciences, 10, 285.

Martola, E.L. and Baum, J.L. (1968) A.M.A. Arch. Ophthal. 79, 28.

Maurice, D.M. (1957) J. Physiol. 136, 263.

Maurice, D.M. (1984) in "The Eye" (Davson, H., ed.) Vol 1B p. 1,
Academic Press, London.

Maurice, D.M. and Giardini, A.A. (1951a) Brit.J. Ophthal 35, 169.

Maurice, D.M. and Giardini, A.A. (1951b) Brit. J. Ophthal 35,791.

Maurice, D.M. and Riley, M.V. (1970) in " Biochemistry of the Eye"
(Graymore, C.N., ed.) p. 1, Academic Press, London.

Mayes, K.R. and Hodson, S. (1979) Exp. Eye Res. 28, 699.

Mays, C. and Rosenberry, T.L. (1981) Biochemistry 20, 2810.

McLaughlin, B.J., Caldwell, R.B., Sasaki, Y. *et al.* (1985) Curr. Eye Res. 4, 951.

Meek, K.M., Elliot, G.F., Sayers, Z. *et al.* (1981) J. Mol. Biol. 149, 477.

Miller, A. (1982) Trends In Biochemical Sciences 7, 13.

Miller, E.J. (1985) Ann. N.Y. Acad. Sci. 460, 1.

Miller, E.J. and Gay, S. (1982) Methods in Enzymology 82, 1.

Miller, E.J. and Lunde, L.G. (1973) Biochemistry 12, 3153.

Miller, E.J. and Rhodes, R.K. (1982) Methods in Enzymology 82, 33.

Mishima, S. and Hedbys, B.O. (1967) Exp. Eye Res. 6, 10.

Mishima, S. and Hedbys, B. O. (1968) in "Corneal Edema" (Dohlman, C.H., ed.).

Mishima, S. and Kudo, T. (1967) Invest. Ophthalmol. 6, 329.
Morgan, P. H., Jacobs, H.G., Segrest, J.P. *et al.* (1970) 245, 5042. *J. Biol. Chem*

Müller, P.K., Kirsch, E., Gauss–Müller, V. *et al.* (1981) Mol. Cell. Biochem. 34, 73

Murray, L.W., Waite, J.H., Tanzer, M.L. *et al.* (1982)
Methods in Enzymology 82, 65.

Myers, J.C., Loidl, H.R., Stolle, C.A. *et al.* (1985a) J. Biol. Chem. 260, 5533.

Myers, J.C., Lodl, H.R., Seyer, J.M. *et al.* (1985b) J. Biol. Chem. 260, 11216.

Myllylä, R., Alitalo, K., Vaheri, A. *et al.* (1981) Biochem. J. 196, 6831

Nakayasu, K., Tanaka, M., Konomi, H. *et al.* (1986)
Ophthalmic Research 18, 1.

Neuberger, A. and Slack, H.G.B. (1953) Biochem. J. 53, 47.

Newsome, D.A., Gross, J. and Hassell, J.R. (1982) Invest.Ophthal.Vis. Sci. 22, 376.

Odell, V., Wegener, L., Peczon, B. *et al.* (1974) J. Chrom. 88, 245.

O'Driscoll, S.W., Salter, R.B. and Keeley, F.W. (1985) Anal. Biochem 145, 277.

Oegema, T.R., Jr., Laidlaw, J., Hascall, V.C. *et al.* (1975)
Arch. Biochem. Biophys. 170, 698

Oikarinen, A. Anttinen, H. and Kivirikko, K.I. (1976) Biochem. J. 156, 545.

Olsen, B.R., Ninomiya, Y., Lozano, G. *et al.* (1985) Ann. N.Y. Acad. Sci. 460, 141.

Panjwani, N.A. and Harding, J.J. (1978a) Biochem. J. 171, 687.

Panjwani, N.A. and Harding, J.J. (1978b) Biochem J. 171, 679 .

Parry, D.A.D.and Creamer, L.K., eds. (1979) "Fibrous Proteins: Scientific, Industrial and Medical Aspects" Vols. 1 and 2, Academic Press (London).

Parry, D.A.D., Craig, A.S. and Barnes, G.R.G. (1979)
in the reference above, Vol 2, p. 77p

Peltonen, L., Halila, R. and Ryhanen (1985) J. Cell. Biochem. 28, 15.

Petruska, J.A. and Hodge, A.J. (1964) Proc. Natl. Acad. Sci USA 51, 871.

Piez, K.A. and Miller, A. (1974) J. Supramol. Structure 2, 121.

Piez, K.A. and Trus, B.L., Bioscience Reports 1, (1981).

Piez, K.A., Eigner, E.A. and Lewis, M.S. (1963) Biochemistry 2, 58.

Poole, A.R., (1986) Biochem. J. 236, 1.

Praus, R., Brettschneider, I., and Adam, M. (1979) Exp. Eye Res. 29, 469.

Prockop, D.J. and Kivirikko, K.I. (1984) N. Engl. J. Med. 311, 376.

Prockop, D.J., Kivirikko, K.I., Tuderman, L. *et al.* (1979)
N. Engl. J. Med 301, 13.

Ramachandran, G.N. and Ramakrishnan, C. (1976) *in the next reference,* p. 45

Ramachandran, G.N. and Reddi, A., eds. (1976) "Biochemistry of Collagen"
Plenum Press, New York.

Rauterberg, J. and Kühn, K. (1971) Eur. J. Biochem. 19, 398.

Reid, K.B.M. and Porter, R.R. (1981) Annu. Rev. Biochem. 50, 433.

Rexrodt, F.W., Hopper, K.E., Fietzek, P.P. *et al.* (1973) Eur. J. Biochem. 38, 384.

Riley, M.V. (1969) Exp. Eye Res. 8, 193.

Riley, M.V. (1977) Exp. Eye Res. 24, 35.

Risteli, L., Risteli, J., Salo, L. *et al.* (1979) Eur. J. Biochem. 97, 297.

Risteli, L., Risteli, J., Ihme, A. *et al.* (1980)
Biochem. Biophys. Res. Commun. 96, 1178.

Robins, S.P. and Bailey, A.J. (1977) Biochem. J. 163, 339.

Royce, P.M. and Barnes, M.J. (1977) Biochim. Biophys. Acta 498, 132.

Royce, P.M. and Barnes, M.J. (1985) Biochem. J. 130, 475.

Runnegar, B. (1985) J. Mol. Evol. 22, 14.

Sage, H. and Bornstein, P. (1979) Biochemistry 18, 3815.

Sarkar, S.K., Young, P.F., Sullivan, C.E. *et al.* (1984)
Proc. Natl. Acad. Sci. USA 81, 4800.

Sayers, Z., Koch, M.H.J., Whitburn, S.B. *et al.* (1982) J. Mol. Biol. 160, 593.

Schmitt, F.O, Gross, J. and Highberger, J.H. (1955) Exp. Cell. Res. Suppl. 3, 326.

Schofield, J.D., Freeman, I.L. and Jackson, D.S. (1971) Biochem. J. 124, 467.

Schroeder, W.A. (1967) Methods in Enzymology 11, 351.

Schuppan, D., Glanville, R.W. and Timpl, R. (1982) Eur. J. Biochem. 123, 505.

Schwarz, W. and Graf Keyserlingk, D. (1969) in "The Cornea" (Langham, M., ed.),
p. 123, The John Hopkins Press Ltd., London.

Scott, J.E. (1984) Biochem. J. 218, 229.

Scott, J.E. and Haigh, M. (1985) Bioscience Reports 5, 765.

Scott, J.E. and Orford, C.R. (1981) Biochem. J. 197, 213.

Scott, P.G and Edwards, H. (1981) Developments in Biochemistry 22, 45.

Scott, P.G. and Veis, A. (1976a) Connect. Tissue Res. 4, 107.

Scott, P.G. and Veis, A. (1976b) Connect. Tissue Res. 4, 117.

Seyer, J.M. and Kang, A.H. (1977) Biochemistry 16, 1158.

Shinkai, H. and Yonemasu, K. (1979) Biochem. J. 177, 847.

Siegel, R.C. (1979) Int. Rev. Connect. Tissue Res. 8, 73.

Slaby, I. and Holmegren, A. (1975) J. Biol. Chem. 250, 1340.

Smelser, G.K., Polack, F.M. and Ozanics, V. (1965) Exp. Eye Res. 4, 349.

Smith, J.W. (1968) Nature 219, 157.

Spiro, R.G. (1967) J. Biol. Chem. 242, 4813.

Spiro, R.G. (1969) J. Biol. Chem. 244, 602.

Spiro, R.G. and Spiro, M.J. (1971) J. Biol. Chem. 246, 4919.

Steinmann, B., Rao, V.H., Vogel, A. *et al.* (1984) J. Biol. Chem. 259, 11129.

Sternberg, M. and Spiro, R.G. (1979) J. Biol. Chem. 254, 10329.

Sykes, B. (1985) Bioessays 3, 111.

Sykes, B.C. and Bailey, A.J. (1971) Biochem. Biophys. Res. Commun. 43, 340.

Sykes, B. and Smith, R. (1985) Quarterly Journal of Medicine, New Series 56, 533.

Sykes, B., Puddle, B. and Francis, M.J.O. (1976)
Biochem. Biophys. Res. Commun. 72, 1472.

Sykes, B., Francis, M.J.O. and Smith, R. (1977) N. Engl. J. Med. 296, 1200.

Sykes, B., Ogilvie, D., Wordsworth, P. *et al.* (1986) Lancet Vol. II, p. 69.

Tang, P.W. and Williams, J.M. (1984) Anal. Biochem. 142, 37.

Tate, V., Finer, M., Boedtker, H. *et al.* (1982)
Cold Spring Harb. Symp. Quant. Biol. 47, 1039.

Traub, W. and Steinmann, B. (1986) FEBS Lett. 198, 213.

Trelstad, R.L. and Hayashi, K. (1979) Dev. Biol. 71, 228.

Trenberth, S.M. and Mishima, S. (1968) Invest. Ophthalmol. 7, 44.

Tryggvason, K., Risteli, J. and Kivirikko, K.I. (1977)
Biochem. Biophys. Res. Commun. 76, 275.

Tseng, S.C.G., Smuckler, D. and Stern, R. (1982) 257, 2627. J. Biol. Chem.

van der Rest, M, Rosenberg, L.C., Olsen, B.R., *et al.* (1986) Biochem. J. 237, 923.

von der Mark, K., von der Mark, H., Timpl, R. *et al.* (1977) Dev. Biol. 59, 75.

von der Mark, K. (1981) Int. Rev. Connect. Tissue Res. 9, 265.

Viidik, A. and Vuust, J., eds. (1980) "Biology of Collagen", Academic Press, London.

Veis, A. and Yuan, L. (1975) Biopolymers 14, 895.

Veis, A., Leibovich, S.J., Evans, J., *et al.* (1985). Proc. Natl. Acad. Sci. USA 82, 3693

Volpin, D. and Veis, A. (1973) Biochemistry 12, 1452.

Vuust, J. and Piez, K.A. (1972) J. Biol. Chem. 247, 856.

Weiss, J.B. and Jayson, M.I.V., eds. (1982) "Collagen in Health and Disease", Churchill Livingstone, Edinburgh.

Weissman, B.A., Fatt, I. and Rasson, J. (1981) Invest. Ophthal. Vis. Sci. 20, 123.

Welsh, C., Gay, S., Rhodes, R.K., *et al.* (1980) Biochim. Biophys. Acta 625, 78.

Wood, G.C. and Keech, M.K. (1960) Biochem. J. 75, 588.

Woodhead–Galloway, J. (1980) "Collagen: The Anatomy of a Protein", *Studies in Biology* series, no. 117.

Woods, K.R. and Wang, K.T. (1967) Biochim. Biophys. Acta 133, 369.

Yamada, Y., Liau, G., Mudryj, M., *et al.* (1984) Nature 310, 333.

Zimmerman, D.R., Trueb, B., Winterhalter, K.H., *et al.* (1986) FEBS Lett. 197, 55.

ADDITIONS

Byers, P.H., Shapiro, J.R., Rowe, D.W. *et al.* (1983) J. Clin. Invest. 71, 689.

Farber, S., Garg, A.K., Birk, D.E. *et al.* (1986) Int. J. Biol. Macromol. 8, 37.

Fietzek, P.P., Allman, H., Rauterberg, J. *et al.* (1979) Hoppe–Seyler's Z. Physiol. Chem. 360, 809.

Goldberger, R.F. and Anfinsen (1962) Biochemistry 1, 401.

Goldman, J.N., Benedek, G.B., Dohlman, C.H. *et al.* (1968) Invest. Ophthalmol. 7, 501.

Kühn, K., Glanville, R.W., Babel, W. *et al.* (1985) Ann. N.Y. Acad. Sci. 460, 14

Meek, K.M. and Holmes, D.F. (1983) Int. J. Biol. Macromol. 5, 17.

Ryhänen, L. and Kivirikko, K.I. (1974) Biochim. Biophys. Acta 343, 121.

Spiro, R.G. (1972) Methods in Enzymology 28B, 3.

Sutton, M.R., Fall, R.R., Nervi, A.M. *et al.* (1977) J. Biol. Chem. 252, 3934.

Vuust, J., Abildstein, D. and Lund, T. (1983) Connect. Tissue Res. 11, 185.

Wood, G.C. (1960) Biochem. J. 75, 598.

APPENDIX

Amino acid sequences available for (pro)α chains of collagen types I, II, III, IV, V and IX : references to be used in conjunction with figure 1.10

1. Galloway, D., (1982) in "Collagen in Health and Disease"
(Weiss, J.B. and Jayson, M.I.V. eds.), p528, Churchill−Livingstone.

2. Bornstein, P. and Traub, W. (1979) in "The Proteins", vol. IV
(Neurath, H. and Hill, R. eds.), p 411, Academic Press, London.

3. Fietzek, P.P. and Kühn, K. (1976). Int. Rev. Connect. Tiss. Res. 7, 1.

4. Fietzek, P.P. and Kühn, K. (1975). Mol. Cell. Biochem. 8, 141.

5. Hulmes, D.J.S., Miller, A., Parry, D.A.D. *et al.* (1973). J. Mol. Biol. 79, 137.

6. Highberger. J.H., Corbett, C., Dixit, S.N. *et al.* (1982). Biochemistry 21, 2048.

7. Palmiter, R.D., Davidson, J.M., Gagnon, J. *et al.* (1979)
J. Biol. Chem. 254, 1433.

8. Glanville, R.W., Breitkreutz, D., Meitinger, M. *et al.* (1983)
Biochem. J. 215, 183.

9. Kleinman, H.K., McGoodwin, E.B., Martin, G.R. *et al.* (1978)
J. Biol. Chem. 253, 5642.

10. Chu, M−L., de Wet, W., Bernard, M. *et al.* (1984) Nature 310, 337.

11. Boedtker, H., Fuller, F. and Tate, V. (1983) Int. Rev. Connect. Tiss. Res. 10, 1.

12. Tate, V., Finer, M.H., Boedtker, H. *et al.* (1983) Nucleic Acids Res. 11, 91.

13. Dickson, L.A., de Wet, W., di Liberto, M. *et al.* (1985)
Nucleic Acids Res. 13, 3427.

14. Francis, G., Butler, W.T. and Finch, J.E., Jr. (1978) Biochem. J. 175, 921.

15. Butler, W.T., Finch, J.E., Jr. and Miller, E.J. (1977) Biochemistry 16, 4981.

16. van der Rest, M., Rozenberg, L.C., Olsen, B.R. *et al.* (1986)
Biochem. J. 237, 923

17. Liau, G., Mudryj, M. and de Crombrugghe, B. (1985) J. Biol. Chem. 260, 3773.

18. Sangiorgi, F.O., Benson−Chanda, V., de Wet, W. *et al.* (1985)
Nucleic Acids Res. 13, 2207

19. Sandell, L.J., Prentice, H.L., Kravis, D. *et al.* (1984) J. Biol. Chem. 259, 7826.

20. Ninomiya, Y., Showalter, A.M., van der Rest, M. *et al.* (1984)
Biochemistry 23, 617.

21. Sangiorgi, F.O., Benson−Chanda, V., de Wet, W. *et al.* (1985)
Nucleic Acids Res. 13, 2185.

22. Kohno, K., Martin, G.R. and Yamada, Y. (1984) J. Biol. Chem. 259, 13668.

23. Cheah, K.S.E., Stoker, N.G., Griffin, J.R. *et al.* (1985)
Proc. Natl. Acad. Sci. USA 82, 2555.

24. Nunez, A.M., Francomano, C., Young, M.F. *et al.* (1985) Biochemistry 24, 6343.

25. Allmann, H., Fietzek, P.P., Glanville, R.W. *et al.* (1979)
Hoppe-Seyler's Z. Physiol. Chem. 360, 861.

26. Brandt, A., Glanville, R.W., Hörlein, D. *et al.* (1984) Biochem. J. 219, 625.

27. Seyer, J.M. and Kang, A.H. (1981) Biochemistry 20, 2621.

28. Glanville, R.W. and Fietzek, P.P. (1976) FEBS Lett. 71, 99.

29. Glanville, R.W., Allman, H. and Fietzek, P.P. (1976)
Hoppe-Seyler's Z. Physiol. Chem. 357, 1663.

30. Yamada, Y., Liau, G., Mudryj, M. *et al.* (1984) Nature 310, 333.

31. Yamada, Y., Kühn, K. and de Crombrugghe, B. (1983)
Nucleic Acids Res. 11, 2733.

32. Myers, J.C., Loidl, H.R., Stolle, C.A. *et al.* (1985) J. Biol. Chem. 260, 5533.

33. Myers, J.C., Loidl, H.R., Seyer, J.M. *et al.* (1985) J. Biol. Chem. 260, 11216.

34. Glanville, R.W., Qian, R-Q., Siebold, B. *et al.* (1985)
Eur. J. Biochem. 152, 213.

35. Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545.

36. Schuppan, D., Glanville, R.W. and Timpl, R. (1982) Eur. J. Biochem. 123, 505.

37. Schuppan, D., Glanville, R.W., Timpl, R. *et al.* (1984) Biochem. J. 220, 227.

38. Oberbäumer, I., Laurent, M., Schwarz *et al.* (1985) Eur. J. Biochem. 147, 217.

39. Pihlajaniemi, T., Trygvasson, K., Myers, J.C. *et al.* (1985)
J. Biol. Chem. 260, 7681.

40. Brinker, J.M., Gudas, L.J., Loidl, H.R. *et al.* (1985)
Proc. Natl. Acad. Sci. USA 82, 3649.

41. Kurkinen, M., Bernard, M.P., Barlow, D.P. *et al.* (1985) Nature 317, 177.

42. Ninomiya, Y. and Olsen, B.R. (1984) Proc. Natl. Acad. Sci. USA 81, 3014.

43. van der Rest, M., Mayne, R., Ninomiya, Y. *et al.* (1985)
J. Biol. Chem. 260, 220.

44. Olsen, B.R., Ninomiya, Y., Lozano, G. *et al.* (1985)
Ann. N.Y. Acad. Sci. 460, 141.

45. Lozano, G., Ninomiya, Y., Thompson, H. *et al.* (1985)
Proc. Natl. Acad. Sci. USA. (1985) 82, 4050.

46. Dèak, F., Argraves, W.S., Kiss, I. *et al.* (1985) Ann. N.Y. Acad. Sci. 460, 409.