CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

Canada

THE UNIVERSITY OF ALBERTA

THE PROTEIN CORE OF PROTEODERMATAN SULPHATE

by

NEIL WINTERBOTTOM

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

ORAL BIOLOGY

FACULTY OF DENTISTRY

EDMONTON, ALBERTA

SPRING, 1987

ISBN 0-315-37682-1

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR NEIL WINTERBOTTOM

TITLE OF THESIS THE PROTEIN CORE OF PROTEODERMATAN SULPHATE

DEGREE FOR WHICH THESIS WAS PRESENTED DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED SPRING, 1987

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) .....

PERMANENT ADDRESS:

272 STRETFORD ROAD,
URMSTON, MANCHESTER,
M31 1WJ, ENGLAND

DATED APRIL 1ST 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled THE PROTEIN CORE OF PROTEODERMATAN SULFATE submitted by NEIL WINTERBOTTOM in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY, ORAL BIOLOGY.

Supervisor

External Examiner

Date February 27, 1987

DEDICATION

I dedicate this thesis to my friend Velta, with many thanks for the great support which she provided through the tough times, and yet with many regrets.

ABSTRACT

Proteodermatan sulphate, a small proteoglycan bearing a single dermatan sulphate side-chain, was extracted from bovine skin, in 7M urea, and purified by ion-exchange and gel-filtration chromatography. The intact molecule, and protein core preparations derived from the molecule by deglycosylation using chondroitinase ABC, β-elimination/sulphite addition or anhydrous hydrogen fluoride, were subjected to amino acid sequencing. The sequence data obtained from the amino-terminal region of the various cores indicated the attachment of an O-linked substituent, thought to be a glycosaminoglycan, at residue 4. The subsequent work of Chopra *et al.* (Chopra, R.K., Pearson, C.H., Pringle, G.A., Fackre, D.S. & Scott, P.G. (1985) *Biochem. J.* 232, 277-279) confirmed this.

Cleavage with cyanogen bromide produced 5 discrete peptides of which 3 were sequenced throughout and 2 were sub-fragmented with trypsin or *Staphylococcus aureus* V8 protease in order to obtain their complete sequences. The peptides produced by these cleavage methods were separated by gel filtration and reverse-phase chromatography using volatile buffers. Sodium dodecylsulphate polyacrylamide gel electrophoresis was used to assess the homogeneity of the larger peptides and two-dimensional thin-layer electrophoresis/thin-layer chromatography for the smaller peptides. Overlapping sequence data from the various sets of peptides, along with information obtained from immunoblotting the peptides with monoclonal antibodies raised against proteodermatan sulphate, was used in the ordering of the peptides corresponding to the sequence of the first 214 residues of the molecule. Preliminary information was obtained regarding the site of attachment of one of the oligosaccharide side-chains and two of the three disulphide bridges in the molecule.

Predictions of secondary structure and hydropathy were made based on the sequence and are discussed. In a search for regions of homology with other proteins that had previously been sequenced, the sequence of proteodermatan sulphate was compared to the sequences in a computer data-bank. No significant homologies were recognised but the particularly close homology identified with the recently reported sequence, deduced from complementary DNA, of a proteoglycan from human fibroblast cultures, is discussed.

ACKNOWLEDGEMENTS

Many people provided assistance to me in so many ways during my work on this project; I'd like to thank them all, even if I don't mention them individually. My first thanks should go to Dr. John Gallagher, who introduced me to glycoconjugate research and encouraged me to work towards a Ph.D., and to Dr. Gary Gibson, my predecessor, who apparently suggested the project for this thesis. Next, Dr. Harold Pearson and, more recently, Dr. Paul Scott, my long suffering supervisors, deserve thanks for providing an environment within which I could make my own mistakes and learn from them. I'm also really grateful for the time that Paul put in proof reading the first draft of this thesis.

The members of the *Room 6065 Tea-Break Club* have been of immeasurable assistance, in every respect, during my time in Edmonton: Dan Fackre's attention to detail in amino acid analysis, sequencing and any other "high-tech" aspects of the work probably forms the foundation of this project. Doctors Ian Carney and Don Gates provided invaluable insight into the Ph.D. degree in general. Elaine Moase, super-tech, frequently amazed me with her technical expertise and the speed with which she took ideas from our discussions, and produced results. Dr. Mike Eggert showed a keen interest in the project throughout, and they were all willing to listen to my moaning and to openly, and occasionally violently, discuss my project, science in general, and most other subjects under the sun.

I'd like to thank Dr. Larry Smillie for allowing me to make use of the Biochemistry Department's sequencer facility, for high-sensitivity analyses, and Mike Carpenter for his expertise in operating that instrument and his willingness to share his wealth of experience with me. I am grateful to Dr. Bob Hodges for allowing me to use his apparatus for deglycosylation with anhydrous HF, and to Dr. Joyce Pearlstone for her assistance with the use of the Bionet system.

I am indebted to Dr. Gordon Pringle for his generosity in providing me with the monoclonal antibodies to PDS, which he prepared and characterised during his Ph.D. project. I would like to thank Mrs. Sophie Lehocky for the many PDS preparations which she has undertaken, for teaching me how to run SDS-PAGE and later for her expertise in running the

same technique and the two-dimensional TLE/TLC on my behalf. I'm very grateful to Elena Edwards for operating the amino acid analyser in the critical; latter days of my project and I'd like to thank Marie Jones for her work on the neutral sugar analyses in Dr. Mike Eggert's lab.

To my friends out there in the real world, a big thank you for your friendship and support, especially over the last few months. To single out just a few without whom I'd have gone nuts: Lona & Greg Ani, Mike & Sheila Clegg, particularly Mike for proof-reading my thesis, Peter & Marge Kirchmier and Wade & Avril Pike. And I couldn't consider this section complete without a "ta" to my many pals in Edmonton's Scouting community, especially at 130th Edmonton (Duggan).

Finally, I would like to thank the Alberta Heritage Foundation for Medical Reearch for their funding of my studentship and the Canadian Medical Research Council for their support of my research project through grants to my supervisors.

Table of Contents

List of Tables

List of Figures

Figure Page

List of Plates

LIST OF ABBREVIATIONS

Abbreviations used in this manuscript are:

A or Ala	Alanine
B or Asx	Aspartic acid or asparagine
C or Cys	Cysteine
Cya	Cysteic acid
Cmc	Carboxymethylcysteine
D or Asp	Aspartic acid
E or Glu	Glutamic acid
F or Phe	Phenylalanine
G or Gly	Glycine
H or His	Histidine
Hse	Homoserine
I or Ile	Isoleucine
K or Lys	Lysine
L or Leu	Leucine
M or Met	Methionine
mD	Methylaspartic acid
mE	Methylglutamic acid
N or Asn	Asparagine
P or Pro	Proline
Q or Gln	Glutamine
R or Arg	Arginine
S or Ser	Serine
T or Thr	Threonine

V or Val	Valine
W or Trp	Tryptophan
X or Xxx	Any unidentified amino acid.
Y or Tyr	Tyrosine
Z or Glx	Glutamic acid or Glutamine
GlcN	Glucosamine
GalN	Galactosamine
PDS	Proteodermatan sulphate
PDS(ABC)core	The protein core derived from PDS by deglycosylation with chondroitinase ABC
PDS(BES)core	The protein core derived from PDS by β-elimination/sulphite addition.
PDS(HF)core	The protein core derived from PDS by deglycosylation in anhydrous hydrogen fluoride.
GAG	Glycosaminoglycan
DS	Dermatan sulphate
HABR	Hyaluronic acid binding region
MAbs	Monoclonal antibodies
HPLC	High pressure liquid chromatography
TLC	Thin layer chromatography
PITC	Phenyl isothiocyanate
ATZ	Anilinothiazolinone
PTH	Phenylthiohydantoin
SDS	Sodium dodecylsulphate
SDS-PAGE	Sodium dodecylsulphate polyacrylamide gel electrophoresis
TBS	Tris-buffered saline

TFA	Trifluoroacetic acid
CNBr	Cyanogen bromide
Tris	Tris(hydroxymethyl)aminomethane
Bis	N,N'-Methylene-bis-acrylamide
TEMED	N,N,N',N'-Tetramethylethylenediamine
Quadrol	N,N,N',N'-Tetrakis (2-hydroxypropyl) ethylenediamine
M_r	Relative molecular weight
M_R	Relative mobility
Y_R	Repetitive yield
cDNA	Complementary deoxyribonucleic acid

1. INTRODUCTION

1.1 Connective Tissue

If by some magic solution one could dissolve all the connective tissue of the body, all that would remain would be a mass of slimy epithelium, quivering muscle and frustrated nerve cells. (Arcadi, 1952).

The connective tissues, which comprise cellular and fibrous components embedded in an amorphous ground substance, provide the supporting and connecting framework for all the other tissues of the body (epithelia, nerves and muscle). The connective tissues include ligaments, tendons, fascia, bone, dentine, cartilage, joint capsules, the sub-epithelial portions of the skin, the vitreous body, intervertebral discs, umbilical cord and elements of the heart valves, aorta and smaller blood vessels. The indigenous cells of the connective tissues include fibroblasts, osteoblasts, chondroblasts, macrophages and mast cells. These cells produce the bulk of the extracellular material of connective tissues but epithelial cells may make some contribution.

Connective tissues are characterised by a high proportion of extracellular material with respect to the number of cells. There is a great variety in the relative amounts and different types of matrix macromolecules and in the organisation of those molecules within the extracellular matrix. These variations are related to the function of the tissue. Thus, the matrix can provide the hard substance of bones, the resilience of cartilage, the tensile strength of tendons or the transparency of the cornea. (Alberts *et al.*, 1983; Goldberg & Rabinovitch, 1983).

1.2 Fibrous Components of the Connective Tissue Matrix

The collagen and elastic fibres which confer, respectively, the tensile strength and recoil ability inherent in various connective tissues, consist of large, insoluble, cross-linked aggregates.

1.2.1 Collagen

To date, eleven types of collagen have been recognised and characterised to varying extents (see Table 1). To the limit of current knowledge, all of these collagen types conform to the definition that they contain sizeable domains exhibiting typical collagen folding, and that they form extracellular supramolecular aggregates whose principal function is support (Miller, 1985).

Collagen types I, II and III are collectively known as the interstitial collagens. They function as large polymeric fibrils of 10-300nm diameter or as larger fibres of up to several μm diameter. Though composed of different gene products, the biosynthesis and fibrillogenesis of these molecules is essentially the same. (Miller, 1985). More information is available concerning type I collagen than any of the other collagen types because 90% of all collagen is type I. This molecule is therefore considered to be a benchmark to which the other collagens are compared.

1.2.1.1 Biosynthesis of type I collagen

Type I collagen is a heterotrimeric molecule comprising two identical α1(I) and one α2(I) chains. The three protein chains are synthesised as pro-α chains which comprise the mature protein with a carboxy-terminal propeptide extension and amino-terminal signal peptide and propeptide extensions. (Monson *et al.*, 1975). The amino acid sequence of the mature protein is of the following form:

$$(Gly\text{-}X\text{-}Y)_n$$

(Hofmann, Fietzek & Kühn, 1978). Proline and hydroxyproline form 23% of the molecule (hydroxyproline is almost entirely restricted to the "Y" position, Highberger *et al.*, 1982). These imino acids impose rotational restrictions on the polypeptides and thus direct the left-handed helical conformation of the α-chains (Cowan & McGavin, 1955). The hydroxylation of some of the proline and lysine residues and the subsequent glycosylation (Hydroxylysine-Galactose-[Glucose]) of some of the hydroxylysine residues occur

TABLE 1
THE COLLAGENS

Type	Chain M_r	Molecular species	Form of aggregates	Major source
I	95K	[α1(I)]$_2$α2(I) [α1(I)]$_3$	Fibres	Virtually all connective tissues
II	95K	[α1(II)]$_3$	Fibres-Fibrils	Hyaline cartilage
III	95-110K	[α1(III)]$_3$	Reticular Networks	Distensible connective tissues
IV	170-185K	[α1(IV)]$_2$α2(IV) [α1(IV)]$_2$ [α2(IV)]$_2$ RC(IV)	Open-mesh-like aggregates of basement membranes	Basement membranes
V	130-200K	[α1(V)]$_2$α2(V) [α1(V)]$_3$ [α1(V), α2(V), α3(V)]	Aggregates of unknown structure in pericellular and perifibrillar zones	Virtually all connective tissues
VI	140-240K	Chains: α1(VI), α2(VI), α3(VI)	Microfibrillar elements	Placental villi
VII	>170K	[α1(VII)]$_3$	?	Placental membranes
VIII	100-180K	?	?	Endothelial cells
IX	85K	?	Aggregates of unknown structure in chondrocyte lacunar regions	Hyaline cartilage
X	59K	?	Aggregates of unknown structure in zones of hypertrophying chondrocytes	Hyaline cartilage
XI	?	Chains: 1α, 2α, 3α	Aggregates of unknown structure in chondrocyte lacunar zones	Hyaline cartilage

adapted from Miller (1985)

post-translationally (Gallop & Paz, 1975). Disulphide bonding of the C-terminal propeptides of the three α-chains (Byers *et al.*, 1975) brings them into register and promotes the formation of the right-handed triple helix, which is stabilised only by non-covalent interactions (Kivirikko & Myllyla, 1984). The glycine at every third residue of the sequence is essential, as only this residue is small enough to occupy the position on the inside of the triple helix (Bornstein & Traub, 1979).

On secretion from the cell the propeptides are cleaved by procollagen peptidases producing the mature triple-helical collagen molecule which is 300nm long and 1.5nm in diameter and which comprises three α-chains of 95,000 dalton (approximately 1050 amino acid residues) each. The cleavage of the propeptides leaves short non-triple-helical regions, known as the telopeptides, at both the N- (16 residues in bovine α1(I)) and C- (26 residues in bovine α1(I)) termini of the molecule (Davidson *et al.*, 1977; Kivirikko & Myllyla, 1984; Miller, 1984). In addition to their function in bringing the three α-chains into register, the propeptides prevent the intracellular formation of collagen fibrils (Bornstein & Traub, 1979).

1.2.1.2 Fibrillogenesis of type I collagen

On cleavage of the propeptides the molecules aggregate to form fibrils and subsequently, fibres. When stained and observed under the electron microscope the fibrils exhibit a pattern of striations at 67nm (referred to as 'D') intervals. Adjacent molecules within the fibril have been shown to be displaced by that distance axially. Individual collagen molecules are 4.4D in length and are arranged in the fibril with a gap of 0.6D between the N- and C-termini of successive molecules (Petruska & Hodge, 1964). This arrangement maximises the non-covalent interactions (electrostatic, hydrophobic) between amino acid side chains on adjacent molecules (Hulmes *et al.*, 1973, Hofman, Fiezek & Kühn, 1978).

The formation of these fibrils appears to be inherent in the structure of the molecule (Cassel, 1966), as fibrils will form *in vitro* . Many studies have considered the interaction of small dermatan sulphate proteoglycans with the surface of collagen fibrils (Scott, 1980; Scott & Haigh, 1985; Scott *et al.*, 1986; Scott & Hughes, 1986) and the relative amounts of this proteoglycan in relation to the diameter of collagen fibrils (Gillard *et al.*, 1977; Scott & Hughes, 1986) The proteoglycan has been shown to be present in greater amounts, per unit weight of collagen, when the collagen is in the form of thin fibrils (Scott *et al.*, 1981) and has been shown to associate specifically with the d band of the collagen fibril, often forming perifibrillar hoops of filamentous proteoglycan (Scott & Orford,

1981). Though the mechanism is not understood, it has been suggested that this proteoglycan may be involved in the regulation of collagen fibrillogenesis. Some control of the arrangement of collagen fibres has also been suggested, possibly involving the orientation of the cytoskeleton of the secretory cell, to explain the organisation of such different forms as the interwoven fibres in skin, the parallel arrangement in tendon and the almost crystalline arrangement in cornea (Goldberg & Rabinovitch, 1983). A better understanding of the control of these processes awaits a better understanding of the interactions of collagen with cells and other macromolecules.

Covalent cross-links form within and between molecules in the fibrils. The ϵ-amino groups of certain lysine and hydroxylysine residues, mostly within the telopeptides (Miller, 1984), are converted to aldehydes by the extracellular, copper-dependent enzyme, lysyl oxidase. When such aldehydes and the ϵ-amino groups of other lysines are brought together in the fibril, the cross-links form non-enzymically. The extent of cross-linking varies dependent on the function of the tissue, with more cross-links formed where greater tensile strength is required (Piez, 1968; Gallop & Paz, 1975).

1.2.1.3 Turnover of the interstitial collagens

At physiological pH and temperature the triple-helix is resistant to most tissue proteases. During growth, remodelling, inflammation and repair there is considerable degradation of collagen. The animal collagenases are enzymes with specific affinity for native interstitial collagens, which catalyse a single cleavage at a site 75nm from the C-terminus of all three chains of the molecule. The fragments thus produced adopt a random coil conformation and are then susceptible to degradation by neutral proteases (Gross & Nagai, 1965; Kang *et al.*, 1966).

1.2.1.4 The other collagens

Types I and III collagen have a relatively low content of hydroxylysine and carbohydrate whereas type II has a relatively high content of both. Type III has a higher

hydroxyproline content. Types IV and V also have high hydroxylysine and carbohydrate contents (Bornstein & Traub, 1979).

Of the minor collagens, type IV is probably the best characterised. It has been suggested that type IV, the major collagenous component of basement membranes, is deposited not as fibrils but in the form of an open mesh-like network. Unlike the interstitial collagens, type IV undergoes very little if any extracellular processing. Rotary shadowing has shown molecules of approximately 400nm in length with a compact helical end and a globular end. Multimers studied in the same system showed tetramers joined at the helical ends. This would be an appropriate form from which to assemble the mesh-like network which is apparently the functional form of this molecule (Miller, 1985).

With the exception of the information in Table 1, very little has been published concerning the other lesser collagens. Of interest recently was a publication reporting that type IX collagen has been shown to have covalently bound chondroitin sulphate chains and is thus also a proteoglycan (Bruckner *et al.*, 1986).

1.2.2 Elastic Fibres

Elastic fibres are present in those tissues which require the ability to recoil after transient stretching. The fibres occur in varying amounts (up to 50% of the dry weight of aorta) and are arranged differently in different tissues. Two thick concentric lammellae of fibres are found in arteries in contrast to a highly fenestrated network in mesentery, fasciae and skin (Goldberg & Rabinovitch, 1983).

The fibres are made up of an amorphous substance surrounded by microfibrils. The amorphous substance is the protein elastin, the main component of the fibrils. Elastin comprises approximately one third glycine, 11% proline and more than 40% hydrophobic residues. There is very little hydroxyproline and no hydroxylysine (Sage & Gray, 1979). The individual molecules function in a largely aperiodic structure with desmosine and isodesmosine crosslinks between chains connecting the elastin into filamentous and sheet-like networks. It is this structure which confers the rubber-like elasticity to the fibres (Aaron & Gosline, 1981).

Collagen fibres in the matrix limit the extent of stretching in order to prevent tearing.

Tropoelastin, the 72,000 dalton precursor of elastin (Smith, Weissman & Carnes, 1968) is secreted from fibroblasts by exocytosis into the matrix where the formation of cross-links is initiated by the copper-requiring enzyme, lysyl oxidase (Siegel, 1979), which is also involved in the formation of the cross-links of collagen. The enzyme catalyses the oxidative deamidation of the ε-amino groups of specific lysyl residues producing allysine. Four allysine residues condense to form the carbon and nitrogen ring structure of the desmosines, which could therefore cross-link four chains; usually only two chains are thus linked (Rosenbloom, 1984).

The elastin core of the elastic fibre is surrounded by microfibrils of 10-12nm diameter some of which also occur within the fibre. The microfibrils are composed of a glycoprotein comprising mainly hydrophilic amino acids, a considerable number of half-cystines and approximately 5% neutral sugar. There is no hydroxyproline, desmosine or isodesmosine (Ross & Bornstein, 1969). They may be observed prior to the secretion of the elastin and, it has been suggested, may serve to organise the elastin as it is secreted (Fahrenbach, Sandberg & Cleary, 1966).

1.3 Ground Substance of the connective tissue matrix

The cellular and fibrous elements of the extracellular matrix are surrounded by an apparently amorphous hydrated gel called the ground substance. The gel is maintained by the presence of hyaluronic acid and glycoconjugates (both proteoglycans and glycoproteins). The ground substance permits, and to some extent regulates, the diffusion of molecules passed by the vascular filters (nutrients, salts, metabolites, hormones) (Goldberg & Rabinovitch, 1983) Various components of the ground substance are also involved in associations with other matrix and cellular components (see below).

1.3.1 Glycoproteins

The glycoproteins of the connective tissue matrix form a large heterogeneous class. They all exhibit relatively low solubility, or insolubility, in aqueous buffers, a high content of

acidic and aliphatic amino acids and a variable but relatively high content of cysteine (Robert & Moczar, 1982). Fibronectin and laminin are the best characterised representatives of this group. The amino acid composition of the link glycoprotein involved in the stabilisation of the large proteoglycan aggregates found in cartilage, is similar to that for these glycoproteins (Bonnet, Périn & Jollès, 1978). The link glycoprotein could be considered within this category but will be discussed below with the proteoglycan aggregate (Hascall, 1977) of which it is a component.

These glycoproteins, known as the structural glycoproteins, contain amounts of carbohydrate of up to 15%. They associate with the other macromolecules of the matrix and with cell membranes and are involved in the ordered arrangement of these various components. Fibronectin and laminin are the best characterised of the molecules in this group (see below). Recently identified and as yet not well characterised are chondronectin, which is involved in chondrocyte-collagen adhesion (Hewit *et al.*, 1980), hyalurononectin, which binds hyaluronic acid (Delphench & Halavent, 1981), and the sulphated basement membrane glycoprotein, entactin (Carlin *et al.*, 1981).

1.3.1.1 Fibronectin

Fibronectin is a glycoprotein comprising two similar subunit polypeptide chains of about 220,000 daltons each, which are linked near their carboxy-terminal ends by a disulphide bridge (Kurkinen, Vartio & Vaheri, 1980). In the connective tissue matrix an insoluble form of fibronectin mediates cell-matrix adhesion. A soluble form, formerly called cold-insoluble globulin, occurs in the plasma, where it plays an important role in the formation of the haemostatic plug in the repair of damaged blood vessels (Hynes & Yamada, 1982).

The amino acid sequences of rat and human fibronectins have been deduced from cDNA sequences and they suggest that the protein is the product of a gene which underwent substantial gene duplication during its evolution (Hirano *et al.*, 1983). Domains exist within each subunit of the molecule, which have sites that will bind to fibrin, collagen, cells and heparin (Ruoslahti, Engvall & Hayman, 1981). The molecule can

therefore cause the attachment of any one of these to any other. The binding site for the cell surface has been identified as a sequence of four amino acids (arginine-glycine-aspartic acid-serine) within the fibronectin (Pierschbacher & Ruoslahti, 1984). Cell surface bound fibronectin interacts with the actin filaments inside the cell and the fibres of actin and fibronectin align with each other (Hynes & Yamada, 1982). This phenomenon is thought to play a role in the organisation of collagen fibres in the matrix through the ability of fibronectin to also bind to that molecule. A group of three glycoproteins called the 140K complex is thought to be involved in this trans-membrane interaction of fibronectin and actin (Chen *et al.*, 1985).

1.3.1.2 Laminin

Laminin is a high molecular weight glycoprotein which is an abundant component of almost every basement membrane in the body. On reduction two components are seen: one of 200-220kDa and one of 400-440kDa (Timpl *et al.*, 1979). Preliminary evidence suggests that they are related but not identical in amino acid sequence. Electron microscopy shows a cruciform structure with one 75nm-long arm and three 35nm-long arms (Engel *et al.*, 1981). The molecule binds heparin and is probably involved as an adhesive protein, bound to heparan sulphate proteoglycans in the basement membrane (Sakashita, Engvall & Ruoslahti, 1980).

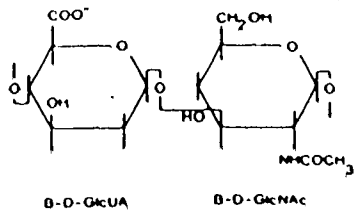
1.3.2 Glycosaminoglycans

Glycosaminoglycans are linear polymers composed of repeated disaccharide units. The disaccharide comprises a hexosamine, which is usually N-acetylated, and a uronic acid, except in keratan sulphate which has galactose in place of uronic acid (see Fig. 1). With the exception of hyaluronic acid the GAGs are sulphated and require a protein acceptor for their biosynthesis (Chakrabarti & Park, 1980). A protein with one or more glycosaminoglycan side chains, attached covalently, is referred to as a proteoglycan. The linkage of the GAG chain to the protein core, is via a short, specialised linkage region which is different from the repeating disaccharide region.

Fig. 1. Structures of the repeating disaccharide units of the glycosaminoglycans. The heteropolymeric dermatan sulphate and heparan sulphate comprise varying amounts of the two disaccharide units shown *n* denotes the number of disaccharide units in the chains.
GlcNAc = N-acetyl glucosamine, *GalNAc* = N-acetyl galactosamine; *Gal* = galactose; *GlcUA* = glucuronic acid; *IdUA* = iduronic acid.
Adapted from Heinegård & Paulsson (1984).

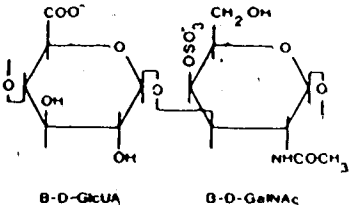
HYALURONIC ACID

(n 50-10000)



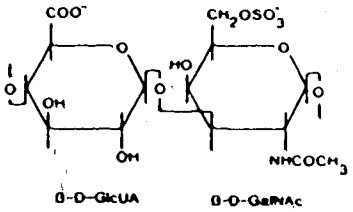
CHONDROITIN 4-SULFATE

(n 20-60)



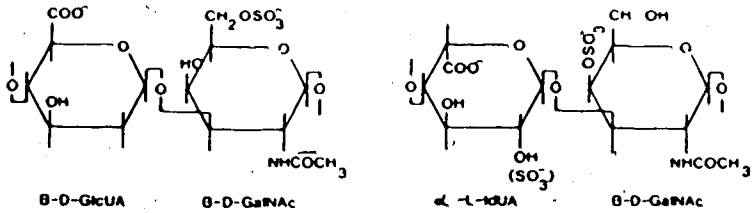
CHONDROITIN 6-SULFATE

(n 20-60)



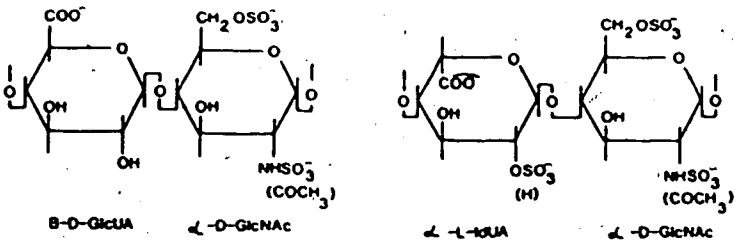
DERMATAN SULFATE

(n 30-80)



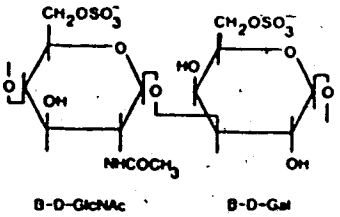
HEPARAN SULFATE

(Heparin)



KERATAN SULFATE

(n 5-40)



1.3.2.1 Hyaluronic Acid

Hyaluronic acid is a polymer consisting of the disaccharide D-glucuronic acid-N-acetylglucosamine repeated many times. It is the only GAG which is neither sulphated nor covalently bound to protein to form a proteoglycan. It is also the GAG with the largest polymer size (often several thousand sugar residues). The molecule is restricted by electrostatic constraints and steric factors, due to its high level of hydration. This prevents the molecule from folding tightly and causes it to adopt an extended random coil conformation which occupies a huge volume relative to its mass. As with the other GAG chains, hyaluronic acid is highly hydrophilic and attracts a large volume of water forming a hydrated gel. This imparts a turgor on the tissue which resists compressive forces (Chakrabarti & Park, 1980). It is believed that such gels have a role in the facilitation of cell migration through tissues, during development or wound repair, in that they form channels through which movement is easier (Alberts *et al.*, 1983).

Hyaluronic acid is also involved as a core on which proteoglycans assemble to form very large aggregates (see proteoglycan section).

1.3.2.2 Chondroitin sulphate

Chondroitin 4-sulphate and chondroitin 6-sulphate differ only in the position at which the ester sulphate group is attached to the galactosamine. The definitions are not absolute as various co-polymers exhibiting sulphation at both sites and over-sulphated forms with sulphated uronic acid groups have also been reported (Suzuki, 1960; Liau *et al.*, 1978). Six different glycosyltransferases and two sulphotransferases (specific for the 4 and 6 positions) are involved in the biosynthesis of chondroitin sulphate. Xylosyl transferase catalyses the transfer of xylose from UDP-xylose to the acceptor serine residue of the protein core. There appears to be a requirement for a glycine residue at the position carboxy-terminal to the linkage serine (Johnson & Baker, 1973; Isemura *et al.*, 1981). Galactosyl transferases I and II catalyse the addition of two galactose residues to the xylose. Transferase I adds to the xylose, II adds to the first galactose. Glucuronosyltransferase I then adds glucuronic acid to the second galactose. The repeating

disaccharides are then added by the alternate actions of N-acetylgalactosaminyltransferase and glucuronosyltransferase II which add N-acetylgalactosamine and glucuronic acid respectively. Sulphation follows at least one step behind chain elongation (Rodén & Horowitz, 1978)

1.3.2.3 Dermatan sulphate

Dermatan sulphate is synthesised by the same enzymes responsible for the assembly of chondroitin sulphate but with the additional action of an unique epimerase which converts some of the D-glucuronic acid to L-iduronic acid by inversion at C-5. Sulphation then follows preventing re-epimerisation, which would cause reversion to D-glucuronic acid (Malmström, 1984).

1.3.2.4 Heparin and heparan sulphate

Heparin and heparan sulphate are synthesised on different protein acceptors. Heparin is generally more highly charged than heparan sulphate. They can both contain N-sulphate in place of N-acetyl groups on the glucosamine. This appears to be a prerequisite for the epimerisation of glucuronic acid to iduronic acid (in these glycosaminoglycans, though not in dermatan sulphate) and as heparin usually contains more N-sulphated glucosamine, it also contains more iduronic acid (50-90% of total hexuronate) (Heinegård & Paulsson, 1984). These two glycosaminoglycans serve very different functions: heparin has an anticoagulant role whereas heparan sulphate proteoglycans are involved in glomerular filtration and the attachment and spreading of cells on substrata.

1.3.2.5 Keratan sulphate

Like the complex oligosaccharides of glycoproteins, keratan sulphate has oligosaccharide branches, contains mannose, and has sialic acid residues in non-reducing terminal positions. Two different linkages to protein have been identified. Type I in cornea is linked to asparagine through an N-glycosidic linkage from N-acetylglucosamine. Type II, found in cartilage and other skeletal tissues, is linked to serine or threonine

through an O-glycosidic linkage from N-acetylgalactosamine. Both polysaccharides have the same repeating disaccharide unit, D-galactosyl-N-acetyl-D-glucosamine (Rodén & Horowitz, 1978).

1.3.3 Proteoglycans

Proteoglycans are glycoconjugates comprising one or more glycosaminoglycan side chains attached to a protein or more often, a glycoprotein core (see Fig. 2). These molecules were originally classified depending upon tissue of origin. The GAG components of these molecules were characterised first, and the proteoglycans have been thus subsequently categorised according to the type of GAG chain attached to the molecule (Poole, 1986). Sufficient data are now available to allow the modification of this GAG-based classification system to incorporate a consideration of amino acid sequence homology, (Hassell, Kimura & Hascall, 1986) and immunological cross-reactivity (Heinegård *et al.*, 1985).

1.3.3.1 The proteoglycans of cartilage

The large, aggregating proteoglycan of cartilage has been the subject of the most investigation. This molecule, which occurs in two major forms of approximate M_r 1,300 and 3,500kDa, has a protein core of 200-300kDa with a globular domain of 60-70kDa at the amino-terminus, which encompasses a hyaluronic acid-binding region (HABR), a site that interacts with the link proteins, most of the 10-15 N-linked oligosaccharides on the molecule, and few if any GAG chains(Hassell, Kimura & Hascall, 1986). The region of the core carboxy-terminal to the globular domain of the smaller, keratan sulphate-rich proteoglycan, is the 30kDa keratan sulphate-binding region where 60% or more of the 130 keratan sulphate chains are attached. The remainder of the molecule, the chondroitin sulphate-attachment region, contains the majority of the approximately 100 chondroitin sulphate chains, the rest of the keratan sulphate chains and several O-linked oligosaccharides (Heinegård & Axelsson, 1977; Heinegård & Paulsson, 1984; Lohmander *et al.*, 1980).

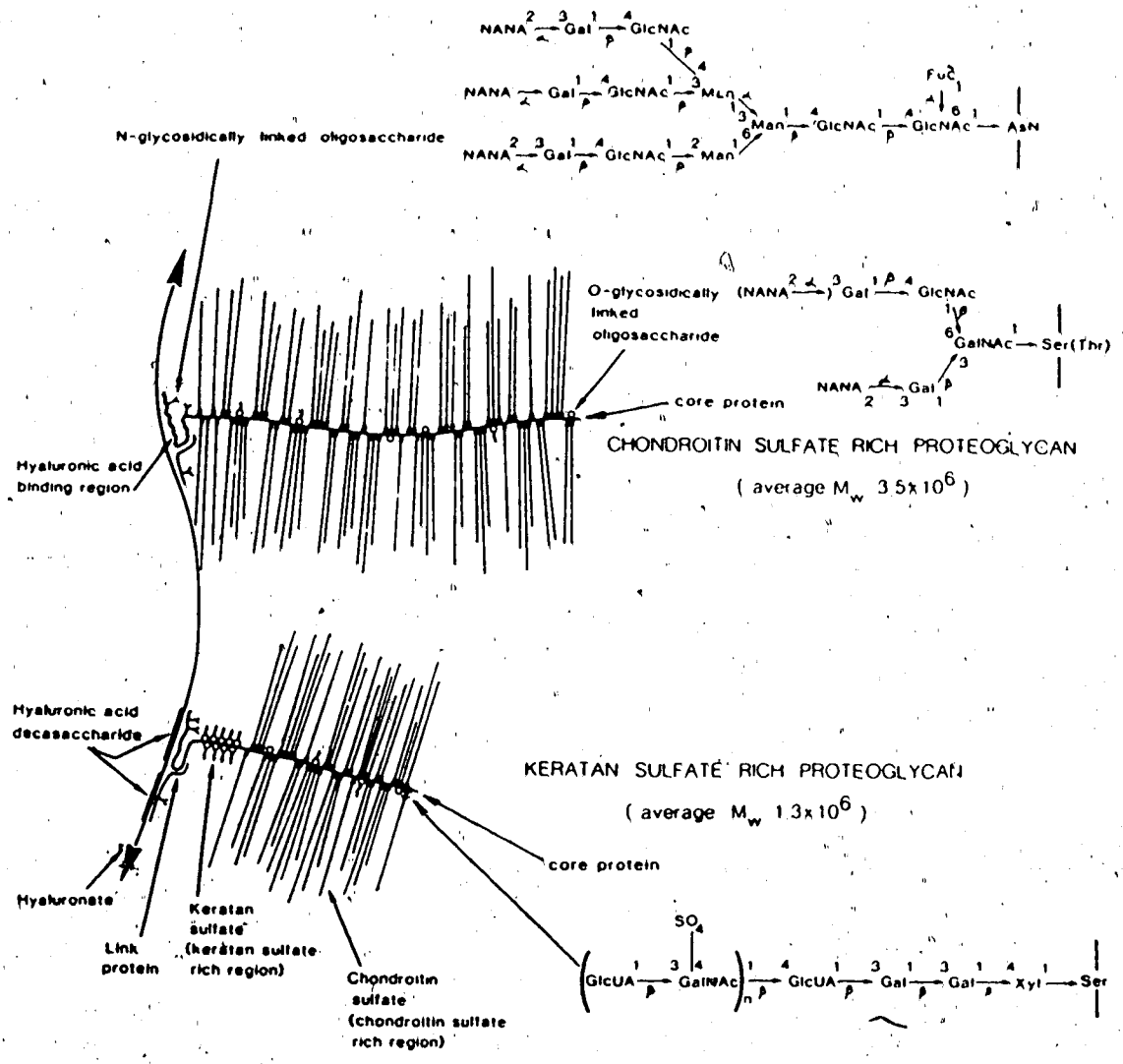


Fig. 2. Tentative structures of proteoglycans. Models are shown for representatives of the two major sub-populations of proteoglycan from cartilage. Adapted from Heinegård & Paulsson (1984)

This molecule functions in cartilage as extremely large aggregates in which some 50 monomers attach via the HABR to hyaluronic acid. This non-covalent interaction is stabilised by a link glycoprotein which binds to the protein core of the proteoglycan monomer and to the hyaluronic acid. Three link proteins of between 40-50kDa have been

identified, though only one is involved in the attachment of each monomer. The interaction of the HABR and the link protein with hyaluronic acid each involves a decasaccharide of the hyaluronic acid. Amino acid sequencing of the link protein and the HABR have shown regions of considerable homology which, it is suggested, may be involved in the binding of both link protein and the proteoglycan to hyaluronic acid (Neame *et al.*, 1985). This proteoglycan aggregate, which comprises 85-90% of the proteoglycan of cartilage, appears to contain more of a larger, chondroitin sulphate-rich monomer in young tissue, whereas older cartilage contains more of a somewhat smaller keratan sulphate-rich monomer (Heinegård *et al.*, 1985a; see Fig. 2). The very high charge density on these proteoglycan aggregates permits them to retain a large amount of water (30-50 times their dry weight; Hascall & Sajdera, 1970). In cartilage this swelling is limited by the network of collagen fibrils. The swelling pressure confers the resilience to cartilage which permits the tissue to withstand repeated loading.

Some 10% of the proteoglycans are large but incapable of forming aggregates (Heinegård & Hascall, 1979), and 2-3% are quite small (Heinegård *et al.*, 1981). The small proteoglycan from bovine nasal cartilage has a core protein, produced by digestion with chondroitinase ABC, of M_r 45,000 with one or two chondroitin sulphate chains attached. Small proteoglycans from adult bovine articular cartilage, which have dermatan sulphate side chains, are divisible into two immunologically distinct sub-populations. They have been separated by the tendency of one of the molecules (PG-I) to self-aggregate. The other proteoglycan, PG-II, electrophoreses with an apparent M_r of 90,000-120,000 and has an (ABC)core which produces a doublet of M_r 43,000 and 47,000 in SDS-PAGE (Rosenberg *et al.*, 1985).

1.3.3.2 Historical review of the occurrence and structure of proteodermatan sulphate

Following a preliminary report of the extraction of a dermatan sulphate-protein complex from bovine heart valves with cold 6M urea (Toole & Lowther, 1965), Toole and Lowther reported the more efficient isolation of the same complex following extraction of heart valves with 6M urea at 60°C (Toole & Lowther, 1968). They observed that this

complex precipitated collagen fibrils from solution, and proposed that its primary biological role may be in the formation and orientation of collagen fibres. Antonopoulos *et al.* (1974) introduced extraction with 4M guanidinium chloride, at 4°C, followed by ion exchange chromatography on DEAE-cellulose. With the inclusion of a cocktail of protease inhibitors in the extraction and column buffers, this has been adopted by many laboratories as a standard method for the extraction of PDS from various tissues (Bovine sclera: Coster & Fransson, 1981; Porcine skin: Damle *et al.*, 1982; Bovine skin: Pearson & Gibson, 1982; Human uterine cervix: Uldbjerg *et al.*, 1983). Fujii & Nagai (1981) isolated PDS from calf skin using 3M magnesium chloride containing protease inhibitors. Nakamura *et al.* (1983) extracted PDS from newborn calf skin using only 0.45M sodium chloride. This non-denaturing system extracted only 30% of the PDS. Anderson (1975) used affinity chromatography on concanavalin A-sepharose, to isolate PDS from bovine tendon. This demonstrated the presence of oligosaccharides containing glucose and/or mannose on the molecule.

Several differences have been observed between the PDS extracted from various tissues. Reported molecular weights range from 70-120kDa. These molecules werereported to comprise between one and four dermatan sulphate chains of 50-85% iduronic acid and 15-29kDa molecular wieght, and a glycoprotein core of 44-56kDa which has several O- and N-linked oligosaccharide side chains attached. Coster *et al.* (1981) reported a self association phenomenon for PDS. They proposed and later confirmed (Fransson *et al.*, 1982) that this aggregation involved interaction between dermatan sulphate chains.

1.3.3.3 Heparin and heparan sulphate proteoglycans

The rat mast cell proteoglycan has a small core protein (20kDa) which, it has been suggested, consists of alternating serine and glycine residues. Heparin chains of approximately 80kDa are attached to about two of every three serine residues (Robinson *et al.*, 1978).

Heparan sulphate proteoglycans form a very diverse group of molecules which have been identified at the cell surface and in basement membranes. Of those at the cell surface some are intercalated in the lipid bilayer, presumably by means of an hydrophobic region of the core, whilst others are thought to be attached by interaction with other cell surface macromolecules. These cell surface-proteoglycans vary in size from the M_r 75,000 molecule of liver cells, which has three or four heparan sulphate chains of M_r 14,000 each attached to a protein core of M_r 17-30,000 (Oldberg *et al.*, 1979), to the M_r 350,000 molecule of fibroblasts which has 2 protein cores of M_r 75,000 each, linked by a disulphide bridge, and which has 8-12 heparan sulphate chains of M_r 20,000 (Carlstedt *et al.*, 1983; Cöster *et al.*, 1983). They are believed to have a role in cell-cell and cell-basement membrane interactions (Gallagher, Lyon & Steward, 1986). The basement membrane heparan sulphate proteoglycan appears to be distinct from the cell surface molecules. The molecule isolated from glomerular basement membranes has four or five GAG chains attached to a protein core of M_r approximately 30,000 (Stow *et al.*, 1983).

1.3.3.4 Other proteoglycans

Small proteoglycans similar to the dermatan sulphate-bearing proteoglycan of cartilage referred to as PG-I, have been identified in several other tissues (tendon, bone & skin. Hassell, Kimura & Hascall, 1986). A small keratan sulphate proteoglycan thought to be unique to cornea has been identified. The protein core preparation that was isolated following digestion of the proteoglycan with keratanase appears as two bands of 53,000-55,000 (major band) and about 40,000 (minor band) in SDS-polyacrylamide gel electrophoresis. There are two to three branched oligosaccharide structures each of which carries two keratan sulphate chains. High mannose N-linked oligosaccharides are also present. This keratan sulphate proteoglycan is also thought to be involved in the regulation of collagen fibrillogenesis especially with regard to the unique arrangement in the cornea which confers transparency on that tissue (Axelsson, 1984).

1.4 Amino acid sequencing of proteoglycan core proteins

Based on the same chemistry as the manual sequencing method (Edman, 1950, see Fig. 3) Edman and Begg (1967) published a description of an automated instrument, the protein sequenator, capable of carrying out degradations of up to 60 cycles in a single run. This instrument, subsequently made available commercially, was used for most of the sequencing undertaken in this project.

At the start of this project there were no reliable sequence data available for any proteoglycan. As it became apparent that the classification of the proteoglycans would necessitate an understanding of their protein cores, and that the function of the proteoglycans may involve properties of the cores, projects were initiated to determine the amino acid sequences of some of these molecules. The availability of recombinant DNA technology made the determination of deduced amino acid sequence far faster, and made it possible to attempt the sequencing of the larger proteoglycans and those proteoglycans not readily available in quantities large enough for amino acid sequencing.

The only sequence data to come from early work were derived from the short peptides which remained attached to GAG chains following pronase digestion of cartilage (Johnson & Baker, 1973; Isemura *et al.*, 1981). From the very simple amino acid composition of the protein core of the heparin proteoglycan (essentially glycine and serine in a ratio of approximately 1:1), it was deduced that the sequence was probably $(Ser\text{-}Gly)_n$ (Robinson *et al.*, 1978). Walton, Volger & Jaynes (1979) reported the amino-terminal amino acid sequence of the protein core of the large aggregating proteoglycan from cartilage. Indicative of the difficulties involved in the sequencing of such a large, heavily glycosylated molecule is the fact that they also predicted that this was the chondroitin sulphate binding region end whereas subsequent workers agree that the amino-terminus is at the hyaluronic acid-binding region end (Doege *et al.*, 1986).

More recent studies of the cartilage proteoglycan monomer have produced short amino acid sequences from known regions of the molecule (Bonnet *et al.*, 1983; Périn *et al.*, 1984). This has aided those investigating the sequence of cDNA coding for the same molecule. Two

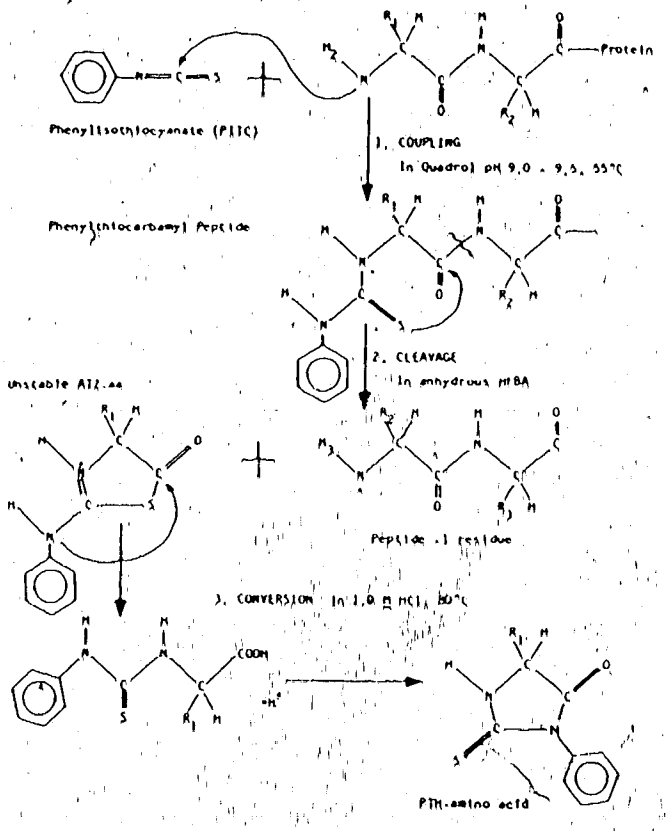


Fig. 3. The Edman degradation.

groups have recently reported cDNA sequences coding for the carboxy-terminal 269 (Doege *et al.*, 1986) and 379 (Sai *et al.*, 1986) amino acid residues of that molecule. Both of these groups recognised a region of sequence from the work of Périn *et al.*, (1984).

The primary structure of the entire link protein from the cartilage proteoglycan aggregate of both the rat chondrosarcoma (amino acid sequencing: Neame, Christner & Baker, 1986) and chicken cartilage (cDNA sequencing: Deak *et al.*, 1986) have been determined. It was observed that a sequence homology exists between the link protein and the proteoglycan core protein, and it has been suggested that this homologous region is involved in the binding of

both proteoglycan and link protein to hyaluronic acid (Neame *et al.*, 1985).

The amino acid sequence of a small chondroitin sulphate proteoglycan from a rat yolk sac tumor, which was deduced from the cDNA sequence, was the first complete proteoglycan sequence to be published (Bourdon *et al.*, 1985). Subsequently the sequence of the precursor form of this proteoglycan has also been determined (Bourdon *et al.*, 1986). This proteoglycan has a central region of its sequence composed of 49 alternating serine and glycine residues of which at least 14 serines have chondroitin sulphate chains attached.

1.5 Objective of this Investigation

With the aim of completing the characterisation of a PDS in order that studies of function could proceed from a better understanding of the molecule, the objective of this investigation was to determine the amino acid sequence of the protein core of the PDS from bovine skin. As this proteoglycan has a relatively low carbohydrate content, it was hoped that fewer problems would be encountered than in the study of the core of one of the more highly glycosylated proteoglycans.

2. MATERIALS AND METHODS

2.1 MATERIALS.

2.1.1 Enzymes.

Chondroitinase ABC (EC 4.2.2.4) from *Proteus vulgaris* and *Staphylococcus aureus* V8 protease were from Miles Scientific, Rexdale, Ontario. TPCK-trypsin (EC 3.4.4.4) was from Worthington, Freehold, NJ, U.S.A., and carboxypeptidase Y was from Pierce Chemical Co., Rockford, Illinois, U.S.A. Cathepsin D (EC 3.4.23.5) was isolated from bovine thymus and purified in this department (Scott & Pearson, 1978).

2.1.2 Chromatography.

Sepharose® CL-4B, Sephacryl® S-300 and S-400 and Sephadex® G-25, G-50 and G-75 were from Pharmacia, Dorval, Quebec. DEAE-cellulose was from Whatman, Maidstone, Kent, U.K., and Fractogel® TSK-HW-50-F was from EM Science, Gibbstown, NJ, U.S.A. Cellulose TLC plates (#13255) were from Eastman Kodak, Rochester, NY, U.S.A., and Schleicher & Schuell Micropolyamide A1700 plates were obtained from Pierce.

2.1.2.1 High pressure liquid chromatography

The Gilson HPLC, used for peptide separations, which consisted of model 302 pumps and an Apple II plus micro computer controller with gradient manager 702 V 1.2 software, was obtained from Gilson Medical Electronics, Middletown, WI, U.S.A. A variable wavelength UV detector (model ERC 7210), used with the Gilson chromatography system, was from Erma Optical Works, Tokyo, Japan. The Beckman Liquid Chromatograph Model 332, used for the PTH-amino acid analyses, was from Beckman Instruments, Mississauga, Ontario. The WISP® model 710B automated sample processor used with the Beckman HPLC was from Waters, Mississauga, Ontario. The FPLC system, comprising P-500 pumps, a GP-250 programmer, UV-1 and UV-1(214) detectors and

FRAC-100 fraction collector was from Pharmacia. HPLC grade acetone, acetonitrile and methanol were from Caledon, Georgetown, Ontario or Fisher Scientific, Edmonton. Tetrahydrofuran was from Burdick and Jackson, Muskegon, Michigan, U.S.A. Pro-RPC, Pep-RPC, Poyanion-SI and Superose-12 prepacked columns were from Pharmacia. Vydac® RP-201-TP and RP-214-TP prepacked columns were from The Separations Group, Hesperia, CA, U.S.A. Altex ODS and Ultrapore RPSC (C_3) prepacked columns were from Beckman and the Zorbax® Bioseries PTH column was from Dupont, Wilmington, DE, U.S.A.

2.1.3 SDS-Polyacrylamide Gel Electrophoresis and Immunoblotting.

Protein standards (low molecular weight range, 10,000-100,000 kDa), electrophoresis grade acrylamide, Bis, SDS, N,N,N',N'-tetramethylethylenediamine, Tris, and urea were from Bio-Rad Laboratories, Mississauga, Ontario. 2-mercaptoethanol was from Sigma Chemical Co., St. Louis, MO, U.S.A. Coomassie Blue R250 was from Bio-Rad, Toluidine Blue from Fisher, Bromophenol Blue from B.D.H., Toronto, Ontario, Basic Fuchsin from Baker, Phillipsburg, NJ, U.S.A., Pyronin Y from Matheson, Cincinatti, OH, U.S.A. and Amido Black was from Dr. G. Grübler & Co., Leipzig, D.D.R. Periodic acid was from G.S. Smith Co., Columbus, OH, U.S.A. Nitrocellulose was obtained from Biorad. Bovine serum albumin and diaminobenzidine tetrahydrochloride were from Sigma. Peroxidase-conjugated rabbit anti-mouse Ig G was from Cooper Biomedical, West Chester, PA, U.S.A. The monoclonal antibodies to PDS were raised in this department, and kindly made available to me by Dr. G.A. Pringle (Pringle *et.al.*, 1985). The electrophoresis cell (a Protean®, 16cm cell) was from Bio-Rad. The power supply used with this cell was from Buchler, Fort Lee, NJ, U.S,A. The trans-blot cell and 160/1.6 power supply were from Bio-Rad.

2.1.4 Sequencing.

The 890C amino acid sequencer was from Beckman. The Sequemat® SC-510 Programmer and P-6 Autoconverter were from Genetic Design, Watertown, MA, U.S.A. The

model 470A gas-phase sequencer was from Applied Biosystems, Foster City, CA, U.S.A. The routine sequencer reagents, phenylisothiocyanate, heptane, 1.0M Quadrol® TFA buffer in n-propanol/water (3:4 v/v) pH 9.0, heptafluorobutyric acid, benzene, ethyl acetate and butyl chloride were from Beckman or Pierce. Sequanal grade trifluoroacetic acid, 4-N,N-dimethylaminoazobenzene 4'-isothiocyanate, pyridine, butyl acetate, n-propanol, and Polybrene® were from Pierce. 1,2-dichloroethane (distilled in glass) was from Caledon. O-phthalaldehyde was from Sigma. The standard PTH-amino acid kit was from Pierce. The vacuum centrifuge (Speedvac®) was from Savant Industries, Hicksville, NY, U.S.A.

2.1.5 Other Chemicals, Apparatus & Sundries.

Benzamidine HCl, phenylmethylsulphonyl fluoride, N-ethylmaleimide and 6-aminohexanoic acid, were from Sigma. Leupeptin and pepstatin were from Institut Armand-Frappier, Laval, Quebec. Trifluoromethanesulphonic acid and cyanogen bromide were from Aldrich, Milwaukee, WI, U.S.A. Sperm whale apomyoglobin was from Beckman and cytochrome c from Sigma. Iodoacetic acid, dithiothreitol, Triton-X-100, fluorescamine and Tris were obtained from Sigma. Sequanal grade guanidinium chloride, constant boiling HCl, constant boiling HCl containing 0.1% (w/v) stannous chloride, triethylamine and 4N methanesulphonic acid containing 0.2% 3-(2-aminoethyl) indole were from Pierce. Hydrogen iodide was from Anachemia, Mississauga, Ontario, cetylpyridinium chloride was from Calbiochem, La Jolla, California, U.S.A. Formic acid was from Eastman Kodak.

The Wiley mill (Model 4) was obtained from A.H. Thomas Co., Philadelphia, PA, U.S.A. The Minitan® concentrator and Milli-Q® water system were from Millipore, Mississauga, Ontario. The ultrafiltration apparatus and membranes were from Amicon, Oakville, Ontario and dialysis membranes were from Spectrum Medical Industries, Los Angeles, CA, U.S.A. Flat bed electrophoresis apparatus was from Pharmacia. The hydrogen fluoride apparatus (type 1) was from Protein Research Foundation, Osaka, Japan. Protein sequence homology and data bank searches were undertaken using MicroGenie from Beckman and IFind® from Intelligenetics, Palo Alto, CA, U.S.A.

All other reagents and supplies were from Fisher Scientific or Canlab in Edmonton.

2.2 METHODS

2.2.1 Extraction and Purification of Proteodermatan Sulphate

This was based on the method of Pearson & Gibson (1982). One kilogram of adult bovine skin, which had been mechanically treated to remove the greater part of the epidermis, was cut into 1cm squares, mixed with dry-ice and ground to a powder in a Wiley mill which had been cooled with liquid nitrogen. The powder was extracted by shaking for 24 hours at 4°C with 30 litres of 7M urea/0.05M Tris-HCl, pH 6.5/0.15M sodium chloride/0.02% (w/v) sodium azide, which contained the following protease inhibitors: 0.05M benzamidine hydrochloride/0.001M phenylmethylsulphonyl fluoride/0.01M disodium ethylenediaminetetraacetate/500μM N-ethylmaleimide/0.1M 6-aminohexanoic acid/5μg/ml pepstatin/5μg/ml leupeptin. Insoluble material was removed by centrifugation (17,000xg) and the supernatant (20 litres) was reduced to 4 litres on a Minitan concentrator (10KDa cut-off membrane). This sample was loaded onto a 5cm x 46cm DEAE-cellulose column, eluted initially with extraction buffer, to remove protein and hyaluronic acid, and subsequently with the same buffer in which the sodium chloride content had been increased to 2M in order to release the more anionic material including the proteoglycans. Uronic acid-containing (Rosenthal, *et. al.*, 1976) fractions were pooled and the crude proteoglycan preparation was precipitated by the addition of 3 volumes of cold ethanol. The precipitate was dissolved in 0.5M sodium chloride and cetyl pyridinium chloride was added to 0.1% (w/v). The precipitate obtained at this stage, containing dermatan sulphate and chondroitin sulphate-bearing material, was dissolved in 1M magnesium chloride and three volumes of ethanol were added. The resulting precipitate which was then free of cetyl pyridinium chloride was dissolved in 7M urea/0.05M sodium acetate/0.05M Tris-HCl, pH 6.5, and was loaded onto a 5cm x 90cm column of Sepharose CL-4B, which was eluted with the same buffer. Two peaks from this chromatography contained uronic acid. The second peak proved to be proteodermatan sulphate and a pool was selected, for sequencing, that was from the central region of that peak. The proteoglycan in this pool was precipitated by the addition of three volumes of ethanol. The

precipitate thus obtained was dissolved in water and then lyophilised. The homogeneity of the proteoglycan was confirmed by amino acid analysis, SDS-polyacrylamide gel electrophoresis, and cellulose acetate electrophoresis of the glycosaminoglycans released by digestion with papain.

2.2.2 Deglycosylation of Proteodermatan Sulphate

Several deglycosylation methods were employed, each with a different cleavage mechanism. The comparison of the various cores, thus produced, provided information about the type of glycosylation and the site of attachment of the carbohydrate moieties of this molecule (see Fig. 4).

2.2.2.1 Chondroitinase ABC

Following the method of Oike *et al.* (1980), PDS, at 3mg/ml was incubated in 0.03M sodium acetate/0.1M Tris-HCl, pH 8.0, in the presence of chondroitinase ABC (1.0 unit/ml), for 45 minutes at 37C. The following proteolytic inhibitors were included in the digestion buffer: 0.01M N-ethylmaleimide/0.01M di-sodium ethylenediaminetetraacetate/0.005M phenylmethylsulphonyl fluoride/360μM pepstatin. The reaction was terminated by the addition of an equal volume of 8M guanidinium chloride, 0.4% Triton X-100, containing an equal concentration of the inhibitors. The resulting small carbohydrate cleavage products (largely disaccharides) were removed by repeated concentration/dilution of the retentate in an Amicon ultrafiltration apparatus (PM10 membrane, 10KDa cutoff) until a theoretical dilution factor of at least 10,000 was achieved. The PDS(ABC)core was then lyophilised. Subsequently, in analytical scale deglycosylations (*e.g.* for SDS-PAGE), the protease inhibitors were omitted.

2.2.2.2 β-elimination/sulphite addition

PDS(ABC)core was dissolved at 1mg/ml in 0.2M sodium sulphite, the pH of which had been adjusted to 11.5 with sodium hydroxide (Isemura *et. al.*, 1981). The reaction mixture was incubated for 24 hours at room temperature, with occasional adjustment to

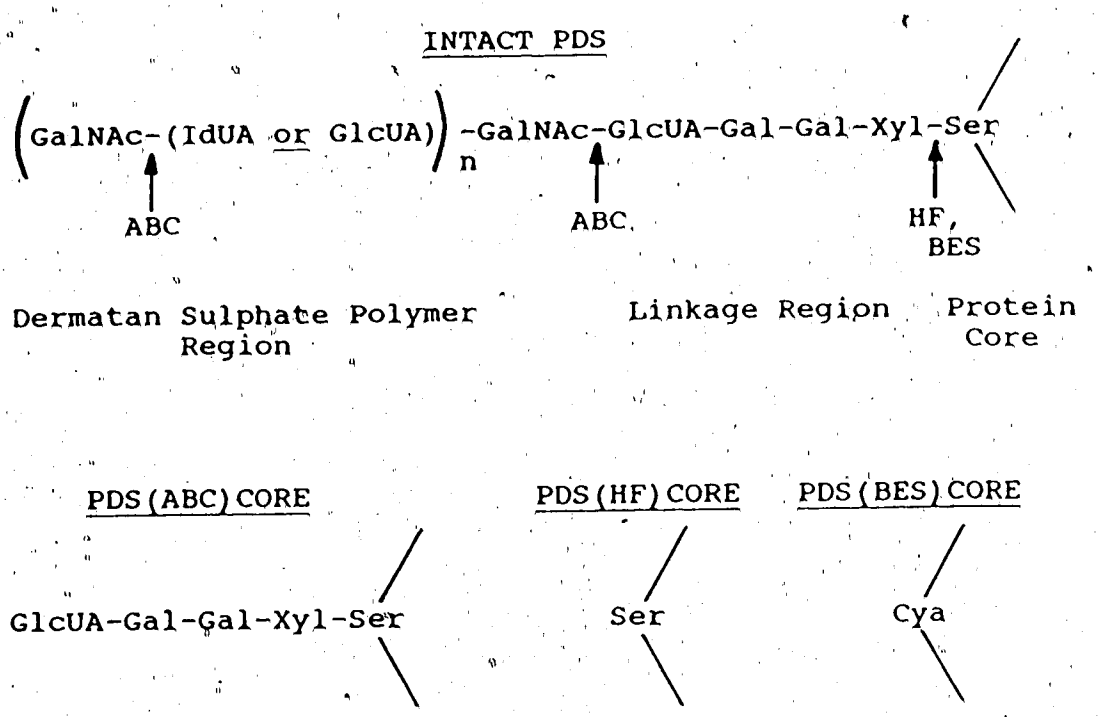


Fig. 4. This diagram shows the structure of the glycosaminoglycan side chain attached to serine-4 of the protein core of PDS, the cleavage sites of the various deglycosylation agents used in this project, and the theoretical products of deglycosylation. Chondroitinase ABC (ABC) cleaves the GAG into disaccharides. It is believed that the last one or two sites fail to cleave leaving one or two of the disaccharides attached to the linkage region. Anhydrous HF (HF) cleaves all O-glycosidic linkages in the GAG (though only one is shown) and any glycoprotein-like oligosaccharide side chains, producing monosaccharides. N-glycosidic linkages are resistant to this treatment and a single glucosamine residue is left attached to the linkage asparagine at the site of each N-linked oligosaccharide. (The enzyme N-glycanase cleaves this N-linkage releasing the oligosaccharide intact and converting the linkage asparagine to aspartic acid.) β-elimination/sulphite addition (BES) cleaves the xylosyl-serine linkage releasing the intact GAG (or GAG stub in this case as PDS(ABC)core was the starting material for the preparation of the PDS(BES)core). IdUA - iduronic acid; GlcUA - glucuronic acid; Gal - galactose; Xyl - xylose.

maintain the pH at 11.5. The reaction was terminated by adjusting the pH to 7.0 with acetic acid. The cleaved glycosaminoglycan was removed by repeated

concentration/dilution in an Amicon ultrafiltration apparatus, as described above for the preparation of PDS(ABC)core. The PDS(BES)core was then lyophilised.

2.2.2.3 Anhydrous hydrogen fluoride.

Following the method of Mort and Lamport (1977), and using a closed-system apparatus designed for use in reactions involving HF, 10ml of anhydrous hydrogen fluoride was distilled into the liquid nitrogen-cooled, teflon reaction vessel containing the PDS sample (15mg, dried over phosphorus pentoxide). The vessel temperature was raised to 0°C (ice water bath) and the reaction was allowed to continue for 2 hours with magnetic stirring. The hydrogen fluoride was removed under gentle vacuum with continued stirring, to prevent bumping, then a higher vacuum was applied for a further hour to ensure total removal. The PDS(HF)core, which is not readily water soluble, was removed from the reaction vessel in 4M guanidinium chloride. Low molecular weight digestion products and the guanidinium chloride were removed by ultrafiltration as described above.

2.2.3 Chemical Modifications of PDS

2.2.3.1 Reduction and alkylation

Following the method of Hardingham *et al.* (1976), a sample of PDS (4mg/ml) in 4M guanidinium chloride/ 0.05M Tris HCl, pH 9.0, was incubated at 37°C for 3 hours in the presence of 0.01M dithiothreitol. The solution was made 0.04M in iodoacetic acid and incubated at room temperature for 16 hours in the dark. The solution was then dialysed or subjected to cycles of concentration and dilution in an ultrafiltration apparatus (PM10 membrane). The sample was then lyophilised.

2.2.3.2 Performic acid oxidation

Performic acid was prepared by mixing 1 part of 30% hydrogen peroxide with 9 parts of 88% formic acid and letting the mixture stand for 2 hours at 25 °C (Moore, 1963). The performic acid solution was then cooled to 0°C and 0.5ml aliquots were added to samples of approximately 500μg. Samples were incubated at 0°C for 60-90 mins. The

reaction was terminated by the addition of 15 volumes of water followed by freezing and lyophilisation.

2.2.4 Analytical Methods

2.2.4.1 Determination of protein content

Protein content was determined by the method of Lowry *et al.* (1951), as follows: To 1ml of sample or standard (2-100μg Bovine Serum Albumin per ml) was added 5 ml of 2%(w/v) sodium carbonate/0.01%(w/v) copper sulphate/0.02%(w/v) sodium potassium tartrate/0.1M sodium hydroxide. After 10 minutes, 1ml of 0.5N Phenol Reagent (Folin-Ciocalteau's) was added. The solutions were thoroughly mixed and allowed to stand for 40 minutes, after which the absorbance of the solutions was read at 740nm.

Occasionally a Bio-Rad protein assay kit was used; however, the protein values obtained from amino acid analyses were probably more accurate than the Bio-Rad kit or the Lowry values.

2.2.4.2 SDS Polyacrylamide gel electrophoresis

The method of Weber & Osborn (1975) was used. Samples were dissolved in a sample solvent comprising 0.007M sodium phosphate/0.15% SDS/1.35M urea/1% 2-mercaptoethanol/0.17% Bromophenol Blue and boiled for 3 minutes. The gel slabs, of various acrylamide contents (Bis:acrylamide, 1:30, 0.15% (v/v) TEMED/0.025% (w/v) ammonium persulphate), and the electrophoresis buffer contained the same buffer solution: 0.1% SDS/0.05M sodium phosphate, pH 7.2. The slabs were electrophoresed at 50V for the first 30 minutes, or longer if necessary to ensure that all of the sample was in the gel, then at 105V for the duration of the run. Earlier in the project, this method was also run in tube gels which are more amenable to scanning in a system available in this laboratory.

Molecular weight standards used were as follows:

1, phosphorylase B (92.5kDa); 2, bovine serum albumin (66.2kDa); 3, ovalbumin

(45kDa); 4. carbonic anhydrase (31kDa); 5. soybean trypsin inhibitor (21.5kDa); 6. lysozyme (14.4kDa).

2.2.4.3 Electrophoresis in highly cross-linked urea/SDS polyacrylamide gels

Following the method of Swank & Munkres (1971), samples were dissolved in 1% SDS/8M urea/1% 2-mercaptoethanol/0.01M phosphoric acid adjusted to pH 6.8 with Tris, and boiled for 5 minutes prior to electrophoresis. Samples were made 10% (w/v) in sucrose and 0.01% (v/v) Bromophenol Blue, prior to electrophoresis. Acrylamide gels (12.5% (w/v), Bis:acrylamide, 1:10) were made, containing 0.1% (w/v) SDS/0.075% (v/v) TEMED/0.07% (w/v) ammonium persulphate/8M urea/0.1M phosphoric acid, adjusted to pH 6.8 with Tris. The electrophoresis buffer was 0.1% (w/v) SDS/0.1M phosphoric acid, adjusted to pH 6.8 with Tris. Electrophoresis at 1-2 mA per tube gel required an overnight run. Cyanogen bromide digests of cytochrome c and sperm whale apomyoglobin were prepared as low molecular weight standards (1,800-15,000 KDa).

2.2.4.4 Staining of SDS-polyacrylamide gels.

Coomassie Blue. The position of the Bromophenol Blue band was marked with India Ink and the slab was then stained in 0.125% (w/v) Coomassie Blue R250/50% (v/v) methanol/7.5% (v/v) acetic acid for one hour (Mechanic, 1979). The gel was then destained in 7.5% (v/v) acetic acid/5% (v/v) methanol until the background was almost clear. Destaining was completed and the gel was stored in 10% (v/v) acetic acid.

Toluidine Blue. The positon of the Bromophenol Blue band was marked as above and the slab was stained in 0.05% (w/v) Toluidine Blue for one hour. Destaining was initiated in 3% (v/v) acetic acid, until the stained bands could be clearly discerned, and continued in water.

Periodic Acid-Schiff. (Konat *et al.*, 1984) Gels were fixed overnight in 40% (v/v) ethanol/5% acetic acid and then transferred to 0.7% (w/v) periodic acid/5% (v/v) acetic acid for 2-3 hours. Excess periodate was reduced in 0.2% (w/v) sodium metabisulphite/5% (v/v) acetic acid for 2-3 hours, with one change of solution after 30 minutes. The

oxidised glycoproteins were then stained in Schiffs reagent overnight. Staining was enhanced in 0.2% (w/v) potassium metabisulphite/40% (v/v) ethanol/5% (v/v) acetic acid, for 90 minutes at 55°C. Destaining in 40% (v/v) ethanol/5% (v/v) acetic acid required about 24 hours.

2.2.4.5 Immunoblotting

Three well characterised monoclonal antibodies, raised against PDS, were made available to me by Dr. G.A. Pringle. These antibodies, 6D6, 7B1 and 3B3, have been shown to have no affinity for collagen types I, II, III or IV, fibronectin, or dermatan sulphate. The epitopes recognised by these antibodies are distinct from each other and are on the protein core of PDS (Pringle *et al.*, 1985). The epitope recognised by 7B1 is also known to be species specific and near the attachment site of the dermatan sulphate chain, probably in the region of residues 9-20 of the protein core (Pringle, 1985). Another monoclonal antibody, 5D1, was also made available to me. This antibody was also raised against PDS but had not been thoroughly characterised.

Following Pringle (1985), the Bromophenol Blue dye front was marked with Pyronin Y and the gel transferred onto a sheet of nitrocellulose paper which had been soaked in 0.05M sodium phosphate buffer, pH 7.2. The protein bands were then transferred to the nitrocellulose in a trans-blot cell at 400mA overnight. The gel was stained in Coomassie Blue, as above, to check the efficiency of the transfer. The nitrocellulose paper was then cut into strips. One strip, containing the protein standards was stained in Amido Black (0.5% (w/v) in 7% (v/v) acetic acid, destained in 3% (v/v) acetic acid). The other strips were stained with monoclonal antibodies raised against the intact PDS. The nitrocellulose strips were first incubated with Tris-buffered saline (TBS: 0.15M sodium chloride/0.05M Tris HCl, pH 7.6) containing 2% (w/v) bovine serum albumin, at 40°C for 45 minutes. They were then washed three times with TBS. Each strip was then incubated, at room temperature, for two hours with 4ml per track of primary antibody solution, comprising 4µl monoclonal ascites fluid, 2ml TBS, 0.4ml normal goat serum and 1.6ml 2% (w/v) bovine serum albumin in TBS. The strips were washed four

times in TBS and incubated with the secondary antibody (peroxidase conjugated rabbit anti-mouse IgG, 1:600) in the same solution used for the primary antibody. Four washes with TBS were then followed by incubation in 6ml per track of 0.05% (w/v) diaminobenzidine tetrahydrochloride/ 0.01% (v/v) hydrogen peroxide, in TBS until the desired staining intensity developed. The strips were then washed thoroughly in running water and dried between paper towels.

2.2.4.6 Thin-layer electrophoresis/thin-layer chromatography.

Kodak #6064 cellulose plates (10cm x 10cm) were loaded in one corner with samples of peptide (2-3nmoles in 5μl). They were then sprayed with acetic acid/formic acid/water (75:25:400) and electrophoresed in the same buffer, on a Pharmacia Flat-Bed Electrophoresis system, at 950V for 12 minutes. The plates were dried with hot air and subjected to ascending chromatography, in the second dimension, in butanol/pyridine/acetic acid/water (35.2:25:5:20). The plates were washed with HPLC-grade acetone and dried. They were then sprayed with 1% (v/v) triethylamine in acetone and, whilst still wet, they were sprayed with fluorescamine (0.1mg/ml in acetone). Drying, acetone washing and drying again, were followed by observation under long wavelength ultraviolet light.

2.2.4.7 Total acid hydrolysis

Protein samples of up to 200μg in acid-washed, screw-cap culture tubes (10x100mm), were dissolved in 200μl of constant-boiling hydrochloric acid containing 0.2% (w/v) phenol. The tubes were sealed under nitrogen and incubated at 110°C for 22 hours. The hydrolysates were dried down and submitted for amino acid analysis. If the presence of homoserine was suspected, the dried hydrolysate was treated to ensure that all the homoserine was in the open-chain form. The sample was dissolved in 80μl of 0.01M hydrochloric acid. Sodium hydroxide (10μl, 2M) was added, and after shaking, the sample was allowed to stand for 5 minutes before neutralisation by the addition of 10μl of 2M hydrochloric acid. The sample was then analysed as soon as possible.

2.2.4.8 Hexosamine hydrolysis

Samples, in tubes as above, were hydrolysed in 200μl of 4M hydrochloric acid at 100°C for 18 Hours.

2.2.4.9 Methanesulphonic acid hydrolysis

Samples, in tubes as above, were hydrolysed in 300μl 4N methanesulphonic acid containing 0.2% 3-(2-aminoethyl) indole at 115°C for 22, 48 and 72 hours (Simpson, *et.al.*, 1976). A graph of tryptophan released versus time was extrapolated to zero time.

2.2.4.10 Amino acid analysis

Amino acid analyses were carried out on a Beckman 121MB amino acid analyser using a method which separates glucosamine and galactosamine as well as the amino acids.

2.2.5 Protein Cleavage Methods

2.2.5.1 Cyanogen bromide

Protein at 1mg/ml in 70% (v/v) formic acid was incubated under nitrogen, at room temperature for four hours in the presence of cyanogen bromide (12mg/ml). The reaction was terminated by dilution with 20 volumes of water followed by lyophilisation.

2.2.5.2 Cathepsin D

Protein was dissolved at 1mg/ml in 0.05 M ammonium formate, pH 4.0. This solution was warmed to 37°C and enzyme solution, prepared in this laboratory, was added (enzyme:substrate, 1:666, w/w). The mixture was incubated at 37°C for 30 minutes and digestion was terminated by freezing and lyophilisation.

2.2.5.3 Trypsin

Protein at 2mg/ml was dissolved in water and the solution boiled for 1 minute to ensure complete denaturation. An equal volume of 2% (w/v) ammonium bicarbonate, pH 8.0, and sufficient TPCK-treated trypsin to produce an enzyme to substrate ratio of 1:100

(w/w) were added. The mixture was incubated at 25°C for 3 hours and then frozen and lyophilised.

2.2.5.4 Staphylococcus aureus V8 protease

Following the method of Drapeau *et. al.*, (1972) protein, at 1mg/ml in 0.1M ammonium bicarbonate/ 0.002M disodium ethylenediaminetetraacetate, pH 7.8 was incubated at 37°C for 18 hours (enzyme:substrate, 1:30, w/w). Digestion was terminated by freezing and lyophilisation.

2.2.6 Separation of Peptides

2.2.6.1 Sephacryl S-300

Sephacryl S-300, a gel filtration medium, eluted with 0.01% trifluoroacetic acid was used in the initial separation of the CNBr peptides of the PDS(ABC)core.

2.2.6.2 Pro-RPC

Pro-RPC, a large-pore, C_1/C_4 reversed-phase chromatography column, eluted with acetonitrile gradients in the presence of heptafluorobutyric acid was used in the preparation of the two CNBr peptide groups, A and B, and in the preparation of CB-6.

2.2.6.3 201-TP

Vydac RP-201-TP, a large-pore, C_{18} reversed-phase chromatography column, eluted with acetonitrile gradients in the presence of trifluoroacetic acid, was used in the separation of small peptides (1-50 amino acids) such as CB-7 and CB-8 and the tryptic and V8 digests.

2.2.6.4 Polyanion-SI

A cathepsin D-digest of PDS was chromatographed on Polyanion SI, an anion exchange medium, eluted with a linear gradient of 0.15M to 2.0M NaCl in 7M Urea/0.05M Tris HCl, pH 6.5. All the retarded material was collected as a single pool referred to as

"CD-N".

2.2.7 Sequencing Methods

2.2.7.1 Manual sequencing

Several manual sequencing methods were tried before any success was achieved (Edman & Henschen, 1975; Tarr, 1977). The following method, based on the method of Chang (1983) was used successfully in the sequencing of several short tryptic peptides. Peptide (2-3nmoles) in a muffle furnace-cleaned, 6 x 50mm pyrex tube was dissolved in 10μl of Milli-Q water. Twenty microlitres of 4-N,N-dimethylaminozobenzene 4'-isothiocyanate (DABITC) in pyridine (1.41mg/ml) was added, the tube was centrifuged, sealed under nitrogen, vortexed and incubated at 55°C for 20 minutes. PITC (2μl) was added, the tube was centrifuged, sealed under nitrogen, vortexed and incubated at 55°C for a further 20 minutes in the dark. Heptane: ethyl acetate (2:1, 300μl) was added, the tube was vortexed and centrifuged to separate the phases. The upper phase, containing unreacted coupling reagents, was removed with a drawn-out Pasteur pipette and discarded. This extraction was repeated twice, the third extraction used only 150μl of solvent and the tube was not vortexed. The necks of the tubes were wiped with a Kimwipe, which had been soaked in butyl acetate, and then solvent was removed on the vacuum centrifuge.

Cleavage was effected by the addition of 40μl of trifluroacetic acid (TFA), sealing the tube under nitrogen and incubating at 55°C for 12 minutes. The TFA was removed by evaporation in a stream of nitrogen followed by drying on the vacuum centrifuge. The dried residue was dissolved in 10μl of water. Butyl acetate (150μl) was added, the solution was mixed, centrifuged and the upper phase, containing the DABTZ-amino acid, was removed to a clean tube. The aqueous phase was dried on the vacuum centrifuge and reserved for the second sequencing cycle. The organic phase was dried on the vacuum centrifuge, converted in 50μl of 50% (v/v) TFA , under nitrogen for 15 minutes at 70°C and redried. The cleaved DABTH-amino acids were identified by TLC (see below).

2.2.7.2 Liquid-phase automated sequencing.

The majority of automated sequencing was undertaken in this department on a Beckman 890C sequencer (based on the work of Edman & Begg, 1967). During the course of this project the instrument was upgraded by the addition of a cold trap in the vacuum system, a Sequemat SC-510 Sequential Controller and a Sequemat P-6 Auto-Converter. The amount of protein required to produce useful sequence data decreased with successive improvements to the instrument reaching the 5-10 nmol level by the end of this project. The conversion reagent used by the auto-converter was a methanolic solution of hydrochloric acid made by the dropwise addition of one part of cold acetyl chloride to seven parts of cold methanol. The extraction solvent was a mixture of 1,2-dichloroethane and methanol (7:3). The routine sequencer programme used on this project was a single coupling, single cleavage method. Double coupling was used at the first cycle for longer peptides and double cleavage was used whenever a residue was known from a previous run, to be proline. Polybrene (3mg) was used in all runs.

2.2.7.3 Precycling of Polybrene

In order to remove any impurities from the Polybrene, which could have been extracted with the derivatised amino acid, it was subjected to six cycles in the sequencer. 55mg of Polybrene was dissolved in 770μl of water. Of this, 700μl was loaded into the sequencer cup as for a protein run. Six sequencer cycles were run and the dried Polybrene was then dissolved in 5ml of water and removed. 300μl aliquots, containing approximately 3mg each, were stored in vials at -20°C until required.

2.2.7.4 Attempted suppression of background in automated sequencing using o-phthalaldehyde

This experiment was undertaken in an attempt to reduce the PTH-amino acid background signal in late cycles of sequencer runs and allow the recognition of more sequence data. O-phthalaldehyde (OPA) reacts with amino groups but not imino groups. Consequently, all the common amino acids except proline will react with this reagent.

O-phthalaldehyde labelled amino-terminal-amino acids are not cleaved in the Edman degradation.

Sequencing was interupted at cycles where Proline was known, from previous sequencer runs, to be the amino-terminal residue. Following the method of Brauer *et.al.*, (1984) the sequencer was programmed to deliver OPA to the spinning cup, allow the reaction of OPA with any non-proline residues and wash out any unbound OPA. Sequencing was then continued.

2.2.7.5 Gas-phase automated sequencing.

Several samples, especially those available only in small amounts, were analysed on the Applied Biosystems model 470A gas-phase sequencer in the Biochemistry Department at this university. This instrument, recently upgraded by the addition of automated, in-line HPLC analysis of the PTH-amino acids (Applied Biosystems model 120A PTH Analyser), is capable of routine operation with samples of only 100 pmol.

2.2.7.6 Carboxy-terminal sequencing with carboxypeptidase Y

Based on the method of Hayashi (1977), samples were exhaustively dialysed against Milli-Q water to remove free amino acids, lyophilised and dissolved in 0.05M pyridine acetate buffer, pH 5.5 containing norleucine as an internal standard. Carboxypeptidase Y at an enzyme:substrate ratio of 1:400 was added and digestion allowed to procede at 37°C. Aliquots were removed at various intervals over 24 hours and frozen. They were subsequently analysed on the amino acid analyser.

2.2.8 Identification of Amino Acid Derivatives Produced by Sequential Degradation of Proteins and Peptides.

2.2.8.1 High-pressure liquid chromatographic analysis of PTH-amino acids.

Several methods have been used during the project (Zimmerman, *et. al.*, 1977; Somack, 1980; Tarr, 1981; Glajch,*et.al.*, 1985). The current routine method (Glajch,*et.al.*, 1985) uses a Zorbax BioSeries PTH (CN) column (see Fig. 5). Two eluents (A and B) are

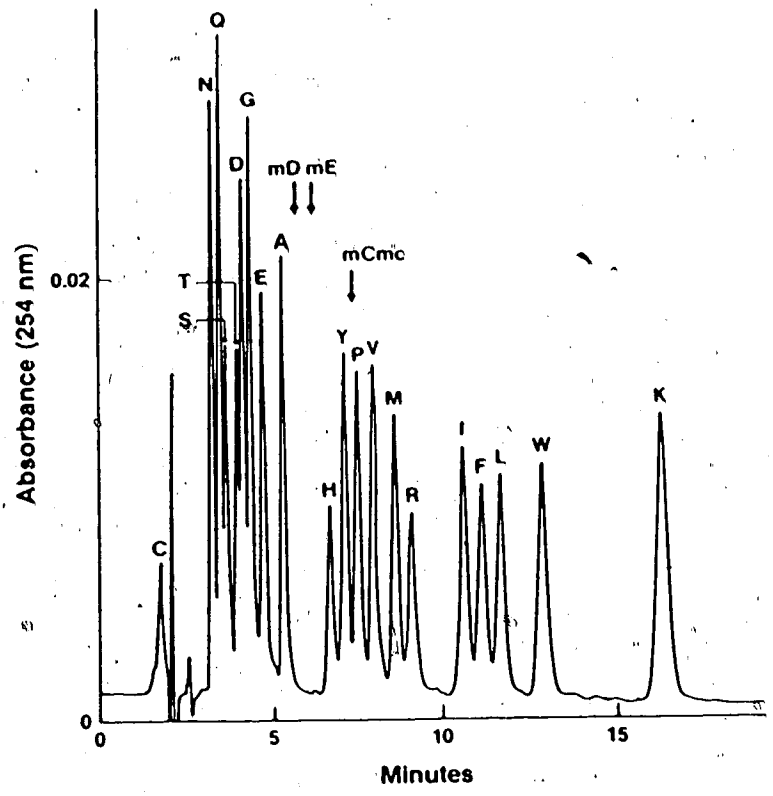


Fig. 5. Chromatography of 0.5nmoles of each of the PTH-amino acids on a Zorbax BioSeries PTH column. The elution positions of methylaspartic acid, methylglutamic acid and methylcarboxymethylcysteine phenylthiohydantoins, made in this laboratory, are shown.

made fresh daily as follows: acetonitrile/tetrahydrofuran/stock buffer (18:16:66). Two stock buffers are used (16mM and 4mM phosphoric acid, pH adjusted to 3.15 with sodium hydroxide) to prepare eluents A and B respectively. The optimum separation of PTH-amino acids is achieved at approximately 80% B. The exact proportion is determined from the first standard of the day. An adjustment of ±2% B is often required.

2.2.8.2 Thin layer chromatographic identification of DABTH-amino acids produced by manual sequencing.

Two-dimensional TLC on 2.5cm square, double-sided, micropolyamide plates was used to identify the DABTH-amino acids released during manual sequencing. The first dimension was eluted in acetic acid:water (1:2), the second in toluene:n-hexane:acetic acid (2:3:1). (see Fig. 6)

DABTH-amino acid standards are made as follows:

To 100nmoles of amino acid, lyophilised in a 6x50mm test tube, add 50μl of buffer (3.56mM triethylamineacetic acid pH 10.65, 45% acetone) and 25μl of DABITC solution (4μM in acetone), seal and incubate at 54°C for 60 minutes. Dry on the vacuum centrifuge. Redissolve in 50μl 50% TFA, seal and incubate at 54°C for 60 minutes. Dry on the vacuum centrifuge and dissolve in 80% ethanol for spotting onto TLC plates. The internal standard, DABTC-diethylamine is made by mixing 1ml of DABITC solution (1nmol/μl in ethanol), 10μl diethylamine and 100μl water. This solution, stored at -20°C, is used without dilution.

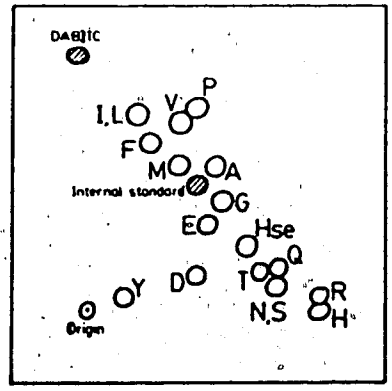


Fig. 6. Two dimensional TLC identification of the DABTH-amino acids produced by manual sequencing. 1: (horizontal) acetic acid:water (1:2); 2: (vertical) toluene:n-hexane:acetic acid (2:3:1)

2.2.8.3 Back hydrolysis of PTH-amino acids.

If problems of identification arose, the PTH-amino acids were "back-hydrolysed" to the free amino acid which could then be identified by amino acid analysis. Two methods were used:

1. 57% hydroiodic acid (Smithies, 1971), 127°C, 18 hours.

2. Constant boiling hydrochloric acid containing 0.1% (w/v) stannous chloride (Lai, 1977), 150°C, 4 hours.

3. RESULTS

3.1 Characterisation of Proteodermatan Sulphate and the Various Glycoprotein Cores derived from PDS

Several deglycosylation methods were employed and the comparison of sequence data from the various cores assisted in the elucidation of the nature of residue 4. Deglycosylation using trifluoromethanesulphonic acid (Edge *et al.*, 1981) was attempted but abandoned as degradation of the protein core was observed.

3.1.1 Proteodermatan Sulphate

Routine analyses, related to lyophilised weight, of the various preparations of PDS used in this project, yielded very consistent values. Protein content averaged 48% by the method of Lowry *et al.* (1951) and values derived from amino acid analyses were always in excellent agreement with these. Uronic acid contents averaged 9% (corresponding to a glycosaminoglycan content of about 32%). Analyses for neutral sugars in PDS, using derivatisation of hydrolysates with dansyl-hydrazine, followed by HPLC determination of the derivatised sugars (Eggert & Jones, 1985), gave values of: galactose, 1.8%; mannose, 1.8%; xylose, 0.2% and fucose, 0.3%. On electrophoresis in 5% SDS-PAGE the apparent M_r, estimated using protein standards, averaged 87,000 (Fig. 7). The amino acid and hexosamine composition was also very consistent (Table 2).

Sequence analysis of intact PDS was not very successful. Six of the first eight residues were tentatively identified (Table 3) from a run in which the sample film in the sequencer cup deteriorated rapidly to a sticky mass which apparently interfered with the smooth flow of reagents necessary for efficient sequencing. Digestion of PDS with carboxypeptidase Y released Leu, Val, Tyr, Arg, Gly, Ser and Ala (See carboxy-terminal sequencing of peptide CB-1 (Fig. 20) which produced almost identical results to the carboxy-terminal sequencing of PDS).

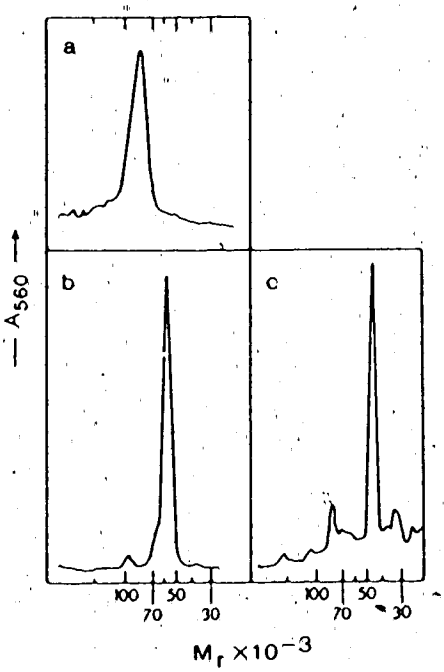


Fig. 7. SDS-polyacrylamide gel electrophoresis. a, proteodermatan sulphate; b, PDS(ABC)core; c, PDS(HF)core. Electrophoresis was carried out at pH 7.2 in 5% gels which were then stained with Coomassie Blue R-250, destained and scanned at 560nm.

TABLE 2
AMINO ACID COMPOSITIONS OF PDS AND CORES

Amino Acid	PDS	(ABC)Core	(BES)Core	(HF)Core
	residues/1000 residues			
Asx	118	119	126	119
Thr	42	43	44	41
Ser	64	64	64	61
Glx	102	103	103	110
Pro	75	68	62	70
Gly	69	69	68	74
Ala	52	53	52	54
Val	62	63	64	57
Met	6	13[1]	4	10
Ile	64	63	63	56
Leu	132	130	133	120
Tyr	27	28	29	26
Phe	30	32	34	31
Lys	85	76	78	89
His	28	27	28	25
Arg	29	29	30	32
Cys	13[2]	18[2,3]	ND	ND
Trp	ND	3[4]	ND	ND
Cya	-	-	-5	
GlcN[5]	30	23	26	9
GalN[5]	109	4	3	5

[1] Determined as methionine sulphone following performic acid oxidation.
[2] Determined as cysteic acid following performic acid oxidation.
[3] Determined as S-carboxymethylcysteine following reduction and alkylation.
[4] From methanesulphonic acid hydrolysis.
[5] Hexosamines expressed as residues/1000 amino acid residues.
ND = Not determined

TABLE 3
SEQUENCE OF THE AMINO TERMINAL REGION INCLUDING RESIDUE 4

sample	Cycle 1	2	3	4	5	6	7	8
PDS	D	E	A/S	-	G	-	G	P
(ABC)Core	D	E	A	-	G	I	G	P
R&A[1] (ABC)Core	D	E	A	-	G	I	G	P
CB-8	D	E	A	-	G	I	G	P
(BES)Core	D	E	A	Cya[2]	G	I	G	P
(HF)Core	D	E	A	S	G	I	G	P

[1] Reduced and alkylated
[2] Cysteic acid

3.1.2 PDS(ABC)Core

The apparent M_r of the PDS(ABC)core, in SDS-PAGE on 5% gels(Fig. 7), was 53,000-55,000. This value was found to be dependent on the concentration of the gel. Consequently the method of Ferguson (1964) was employed to obtain a more reliable estimate. This method is based on the relationship between mobility and gel concentration which can be expressed as:

$$\mathrm{Log}M_R = -K_R C + \mathrm{Log}M_0$$

A plot of Log relative mobility (M_R) for the unknown sample and the standards, at various gel concentrations (C) (Fig. 8) is used to derive the retardation coefficient, K_R (the slope of the graph is $-K_R$). Plotting K_R versus M_r for the standards produces a straight line from which the M_r of PDS(ABC)core was estimated to be 45,000. The y-intercept of the lines in Fig. 8 is M_0, the relative mobility at zero gel concentration. This is a measure of the "normality" of proteins (*ie* their utility as standards in relation to the unknown) as the graphs for the unknown and the standards should converge at a common M_0. This graph suggests that the standards employed in this experiment were, with the exception of phosphorylase B (which was omitted from subsequent calculations), suitable for this purpose.

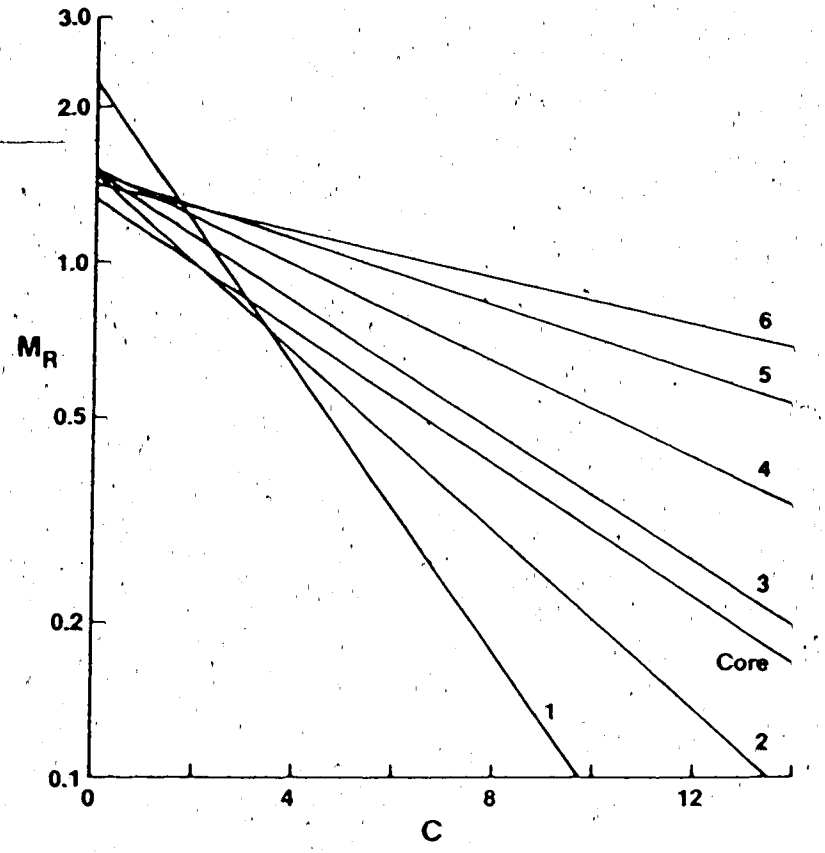


Fig. 8. Ferguson Plot. A plot of relative mobility (M_R) versus gel concentration (C) for PDS(ABC)core (core) in SDS-PAGE. Electrophoresis was carried out at pH 7.2 in gels of 5-12% acrylamide. Standards used were as numbered in section 2.2.4.2.

The PDS(ABC)core showed no significant differences in amino acid composition from the intact proteoglycan (Table 2). The hexosamine composition however, showed a reduction of galactosamine to only 4% of the level found in PDS. Neutral sugar contents were the same as for intact PDS. These data are consistent with the removal of almost all of the dermatan sulphate side chain by chondroitinase ABC. Accurate values for methionine (as methionine sulphone following performic acid oxidation), cysteine (as cysteic acid following performic acid oxidation and as S-carboxymethylcysteine following reduction and alkylation) and tryptophan

(following methanesulphonic acid hydrolysis) were determined for this preparation.

Sequencing identified 26 of the first 30 residues of the PDS(ABC)core and tentatively identified three subsequent residues (Fig. 9). Residue four was not identified from the (ABC)core (Table 3).

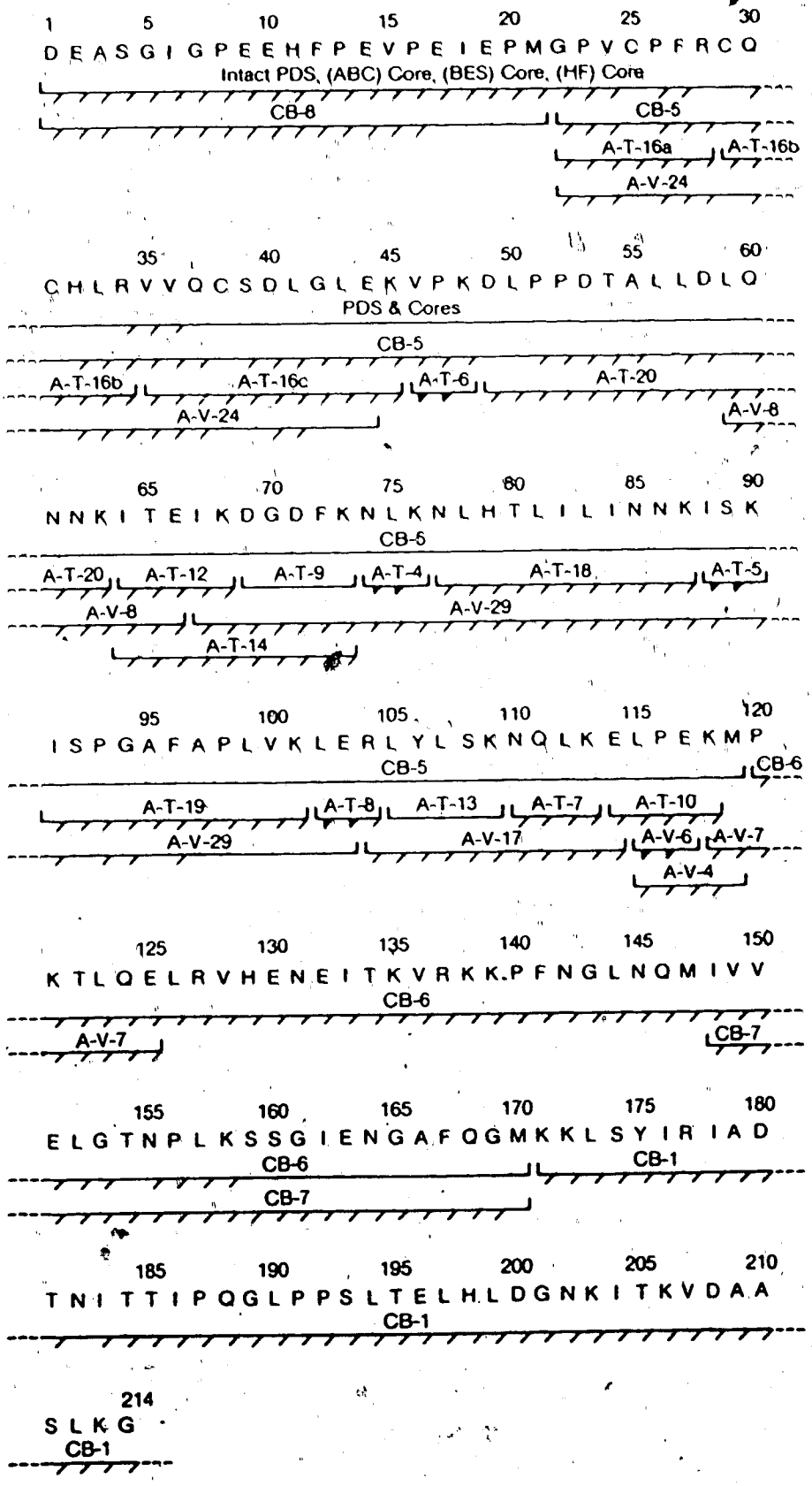
3.1.3 PDS(BES)Core

In SDS-PAGE the PDS(BES)core produced a single band corresponding to the same molecular weight as the (ABC)core. The small change in molecular weight due to the removal of the linkage region of the glycosaminoglycan was not detected. This was probably due to the small size of the carbohydrate that was removed (approximately 2kDa) and the fact that carbohydrate does not bind SDS, such that molecular weight estimates involving carbohydrates are always inaccurate. The amino acid and hexosamine compositions were also identical to those of the (ABC)core, within experimental error, except for the inclusion of five residues of cysteic acid per 1,000 residues (Table 2). This is consistent with the removal of that portion of the dermatan sulphate chain that remains attached to the PDS(ABC)core, and any other O-substituents on serine residues, by β-elimination/sulphite addition. The incorporation of five residues of cysteic acid suggests the removal of 1.6 O-substituents per mole and the expected corresponding decrease in serine content was not observed. These discrepancies fall within the experimental error of amino acid analysis. Sequencing the PDS(BES)core identified 26 of the first 30 residues (Table 3, Fig. 9) including cysteic acid at residue four (Fig. 10). The identification of cysteic acid at residue four indicates that β-elimination, sulphite addition cleaved a substituent which was O-glycosidically linked to serine at this location.

3.1.4 PDS(HF)Core

The apparent M_r of the (HF)core in 5% SDS-PAGE was considerably less than that of the other two cores, at 42,000 (Fig. 7). The amino acid composition again matched that of the intact proteoglycan (Table 2). The hexosamine analysis showed a reduction in the

Fig. 9. A summary of all the sequence data used in the assembly of the amino acid sequence of the amino-terminal 214 residues of PDS. Open ticks indicate data obtained from automated sequencing; closed ticks, manual sequencing. The peptide nomenclature is explained in the text. The data shown on the single line corresponding to intact PDS and the various cores derived from PDS, was largely derived from the (ABC)core and (BES)core. The contributions of the intact PDS and the (HF)core are shown in Table 3.


1 5 10 15 20 25 30
D E A S G I G P E E H F P E V P E I E P M G P V C P F R C Q
Intact PDS, (ABC) Core, (BES) Core, (HF) Core
CB-8
CB-5
A-T-16a
A-T-16b
A-V-24
35 40 45 50 55 60
C H L R V V Q C S D L G L E K V P K D L P P D T A L L D L Q
PDS & Cores
CB-5
A-T-16b
A-T-16c
A-T-6
A-T-20
A-V-24
A-V-8
65 70 75 80 85 90
N N K I T E I K D G D F K N L K N L H T L I L I N N K I S K
CB-5
A-T-20
A-T-12
A-T-9
A-T-4
A-T-18
A-T-5
A-V-8
A-V-29
A-T-14
95 100 105 110 115 120
I S P G A F A P L V K L E R L Y L S K N Q L K E L P E K M P
CB-5
CB-6
A-T-19
A-T-8
A-T-13
A-T-7
A-T-10
A-V-29
A-V-17
A-V-6
A-V-7
A-V-4
125 130 135 140 145 150
K T L Q E L R V H E N E I T K V R K K P F N G L N Q M I V V
CB-6
A-V-7
CB-7
155 160 165 170 175 180
E L G T N P L K S S G I E N G A F Q G M K K L S Y I R I A D
CB-6
CB-1
CB-7
185 190 195 200 205 210
T N I T T I P Q G L P P S L T E L H L D G N K I T K V D A A
CB-1
214
S L K G
CB-1

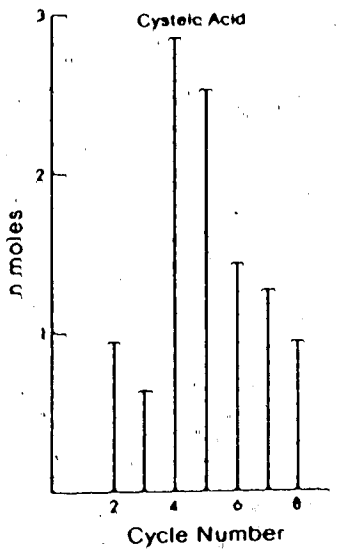


Fig. 10. Identification of residue 4 of the PDS(BES)core. The yield of PTH-cysteic acid produced by the sequencer is shown for residues 1-8.

galactosamine, to the same level as seen in the other two cores, and a considerable reduction in the glucosamine, to about 30% of that in PDS. This, and the considerably lower molecular weight value, are consistent with the expected removal of the short oligosaccharide side chains (a single glucosamine, N-glycosidically linked to asparagine, remains at the site of attachment of each of the N-linked oligosaccharides) in addition to the removal of the dermatan sulphate. Sequence analysis of this core identified the first eight residues (Table 3, Fig. 9) including serine at residue four (Fig. 11). Due to a considerable breakdown of serine derivatives which was encountered in early sequencing experiments, serine was identified by the recognition of both PTH-serine (at 254nm) and a degradation product of PTH-serine (at 313nm), which was tentatively identified as PTH-dehydroserine from its very high A_{313}/A_{254} ratio. The identification of serine at residue four confirms the identification made from the sequence of the (BES)core. The sequence serine-glycine at residues four and five matches the only known

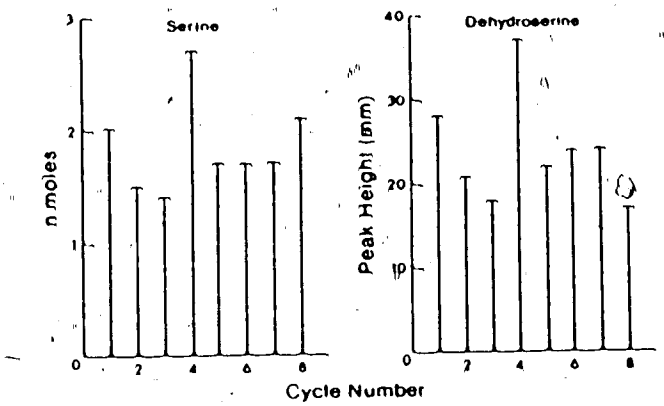


Fig. 11. Identification of residue 4 of the PDS(HF)core. The yields of PTH-serine and PTH-dehydroserine produced by the sequencer are shown for residues 1-8.

biosynthetic sequence requirement for the attachment of a chondroitin sulphate chain (Isemura, *et. al.*, 1981). Chondroitin sulphate is the biosynthetic precursor for dermatan sulphate. This strongly suggested, but did not prove, that the DS chain is attached at this position. Definitive proof of this was obtained by a different approach (see later).

3.2 Isolation and Characterisation of Peptide Fragments.

3.2.1 Cathepsin D Digestion Products of PDS

Cathepsin D, an enzyme that cleaves only a restricted number of bonds (Scott & Pearson, 1981), and which is available in this laboratory, was chosen in order to produce a small number of relatively large peptides from PDS. A cathepsin D digest of PDS was chromatographed on Polyanion SI (Fig. 12). All the retarded material was collected as a single pool referred to as CD-N, which probably included only those peptides which carried a dermatan sulphate chain. The work of Dr. R.K. Chopra, in this department, demonstrated

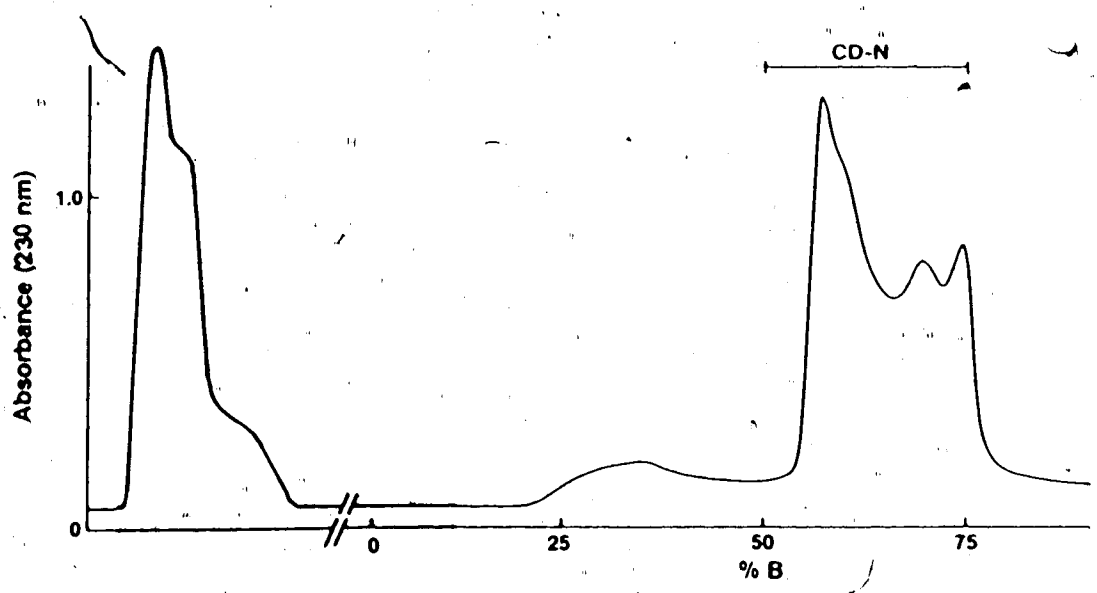


Fig. 12. Polyanion-SI ion exchange chromatography of a cathepsin D digest of PDS. Column, (HR 10/10) 1 x 10cm. Eluents, A: 7M urea, 0.15M sodium chloride, 0.05M Tris pH 6.5; B: 7M urea, 2.0M sodium chloride, 0.05M Tris pH 6.5. Flow rate 1.0ml/min.

that at least 75% of the bovine skin PDS molecules contain only one dermatan sulphate, which is attached to serine-4 (Chopra *et al.*, 1985). Consequently most of the material in the CD-N pool must include the amino-terminus of the intact molecule (Fig. 13).

CD-N is seen as a single diffuse band, on Toluidine Blue staining of SDS-PAGE gels, which yields several smaller bands on Coomassie Blue staining of a chondroitinase ABC digest of CD-N. As so many peptides, all including the amino terminus of PDS, are derived from the cathepsin D digest, it is apparent that incomplete cleavage occurred. A Western blot, from a 10% SDS-PAGE gel, of CD-N, after deglycosylation with chondroitinase ABC, showed (Amido Black staining) a series of peptides of apparent M_r 16,000-34,000, contaminated with a small amount of PDS(ABC)core (Plate 1). All bands stained with the monoclonal antibody 7B1. All but the smallest peptide, detected with Amido Black, stained with 5D1. The two

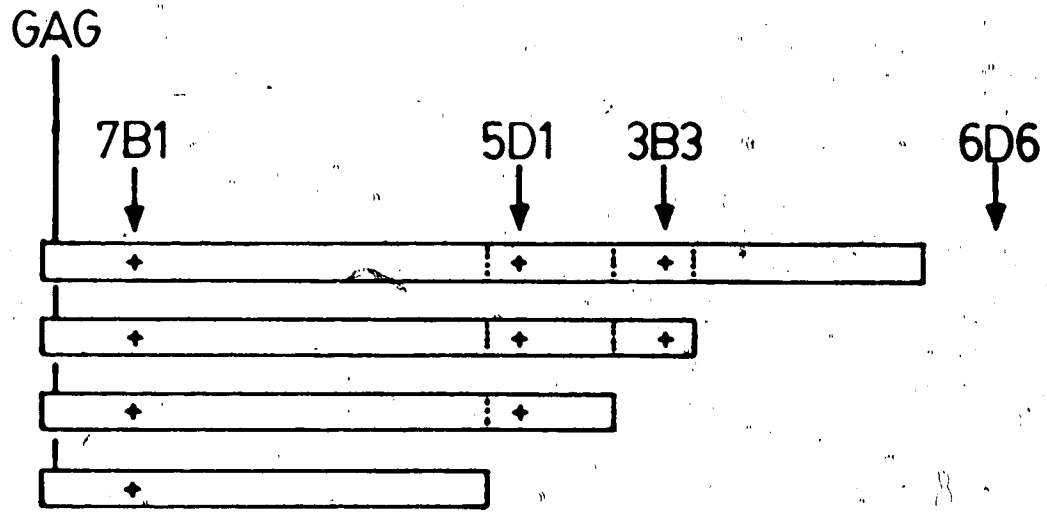


Fig. 13 A model of the structure of CD-N, the series of GAG bearing peptides isolated from a cathepsin-D digest of PDS by chromatography on polyanion-SI, an anion exchange column. The approximate sites of the epitopes recognised by the MAbs used during this project, are also shown (+).

smallest peptides did not stain with 3B3 and only the contaminating PDS(ABC)core stained with 6D6 (Plate 1). These data suggested that the order of the epitopes recognised by these MAbs, on the protein core of PDS, from the amino-terminus, is: 7B1, 5D1, 3B3, 6D6.

3.2.2 Cyanogen Bromide Cleavage Products of PDS(ABC)Core

A series of trial digestions using various samples (PDS, (ABC) and (HF)cores), pretreatments (reduction, reduction and alkylation) and digestion conditions, screened with 10% SDS-PAGE, was used to select the optimal CNBr digestion procedure. The method selected for routine use was the one using the most straightforward set of conditions, producing consistent results and leaving no undigested starting material.

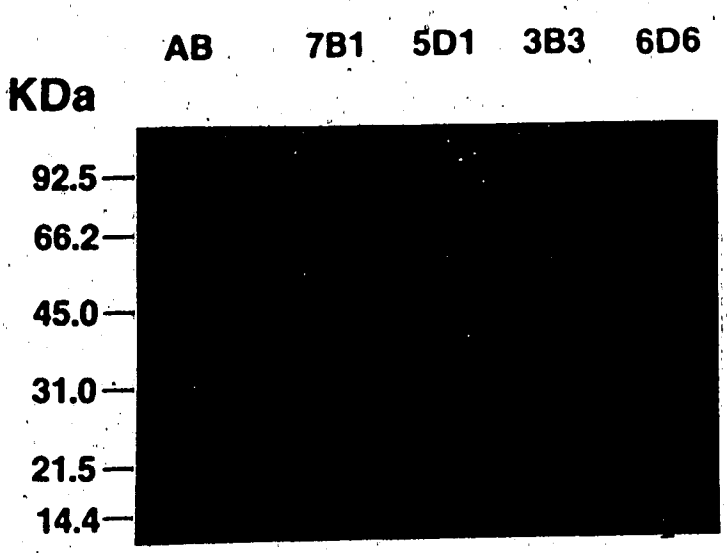


Plate 1. Immunoblot of a chondroitinase ABC digest of CD-N (CD-N is the pool of GAG-bearing peptides from a cathepsin D digest of PDS). The peptides were transferred from a 10% SDS-PAGE slab gel to nitrocellulose paper and stained with Amido Black (AB), or immunochemically using 1:100 dilutions of monoclonal antibody ascites fluid (as indicated). The Amido Black stained track was loaded with 20μg of CD-N and the MAb stained tracks with 1-4μg per track. The left hand track contains molecular weight standards.

Initial screening of the CNBr cleavage products of PDS(ABC)core, with 10% SDS-PAGE, showed 3 major bands of apparent M_r 35,000, 12,000 and 10,000 and several more weakly-stained bands (Fig. 14, Table 4). These bands were numbered in order of increasing mobility from the largest, CB-1 to the smallest CB-7. CB-1 and CB-2 were the only peptides which stained with periodic acid Schiff, suggesting that these were the only peptides with oligosaccharide attached. This method of periodic acid Schiff staining produced an intensity of staining for these peptides, equivalent to that of Coomassie Blue.

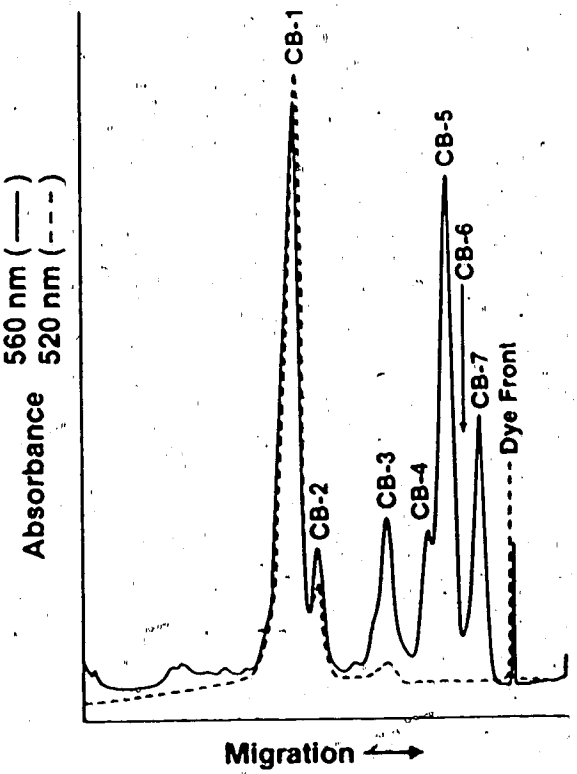


Fig. 14. SDS-PAGE of a CNBr digest of PDS(ABC)core. Aliquots (25μg) of the digest were run on separate tube gels of 10% acrylamide at pH 7.2. The gels were then stained, destained and scanned at 560nm (Coomassie Blue staining) or 520nm (Periodic acid Schiffs staining).

TABLE 4
CNBR PEPTIDES OF PDS(ABC)CORE

CNBr Peptide	CB-1	CB-2	CB-3	CB-4	CB-5	CB-6	CB-7	CB-8
Apparent M_r (x 10^3)	35	30	20	14	12	11	10	ND.
Proportion of Whole Digest[1]	37%	7%	10%	6%	25%	tr	12%	ND

[1]Based on relative peak area of a scan of an SDS-PAGE gel of a CNBr digest of PDS(ABC)Core (Fig. 14).
ND = Not detected.

3.2.2.1 Isolation of cyanogen bromide peptides

The following gel-filtration media were investigated, using several buffer systems: Fractogel TSK-HW-50-F, Sephadex G-50 and G-75, Superose 12, Sephacryl S-300 and S-400. Using 10% SDS-PAGE to monitor the separation of the peptides, Sephacryl S-300, eluted with 0.01% TFA, was selected as providing optimal separation (Fig. 15).

A small proportion of the CB-1 and CB-7 from the digest was isolated in a homogeneous state, as determined by SDS-PAGE, from S-300 chromatography. The true utility of the S-300 chromatography however, was in the separation of the peptides into groups which were further resolved at later stages.

In some S-300 runs CB-5 was successfully isolated from a mixture containing CB-7, CB-6 and CB-5, by rechromatography on the same S-300 column (Fig. 16). Often however, all the S-300 run fractions containing CB-5 also contained CB-4 and CB-3: Therefore only a small amount of CB-5 could be prepared in this manner.

The major pool of peptides which required further fractionation (Pool I) contained CB-1, CB-2, CB-3, CB-4 and CB-5. The following chromatographic media were investigated using a variety of eluent systems (TFA, HFBA, triethylamine), in search of an appropriate fractionation method: rechromatography on Sephacryl S-300, Altex Ultrapore RPSC (C_3), Vydac RP-214-TP(C_4), Pharmacia Pro-RPC (C_1/C_8) and Pep-RPC (C_3/C_{18}). Again 10% SDS-PAGE was used to monitor the resolution and Pro-RPC, eluted with an acetonitrile gradient in 0.1% HFBA, was adopted as the routine method because it afforded the most useful separation (Fig. 17). This separated S-300 pool I into two further groups of peptides, CB-group A consisting of CB-3, CB-4 and CB-5 and CB-group B consisting of CB-1 and CB-2. As the peptides CB-2, CB-3 and CB-4 were partial cleavage products, they were not always present in all digests and when present their amounts fluctuated. When samples run on this column also contained CB-6, small peaks (i and ii in Fig. 17) were observed eluting prior to the CB-group B peak.

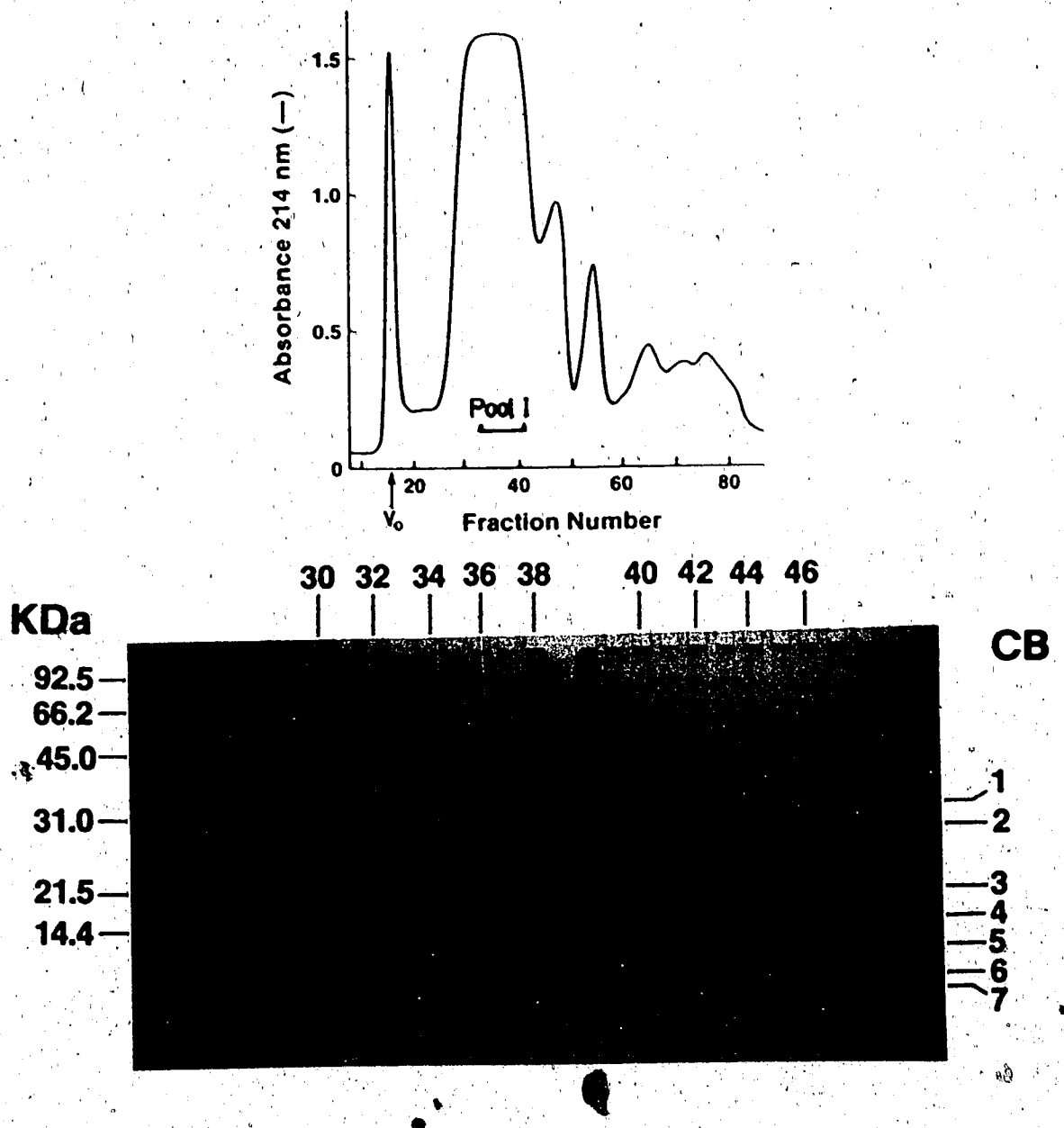


Fig. 15. Above: Sephacryl S-300 gel filtration chromatography of a CNBr digest of PDS(ABC)core. Column, 1.6 x 85cm. Eluent 0.01% TFA, flow rate 6ml/hr. Below: SDS-PAGE of aliquots of the fractions from the above S-300 run. Electrophoresis was at pH 7.2 in 10% acrylamide gels. The left hand track of each of the two slabs shown contains molecular weight standards as listed under section 2.2.4.2.

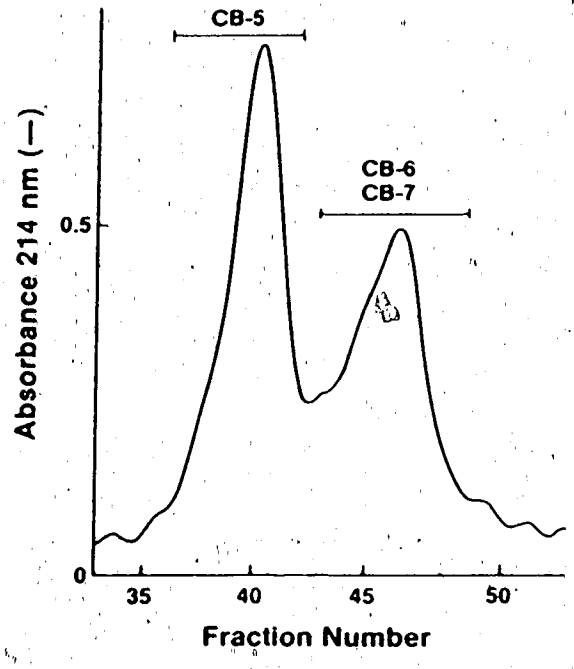


Fig. 16. Sephacryl S-300 gel filtration re-chromatography of a pool containing CB-5, CB-6 and CB-7 from a prior S-300 run of a CNBr digest of PDS(-ABC)core. • Column, 1.6 x 85cm. Eluent 0.01% TFA, flow rate 6ml/hr.

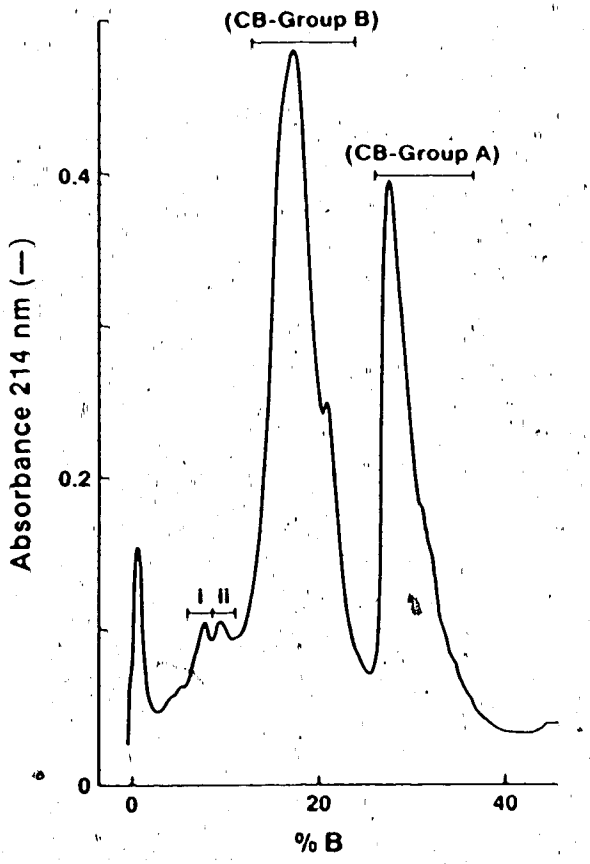


Fig. 17. Pro RPC (C_1/C_8) reverse-phase chromatography of fractions 38-40 from the S-300 chromatogram shown in Fig. 15. Column, HR 5/10 (0.5 x 10cm). Eluents A: 0.1% HFBA, 30% acetonitrile; B: 0.1% HFBA, 50% acetonitrile, flow rate 0.5ml/min.

Any pool containing the trace component CB-6 (*eg:* fractions 43-46 in Fig. 15, pools i and ii in Fig. 17) was run on Pro-RPC (Fig. 18) and produced small amounts of homogeneous CB-6.

The small peptide, CB-7 and another similarly sized peptide, named CB-8, which was not seen in gel-electrophoresis, were isolated from the later fractions of the initial S-300 run (Fig. 15) by chromatography on Vydac RP-201-TP (Fig. 19). Initially, the CB-7/CB-8 pool was first chromatographed on Sephadex G-25 then G-50, before

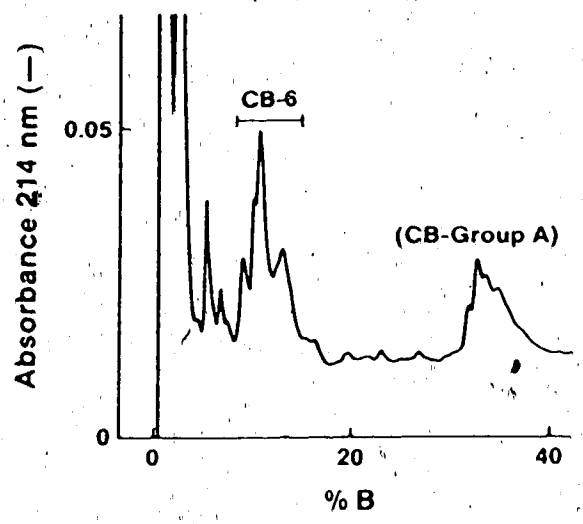


Fig. 18. Pro RPC (C_1/C_4) reverse-phase chromatography of fractions 43-46 from the S-300 chromatogram shown in Fig. 15. Column, HR 5/10 (0.5 x 10cm). Eluents A: 0.1% HFBA, 30% acetonitrile; B: 0.1% HFBA, 50% acetonitrile, flow rate 0.5ml/min.

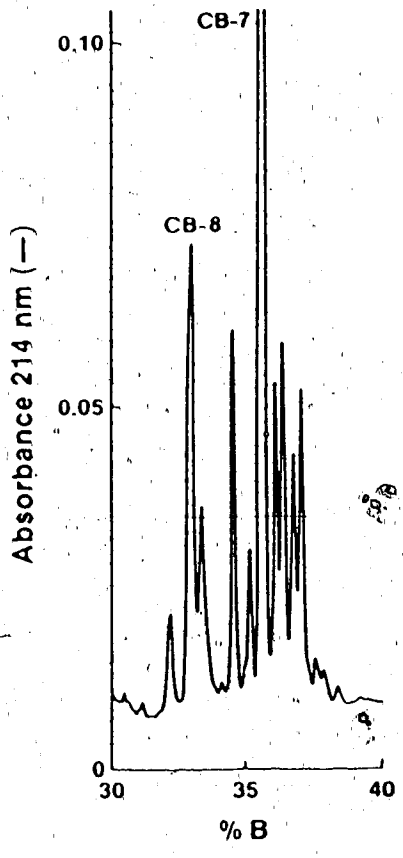


Fig. 19. Vydac RP-201-TP (C_{18}) reverse-phase chromatography of fractions 49-55 from the S-300 chromatogram shown in Fig.13. Column, 0.46 x 25cm. Eluents, A: 0.1% TFA; B: 0.1% TFA, 70% acetonitrile, flow rate 1.0ml/min.

RP-201-TP. These steps did not provide any useful resolution and were subsequently omitted. Unsuccessful attempts were made to observe CB-8 in gel-electrophoresis using shorter running times, precipitation with TCA during fixation, staining and destaining steps, and a highly cross-linked, urea-containing gel system (Swank and Munkres, 1971). Absorbance at 214nm was used routinely to monitor gel-filtration columns and small peptides such as CB-8 were then readily detected. CB-8, the amino-terminal peptide, was identified from its composition which was known in full from the sequence of

PDS(ABC)core.

3.2.2.2 Characterisation of cyanogen bromide peptides

The amino acid compositions of those peptides which were isolated in homogeneous form are shown in Table 5. Residues per mole values were calculated based on the molecular weights known from sequence data. The composition of CB-1 was expressed as residues per 1,000 because an accurate molecular weight value was not available. Sequence data derived from CNBr peptides is summarised in Fig. 9.

TABLE 5
AMINO ACID COMPOSITIONS OF CNBr PEPTIDES

Amino Acid	CB-1 R/1000	CB-5 R/mole	CB-6 R/mole	CB-7 R/mole	CB-8 R/mole
Asx	146	10.9	4.5	2.3	1.0
Thr	53	2.8	2.6	1.0	--
Ser	90	3.8	3.4	2.1	1.1
Hse	--	0.8	0.9	0.8	0.9
Glx	62	9.4	6.6	3.4	6.3
Pro	60	8.2	2.3	1.1	3.9
Gly	76	5.2	5.3	4.3	2.1
Ala	69	3.5	2.1	1.2	1.0
Val	66	5.1	4.1	1.6	1.0
Met	--	--	0.6	--	--
Ile	63	6.3	3.0	1.5	1.8
Leu	125	19.4	6.1	2.0	--
Tyr	47	1.2	0.4	--	--
Phe	27	3.5	2.0	0.9	0.9
Lys	66	12.7	4.2	1.0	0.1
His	27	1.9	1.1	--	0.8
Arg	25	3.3	2.2	--	--
Cys	ND	ND	ND	ND	ND
Total	1000	98	51	23	21
GlcN[1]	39[2]	--	--	0.1	0.1
GalN[1]	1[2]	--	--	--	0.4

[1] Hexosamines determined from total hydrolyses only. The values are therefore depressed by approximately 40% with respect to values determined from the milder hydrolytic conditions employed in the hexosamine hydrolysis.
[2] Hexosamines expressed as residues/1000 amino acid residues.
ND Not Determined.

The largest peptide in a CNBr digest of PDS(HF)core was M_r 23,000. All the other peptides in this digest were also seen in the CNBr digest of PDS(ABC)core. This suggests that CB-1 probably consists of a peptide of M_r 23,000 bearing sufficient oligosaccharide chains to cause it to migrate with an apparent M_r of 35,000. This was the only CNBr peptide which contained glucosamine in significant amounts (Table 5), supporting the suggestion, from periodic acid Schiff stained SDS-PAGE gels, that all the oligosaccharides are attached to this peptide. The amino acid composition of CB-1 (Table 5) showed that it contained no homoserine and thus included the carboxy-terminus of the protein. Digestion of CB-1, using carboxypeptidase Y, released Val, Tyr, Ala, Arg, Gly, Leu and Ser (Fig. 20). Although the results were not identical to those obtained for PDS, they were sufficiently similar to suggest that CB-1 and the intact molecule have the same carboxy-terminal sequence, and thus that CB-1 is indeed the carboxy-terminal CNBr peptide. Particularly long sequencer runs were possible with this peptide. The full data for one such run is presented in Table 6.

The M_r 12,000 estimate for CB-5 was considered to be reliable as the values estimated from SDS-PAGE on 10% gels and calculated from amino acid composition (Table 5), were in good agreement. Thirty of the first 39 residues of the sequence of this peptide were positively identified. Conspicuous failures of identification in the early region of this sequence occurred at residues 4, 8, 10 and 17 (Fig. 9).

Analysis of CB-6 provided sequence of the first 39 amino acid residues with a failure of identification at residue 35 (Fig. 9). These data were sufficient to determine that CB-6 contained a methionine (residue 28) which had not cleaved, and comprised a 28 residue peptide (named CB-6a) and CB-7. The amino acid composition of CB-6 (Table 5) was in good agreement with the composition of the sum of CB-6a (from sequence) and CB-7.

The M_r estimated for CB-7, in highly cross-linked/urea-containing SDS-PAGE, was 2,500. Twenty two continuous residues of sequence were determined for this peptide

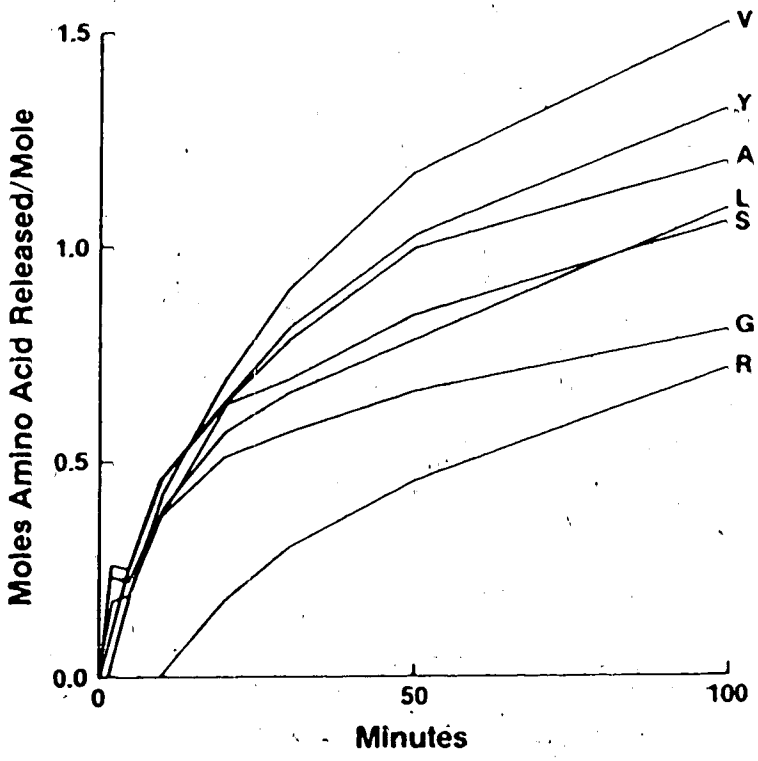


Fig. 20. Carboxy-terminal amino acid sequencing of peptide CB-1. The sample was incubated with carboxypeptidase Y for 24 hours. Aliquots were removed at 0, 2, 5, 10, 20, 30, 50, 100 and 300 minutes and at 24 hours, flash frozen and stored at -60°C until ready for amino acid analysis.

(Fig. 9). The amino acid composition suggested that a single homoserine was unaccounted for, obviously at the carboxy-terminal position. This composition corresponds to a molecular weight of 2,164, in good agreement with the estimate made by electrophoresis.

The peptide CB-8 was known to be 21 residues long from its amino acid composition and the sequence of the intact core. Fifteen of the first sixteen residues of sequence were identified for this peptide, with a failure of identification at residue 4 (Table 3, Fig. 9). The peptide was recognised as the amino-terminal CNBr peptide. The small amount of galactosamine in this peptide supports the attachment of a dermatan sulphate

TABLE 6
CB-1 SEQUENCE DATA

Sequencer Cycle	Residue Number	Residue Identified by HPLC	Yield %	Sequencer Cycle	Residue Number	Residue Identified by HPLC	Yield %
1	171	K	14	23	193	S	12
2	172	K	17	24	194	L	43
3	173	L	87	25	195	T	14
4	174	S	17	26	196	E	•
5	175	Y	75	27	197	L	27
6	176	I	72	28	198	H	8
7	177	R	34	29	199	L	26
8	178	I	67	30	200	D	•
9	179	A	76	31	201	G	8
10	180	D	•	32	202	N	11
11	181	T	33	33	203	K	2
12	182	-	--	34	204	I	10
13	183	I	55	35	205	T	3
14	184	T	24	36	206	K	2
15	185	T	31	37	207	V	8
16	186	I	48	38	208	D	•
17	187	P	44	39	209	A	7
18	188	Q	24	40	210	A	10
19	189	G	28	41	211	-	--
20	190	L	24	42	212	L	5
21	191	P	38	43	213	K	2
22	192	P	47	44	214	G	2

• Indicates a residue which was only determined qualitatively.

Repetitive Yield: S(4-23) 98%; T(11-25) 94%; I(6-16) 96%

chain at residue 4.

Sequence analysis of CB-groups A and B produced two sequences in each case. Group A yielded the sequences of CB-8 and CB-5 in the ratio of 1:3. Group B yielded the sequences of CB-7 and CB-1 in the ratio of 1:9. The disparity in amounts of the peptides in each group, and the fact that the minor sequence was known in full in each case, permitted clear distinction between the two. Eighteen residues of CB-8 and 17 residues of CB-7 were identified from the minor components of CB-groups A and B respectively.

3.2.2.3 Order of cyanogen bromide peptides

Sequence analysis of the intact PDS(ABC)core provided sufficient sequence to allow recognition of the first (21 residue) CNBr peptide and the early part of the sequence of the second. The amino acid composition and sequence of CB-8 were identical to those expected for the amino-terminal peptide. The amino acid sequence of CB-5 was recognised as that of the second CNBr peptide.

A western blot, from 10% SDS-PAGE, of a CNBr digest of PDS(ABC)core showed, on staining with amido black, all the peptides seen in the gel, stained with Coomassie Blue, shown in Fig. 14 (Plate 2). Any differences seen between these two digests are due to variations in proportion of the partial cleavage products, and the differences between direct Coomassie Blue staining of the gel and Amido Black staining of the nitrocellulose paper. Immunoblots of the same digest showed several peptides staining with each monoclonal antibody. This provided information regarding epitopes shared by several peptides, which, together with the sequence and compositional data presented above, permitted the ordering of the CNBr peptides (Fig. 21). Consequently, the epitope recognised by 7B1 is on CB-8, by 5D1 on CB-6a, by 3B3 on CB-7 and by 6D6 on CB-1. The partial cleavage products of the cyanogen bromide digestion were essential to the ordering of the sequence data derived from the discrete peptides.

3.2.3 Cyanogen Bromide Cleavage Products of Intact PDS

Other workers in this department (Chopra *et al.*, 1985) subjected a CNBr digest of PDS to chromatography on Polyanion SI. The retarded, uronic acid-containing material consists of the GAG-bearing CNBr peptides which are dominated by GAG-bearing CB-8. The amino acid composition of this preparation was very similar to to the composition of homogeneous CB-8 but the galactosamine content was substantially higher because the intact DS chain was still present. This, and the residual galactosamine seen from the DS linkage region of CB-8 from PDS(ABC)core, confirms the attachment of a DS chain to CB-8.

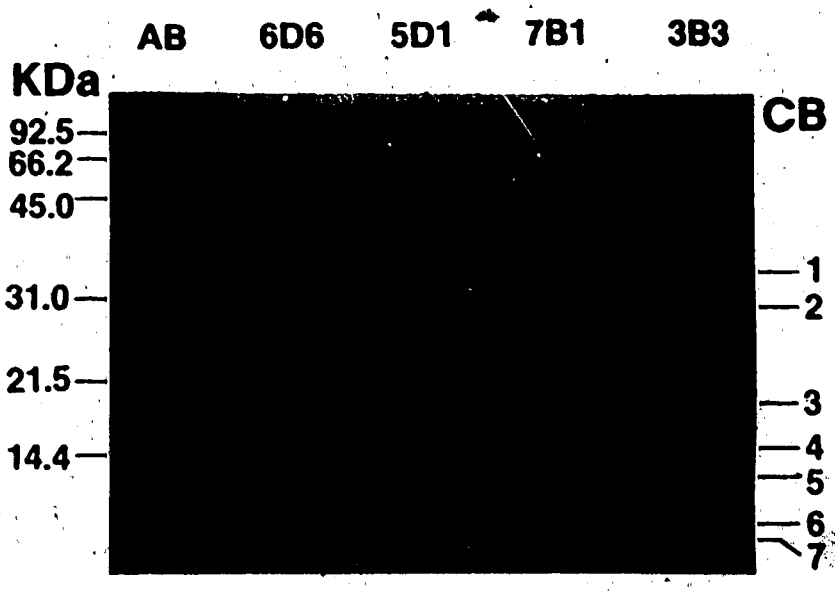


Plate 2. Immunoblot of a CNBr digest of PDS(ABC)core. The peptides were transferred from a 10% SDS-PAGE slab gel to nitrocellulose paper and stained with Amido Black (AB), or immunochemically using 1:100 dilutions of monoclonal antibody ascites fluid (as indicated). The left hand track contains molecular weight standards.

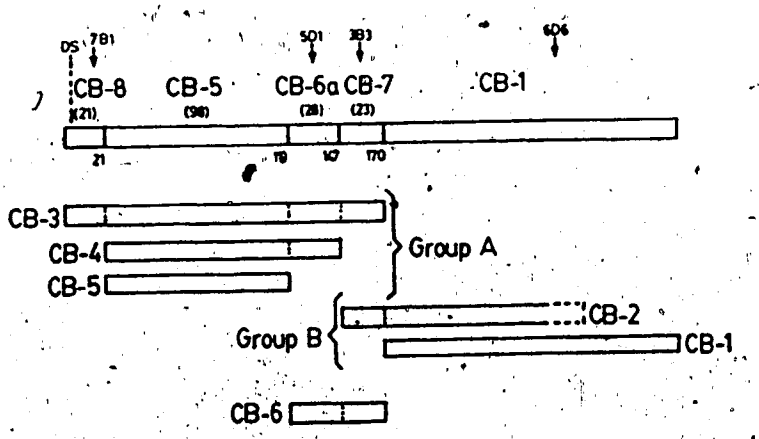


Fig. 21. Model showing the arrangement of the CNBr peptides derived from PDS(ABC)core. The numbers adjacent to the CNBr cleavage sites are the residue numbers of the cleavage methionines. The dotted lines indicate methionine sites which did not cleave. The numbers in parentheses are the lengths of the peptides in amino acid residues. Peptide CB-6a was never isolated, its sequence was obtained from CB-6. The approximate positions of the epitopes recognised by the MAbs used during this project, are shown.

3.2.4 Tryptic Fragments of CB-group A

The tryptic digest of CB-group A was chromatographed on a Vydac RP-201-TP (C_{18} reverse phase HPLC column (Fig. 22). The peptides from this digest were designated A-T-n where n is the number of the peak eluting from the chromatogram shown in Fig. 22. The material from each of the peaks was subjected to a two-dimensional separation comprising thin-layer electrophoresis followed by thin-layer chromatography. If this test produced a single spot, suggestive of homogeneity, or a pattern where one spot clearly dominated, sequencing of the peptide was undertaken. Further separation steps were not found necessary as all the components proved to be suitable for sequencing, according to the above criteria. The chromatographic yield relative to the amount of material which was digested with trypsin, amino acid composition, and amino acid sequence of each of these peptides is shown in Table 7.

Peptides A-T-4, 5, 6 and 8 were sequenced manually. As this is a qualitative method, no repetitive yield values can be reported.

Peptide A-T-16 had an amino acid composition corresponding to residues 22-45 (known from the sequence of CB-5). Previous sequence analyses of this region of the molecule had failed to identify the amino acids at residues 25, 29, 31 and 38. As this peptide contained all of these residues it seemed to be an ideal sample for the elucidation of these sites. Reduction and alkylation of the peptide, followed by chromatography on a Vydac RP-201-TP C_{18} column, yielded three peptides (A-T-16a, 16b & 16c), suggesting that these had been linked together by disulphide bridges. The chromatographic yield, amino acid composition and sequence of these peptides are shown in Table 7. On HPLC analysis of the sequencer fractions from these peptides, the previously unidentified residues all produced PTH-methylcarboxymethylcysteine. This suggests that these residues exist as cysteines in the native molecule and that the three peptides are linked together by two disulphide bridges, from residue 25 to residue 29 or 31 and from residue 38 to residue 31 or 29 (Fig. 23).

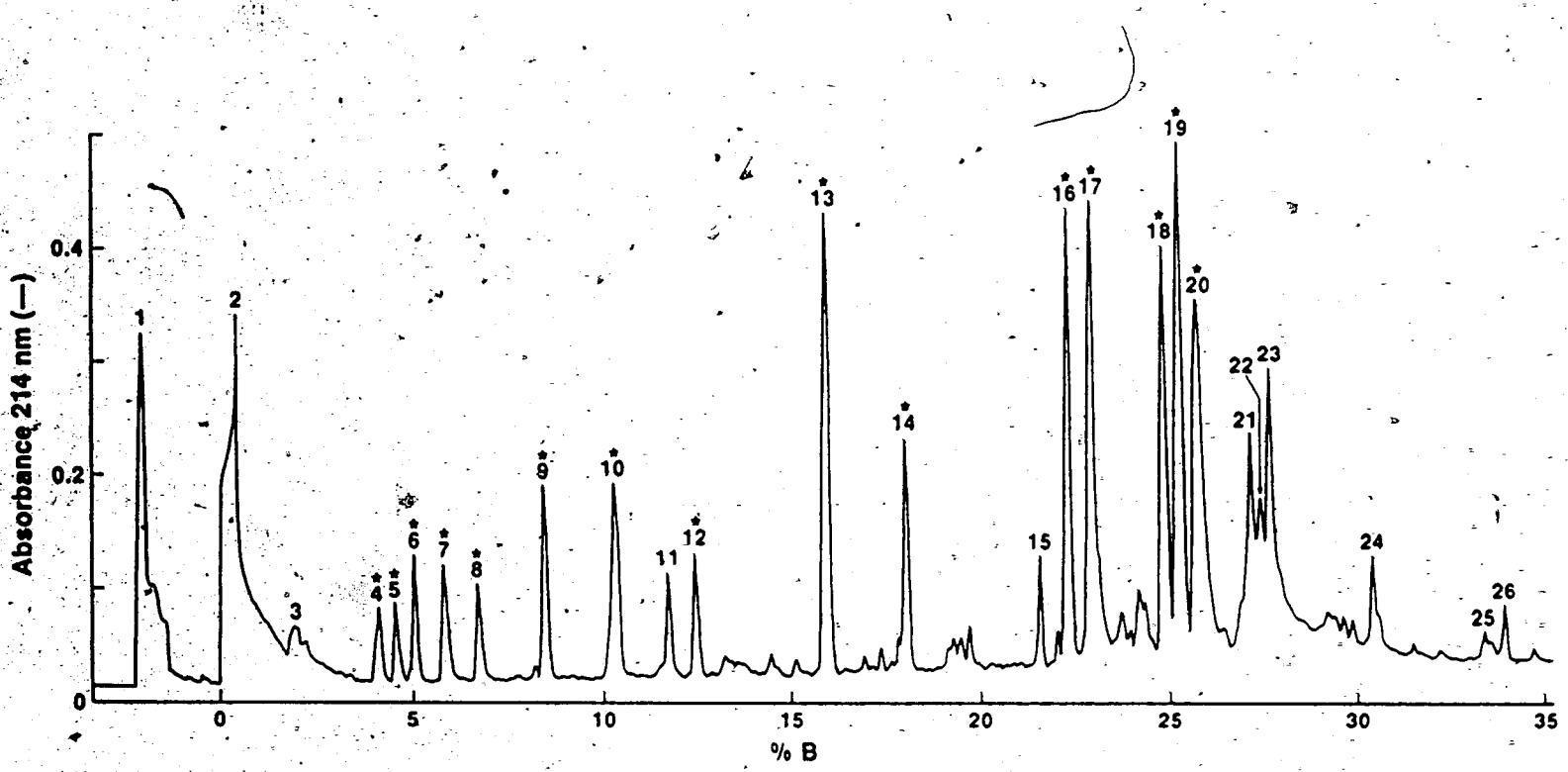


Fig. 22. RP-201-TP (C_{18}) reverse-phase chromatography of a tryptic digest of CB-group A. Column, 0.46 x 25cm. Eluents, A: 0.1% TFA; B: 0.1% TFA, 70% acetonitrile. Flow rate 1.0ml/min. Those peaks indicated by an asterisk were were also found in a tryptic digest of CB-2.

TABLE 7
TRYPTIC PEPTIDES OF CNBr GROUP A

Peptide	Yield %	Y_R	Residue Number	Amino Acid Composition and Sequence Data
A-T-4	36.1		74-76	N L (K) 1.1 0.8 1.2
A-T-5	37.1		88-90	I S (K) 1.0 0.9 1.1
A-T-6	24.7		46-48	V P (K) 1.1 0.8 1.1
A-T-7	37.1	63	110-113	N Q L K 0.9 1.1 0.8 1.2
A-T-8	39.2		102-104	L E (R) 0.8 1.0 1.2
A-T-9	22.7		69-73	(D, G, F, K) 1.8 0.9 1.2 1.1
A-T-10	36.1	79	114-118	E L P E K 1.1 1.2 0.8 1.1 1.1
A-T-10a	36.1		119	Hse 0.5
A-T-12	24.7	69	64-68	I T E I K 1.0 0.9 1.1 1.0 1.0
A-T-13	44.3		105-109	(L, Y, S, K) 1.6 1.3 0.9 1.2
A-T-14	17.5	88	64-73	I T E I K D G D F K 1.0 1.0 1.1 1.0 1.0 0.9 1.0 0.9 1.1 1.0
A-T-16a	12.4	90	29-34	Cmc Q Cmc H L R 1.0 1.0 1.0 1.0 1.0 1.0
A-T-16b	12.4	92	22-28	G P V Cmc P F R 1.1 1.0 1.2 0.7 1.0 1.0 0.9
A-T-16c	12.4	80	35-45	V V Q Cmc S D L G L E 0.8 0.8 1.1 1.2 1.0 1.0 1.1 1.0 1.1 1.1 K 1.1
A-T-18	24.7	81	77-87	N L H T L I L I N N 0.9 0.9 1.2 1.1 0.9 1.1 0.9 1.1 0.9 0.9 K 1.1

Peptide	Yield %	Y_R	Residue Number	Amino Acid Composition and Sequence Data
A-T-19	35.1	81	91-101	I S P G A / F A P L V 1.1 0.9 0.8 0.9 1.1 / 1.1 1.1 0.8 1.2 1.1 K 1.0
A-T-20	37.1	93	49-63	D L P P D T A L L D 1.2 0.8 0.9 0.9 1.2 1.0 1.1 0.8 0.8 1.2 L Q N N K 0.8 1.2 1.2 1.2 1.0

3.2.5 Staphylococcus aureus V8 Protease Cleavage Products of CB-Group A.

The V8 Protease digest of CB-group A was chromatographed on a Vydac RP-201-TP (C_{18}) reverse phase HPLC column (Fig. 24). The peptides from this digest were designated A-V-n, where n is the number of the peak eluting from the chromatogram shown in Fig. 24. The material from each of the peaks was subjected to thin-layer electrophoresis/ thin-layer chromatography and, as with the tryptic peptides of CB-group A, all were found to be suitable for sequencing without further purification. The chromatographic yield, amino acid composition, and amino acid sequence of each of these peptides is shown in Table 8. A-V-6 was sequenced by the manual method, therefore repetitive yield cannot be calculated. The carboxyl-terminal residue of this peptide, identified as "Glx" from the amino acid composition

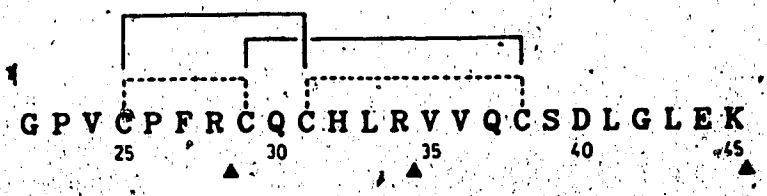


Fig. 23. Residues 22-45 of the sequence of PDS(ABC)core showing the two possible arrangements of the disulphide bridges (solid lines or broken lines). Closed triangular markers indicate tryptic cleavage sites.

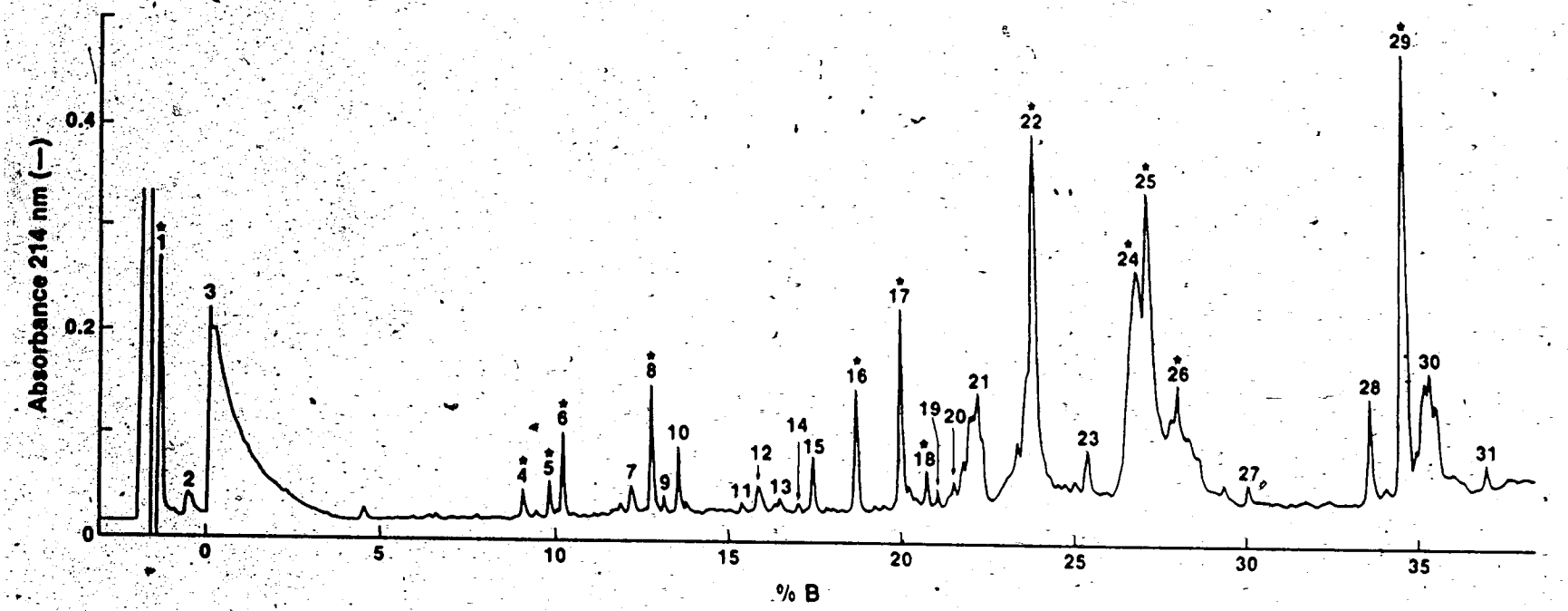


Fig. 24. RP-201-TP (C_{18}) reverse-phase chromatography of a *staph. aureus* V8 protease digest of CB-group A. Column, 0.46 x 25cm. Eluents, A: 0.1% TFA; B: 0.1% TFA, 70% acetonitrile. Flow rate 1.0ml/min. Those peaks indicated by an asterisk were were also found in a *staph. aureus* V8 protease digest of CB-2.

TABLE 8
STAPHYLOCOCCUS AUREUS V8 PROTEASE PEPTIDES OF CNBr GROUP A

Peptide	Yield %	Y_R	Residue Number	Amino Acid Composition and Sequence Data
A-V-4	[redacted]	82	115-119	L P E K (Hse) 1.1 0.8 1.3 1.0 0.9
A-V-6	14.4		115-117	L P (E) 1.2 0.7 1.3
A-V-7	4.6	79	118-125	K M P K T L Q E 0.9 0.6 1.2 0.9 0.8 1.1 1.1 1.1
A-V-8	18.6	93	59-66	L Q N N K I T E 1.0 1.1 1.0 1.0 1.0 0.9 0.9 1.1
A-V-17	15.5	88	104-114	R L Y L S K N Q L K 1.0 1.0 1.0 1.0 0.9 1.0 1.1 1.1 1.0 1.0 E 1.1
A-V-24	20.6	86	22-44	G P V X P F R X Q X 1.0 0.9 0.7 -- 0.9 1.0 0.8 -- 1.0 -- H L R V V Q X X D L 0.7 1.1 0.8 0.7 0.7 1.0 -- -- 1.5 1.1 (S, G, L, E) 0.4 1.0 1.1 1.0
A-V-29	22.7	83	67-103	I K D G D F K N L K 1.0 1.0 0.8 1.0 0.8 1.1 1.0 0.8 1.1 1.0 N L H T L I L I N N 0.8 1.1 1.1 1.0 1.1 1.0 1.1 1.0 0.8 0.8 K I X K I X P G X F 1.0 1.0 -- 1.0 1.0 -- 0.6 1.0 -- 1.1 A P L (S, A, V, K, L, E) 1.2 0.6 1.1 1.8 1.2 1.2 1.0 1.1 1.2

can be designated as glutamic acid, not glutamine, because it is a V8 cleavage site.

The overlapping peptides from tryptic and V8 digests of CB-group A provided sufficient data for the assembly of the sequence of CB-5, and to identify its carboxyl-terminal attachment to CB-6 (See Fig. 9).

3.2.6 Tryptic Fragments of CB-Group B

The tryptic digest of CB-Group B was chromatographed on a Vydac RP-201-TP (C_{18})reverse phase HPLC column (Fig. 25). Amino acid analyses of the peptides have been undertaken but sequence data is not yet available. From these amino acid analyses, several peptides from this digest have been recognised within the region of residues 148-210 (that region of CB-group B, CB-1 and CB-7, which has already been sequenced).

3.2.7 Staphylococcus aureus V8 Protease Cleavage Products of CB-Group B.

The V8 protease digest of CB-group B was chromatographed on Vydac RP-201-TP (C_{18})reverse phase HPLC column (Fig. 26). Amino acid analyses of the peptides have been undertaken but sequence data is not yet available. From these amino acid analyses, several peptides from this digest have been recognised within the region of residues 148-210 (that region of CB-group B, CB-1 and CB-7, which has already been sequenced). Significant among these is B-V-13, which appears to correspond to residues 181-208. Peptide B-V-13 contains one more aspartic acid or asparagine (Asx) residue than was found during the sequencing of the corresponding region of CB-1, and two or three glucosamines per mole (Table 9). This suggests that residue 182 is a glycosylated asparagine. Sequencing of CB-1 failed to identify residue 182, as would be the case were it glycosylated. The known sequence requirement for the biosynthetic attachment of an N-linked oligosaccharide, is Asn-Xxx-Ser/Thr (Marshall, 1972). If residue 182 is a glycosylated Asparagine then 182-184 (Asn-Ile-Thr) fits this pattern.

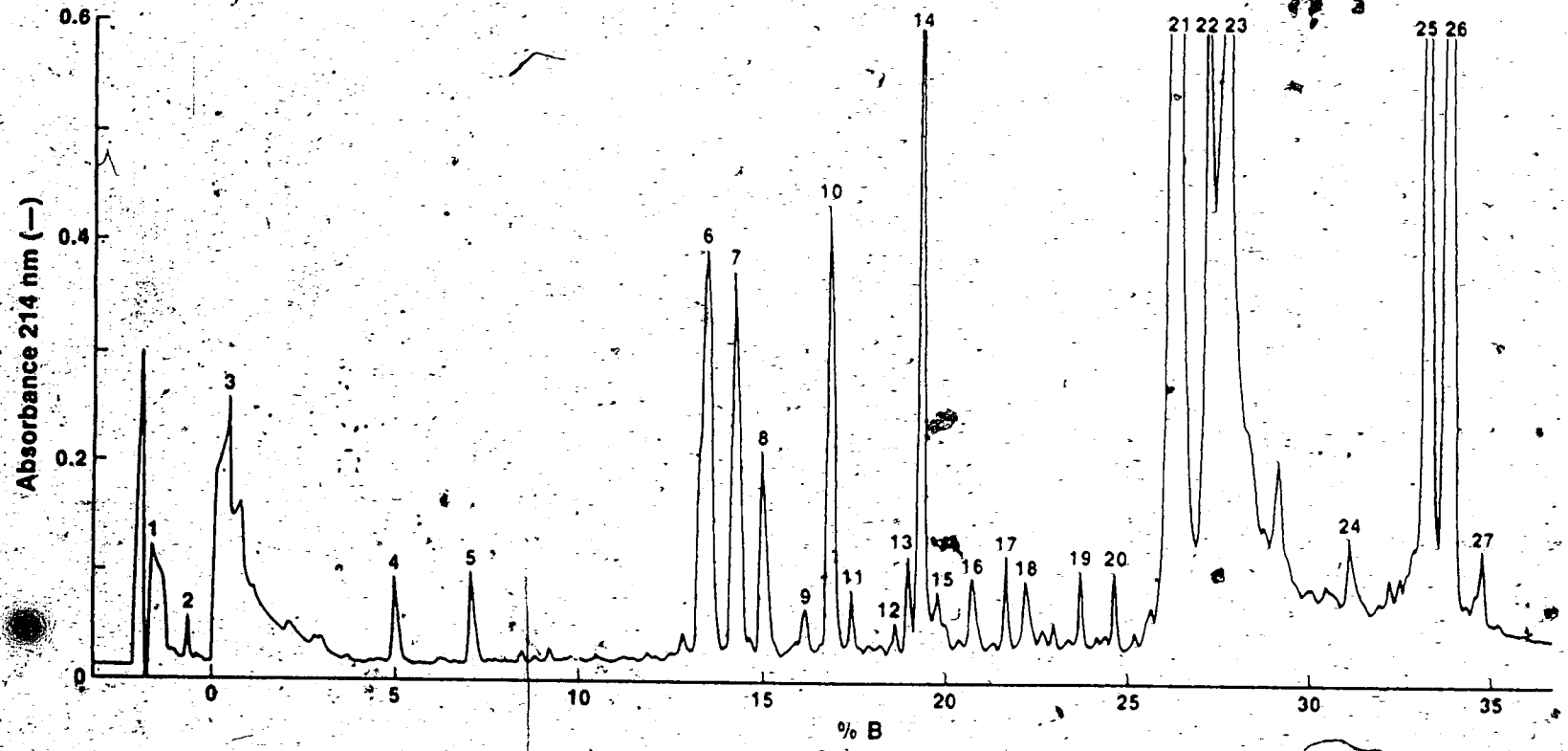


Fig. 25. RP-201-TP (C_{18}) reverse-phase chromatography of a tryptic digest of CB-group B. Column, 0.46 x 25cm. Eluents, A: 0.1% TFA; B: 0.1% TFA, 70% acetonitrile. Flow rate 1.0ml/min.

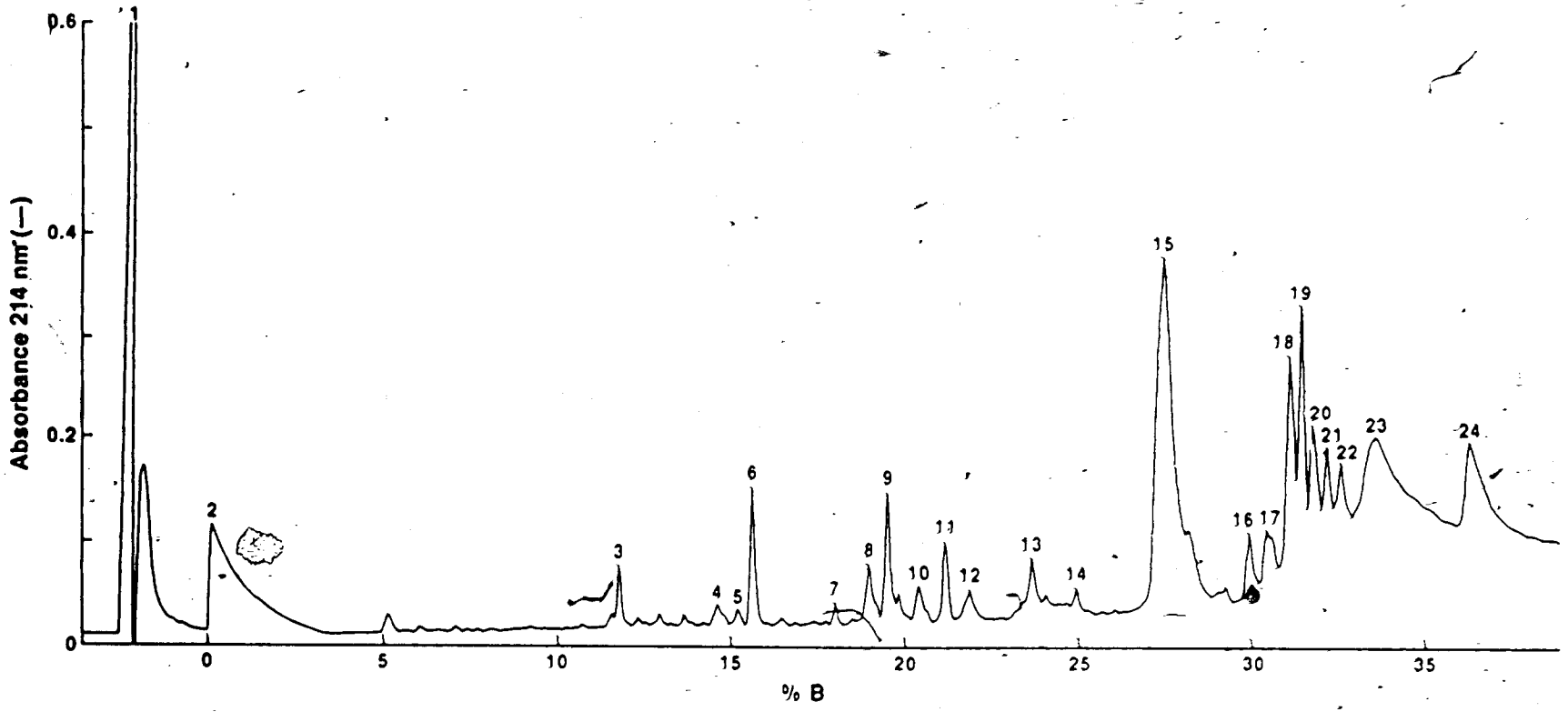


Fig. 26. RP-201-TP (C_{11}) reverse-phase chromatography of a *staph. aureus* V8 protease digest of CB-group B. Column, 0.46 x 25cm. Eluents, A: 0.1% TFA; B: 0.1% TFA, 70% acetonitrile. Flow rate 1.0ml/min.

3.3 Cathepsin C Digestion of Proteodermatan Sulphate

The variety of evidence suggesting the attachment of a DS chain at serine-4 led Dr. R.K. Chopra, working in this department, to study the degradation of PDS with the exo-dipeptidase cathepsin C, which cleaves dipeptides from the amino terminus of protein substrates. Dr. Chopra showed, in densitometric scans of SDS-PAGE gels, that the progress of the digestion was marked by a decrease in the gel band corresponding to PDS (Coomassie Blue and Toluidine Blue staining) and an increase in bands of a size comparable to the PDS(ABC)core (Coomassie Blue staining) and free DS (Toluidine Blue staining). The release of dipeptides was also monitored, in an amino acid analyser, calibrated with the expected dipeptides. Asp-Glu was released rapidly. Gly-Ile and Gly-Pro were released only very slowly. The second, GAG-bearing dipeptide, which is the Toluidine Blue staining component

TABLE 9
COMPOSITION OF PEPTIDE B-V-13

Amino acid	Residues per mole
Asx	4.6
Thr	3.1
Ser	1.2
Glx	2.0
Pro	2.9
Gly	2.3
Ala	1.3
Val	1.7
Ile	1.9
Leu	4.6
Lys	1.8
His	0.7
GlcN[1]	2.1

[1] Hexosamines determined from total hydrolyses only. The values are therefore depressed by approximately 40% with respect to values determined from the milder hydrolytic conditions employed in the hexosamine hydrolysis.

observed in SDS-PAGE, was isolated by cellulose acetate electrophoresis prior to hydrolysis and amino acid analysis which confirmed its composition. Thus, a limited proteolysis of the amino-terminal region of the protein core had removed the DS, confirming that it was located on serine-4 (Chopra, *et. al.*, 1985).

4. DISCUSSION

The aim of this project was to determine the amino acid sequence of the protein component of PDS. Initial attempts at manual sequencing (Methods of Edman & Henschen, 1975, and Tarr, 1977) were unsuccessful, in part because of the nature of the sample and in part because of lack of experience in sequencing techniques. When the emphasis was switched from manual to automated sequencing, it was soon discovered that the intact PDS was unsuitable as a sample for sequencing. The sample problems were overcome and with time, sequencing efficiency was increased.

4.1 Sequencing Efficiency

Thorough studies of sequencer efficiency under various conditions were outside the scope of this project; however, a number of modifications of the instrument or operating conditions, which the literature indicated would provide improved efficiency, were attempted.

Initially, identification of the released PTH-amino acids by HPLC was hampered by a very large peak which co-eluted with several of the PTH-amino acids. It was surmised that this peak was Quadrol. Technical bulletins from the manufacturer of the sequencer, suggested the reduction of the molarity of the Quadrol used from 1.0M to 0.25M and the inclusion of acetic acid in the ethyl acetate (1ml of glacial acetic acid per litre), to aid in the neutralisation of the small amounts of this buffer that may be retained by the sample. The Quadrol was eventually decreased to 0.1M, the acetic acid was incorporated into the ethyl acetate and the interfering peak was reduced to an acceptable level.

The use of the poly(quaternary amine) carrier, Polybrene, was adopted in order to reduce extractive loss of sample, which is a problem especially during the wash with ethyl acetate (Tarr *et al.*, 1978). In order to remove any contaminants of the Polybrene that might be extracted with the PTH-amino acid, it was precycled in the sequencer cup for six cycles with di-glycine (the di-glycine was subsequently omitted).

Proline residues have proven difficult to cleave, (Tarr, 1977) and have been identified as the sites at which sequence determination ceases (probably due to cleavage failure) in several of the experiments run early in this project. A programme was written that provided a second cleavage at cycles known, from preliminary runs, to be proline. This allowed the determination of more residues of sequence beyond the proline in question.

Major changes to the instrument included the rearrangement of the vacuum system, which provided a dedicated vacuum supply for the reaction cell, and supplied this system with a cold-trap to reduce the amount of material in the vapour phase, thus improving the vacuum and protecting the pump from corrosion. The remainder of the vacuum system is used to operate several valves and to provide vacuum for the fraction collector and is thus less critical to the operation of the instrument's chemistry.

The instrument modification that provided the greatest apparent improvement in sequencer efficiency was the installation of an automated conversion module. The sequencer formerly produced ATZ-derivatives of the cleaved amino-terminal amino acids. These were taken from the fraction collector and converted manually. Unfortunately, the ATZ-amino acids are quite unstable (Margolies *et al.*, 1982) and any such derivatives produced by the sequencer during an over-night run had to wait until morning before conversion to the more stable PTH-amino acids. The auto-converter eliminates the break-down of the ATZ-amino acids at this stage by undertaking the conversion immediately the ATZ-amino acid is extracted from the spinning-cup. The more stable PTH-amino acid is then deposited in the fraction collector to await HPLC identification. The methanolic HCl conversion used by the auto-converter was also found to increase the yield of serine, threonine, asparagine and glutamine phenylthiohydantoins.

An electronic programmer for the sequencer was installed at the same time as the auto-converter (the auto-converter has its own programmer), replacing the original electromechanical programmer. This allowed the automatic changing of programmes during a run to accommodate the differential treatment of certain residues such as the use of a second

cleavage step for prolines.

In an attempt to improve the amount of sequence data obtainable from a single run, o-phthalaldehyde was used at cycles where a proline residue is the exposed amino-terminus, to suppress the PTH-amino acid background caused by the sequential degradation of the amino-termini of peptides created by random cleavage of the sample peptide (Brauer *et al.*, 1984). Though a considerable suppression of the background was achieved with each use of OPA, a decrease in yield was also observed and no more sequence data were obtained than in a corresponding run in which OPA was not used. The combined modifications to the instrument and methodology provided sufficient improvement of sequencing efficiency to allow the reduction of sample load from about 100nmoles to 5nmoles.

4.1.1 Manual Sequencing

Late in the project, a manual sequencing method was used in the analysis of the tripeptides derived from CB-group A by digestion with trypsin or V8 protease. This method (Chang, 1983) was radically different from those attempted early in the project and its success provides no explanation for the failure of the early attempts. Factors that probably featured prominently in the success of the later attempts were the routine use of commercially available, highly-purified reagents (sequanal grade) and the use of a specific end-labelling technique through to the identification stage, rather than back-hydrolysis followed by amino acid analysis, as used in the earlier attempts. Back-hydrolysis removes the specific label and one must therefore consider the general amino acid background and maintain appropriately clean technique throughout. It was found that the precautions suggested in the original method with regard to the retention of an oxygen-free atmosphere during the reaction stages of the manual sequencing protocol were not necessary for such short runs. Reaction tubes were merely sealed under Parafilm® after purging with nitrogen.

The identification of the released DABTH-amino acids by thin layer chromatography was straightforward (see Fig. 6). As double-sided plates were used, it was possible to run a

standard on the back of the unknown to aid in identification. The position of the unknown spot relative to the standard on the rear was established by piercing the plate with a pin. The consistency of position on the front and rear of the plates was compared by running an internal standard on both sides. In cases where a residue was one of the two pairs of DABTH-amino acids that were not resolved in this TLC system, the amino acid composition of the peptide was relied upon for the distinction.

4.1.2 HPLC Identification of PTH-Amino Acids

Several HPLC methods were employed for the identification of the PTH-amino acids. The separation of all the PTH-amino acids in a reasonable period of time is quite difficult, and in order to save time, commercial kits comprising a column and a factory optimised elution protocol were occasionally purchased, if it was likely that they would provide an improvement over the method in routine use at that time. Until the installation of the automatic converter, the separation of the PTH-amino acids was somewhat easier, as the manual conversion method separated the three water-soluble PTH-amino acids from the rest.

The first separation method used was that of Zimmerman *et al.*. (1977), using a C_{18} reverse-phase column eluted with a gradient of acetonitrile in 10mM sodium acetate, pH 4.5. The elution of all the PTH-amino acids required 45 minutes between injections.

The next method used was a Beckman kit based on the method of Somack (1980). Using a C_{18} column eluted with an acetonitrile gradient in 4.2mM sodium acetate containing tetrahydrofuran. This method provided better resolution of the PTH-amino acids; however, 45 minutes was still required between injections. Difficulty was experienced in acquiring tetrahydrofuran of sufficient optical transparency to avoid a steeply rising baseline. The method of Tarr (1981), again on C_{18}, was a radical improvement. The method is isocratic and therefore has a flat baseline. The resolution was greatly improved and the run time was reduced to only 18 minutes. Eluent preparation time was also reduced, and buffer composition rendered more consistent, by the use of a stock buffer. Two runs per cycle were still required

whilst the manual conversion was performed as this method separated the derivatives into aqueous and organic phases. Serine and threonine were identified as peaks that absorbed at 313nm. These peaks, presumed to be breakdown products of PTH-serine and PTH-threonine were also found in PTH-serine and PTH-threonine standards that had been allowed to stand overnight at room temperature.

When the automatic conversion module was brought into use, a separation method that would separate all the PTH-amino acids in a single run was desirable, because phase separation of the derivatives no longer occurred. A kit was purchased from Dupont (Glajch *et al.*, 1985) which, in an 18 minute isocratic run on a cyano-column, separated all the PTH-amino acids (see Fig. 5). Two buffers were made up (different ionic strengths one for each pump) and the proportion of each in the eluent, required to optimise the separation, was determined each day. The auto-converter, which uses a methanolic HCl conversion acid, preserved PTH-serine and PTH-threonine far better than the manual conversion method which used aqueous HCl. This alleviated the need for the second detector to identify serine and threonine from their breakdown products. Methanolic HCl conversion produces methyl derivatives of PTH-aspartic acid, PTH-glutamic acid and PTH-carboxymethylcysteine. No commercial standards were available for these derivatives; however, standardisation of elution position, though not of quantity, was possible when methyl derivatives of these amino acids were prepared on the sequencer and run on the PTH-amino acid HPLC system.

4.2 Characterisation of PDS and the Protein Cores Derived from PDS.

One of the principal screening methods used throughout this project was SDS-PAGE. This is a powerful technique capable of resolving components differing by as little as 2KDa in molecular weight. Unfortunately, glycosylated proteins do not behave ideally in this system (Segrest & Jackson, 1972). Intact proteoglycans and even some of the core preparations, although well resolved, migrate anomalously causing over-estimation of their molecular weight. The best available estimate of the molecular weight of the protein component of PDS is 42,000,

obtained for the PDS(HF)core in 5% SDS-PAGE. Anhydrous HF was the most thorough deglycosylation method used, (this was supported by the hexosamine analyses) and therefore produced the species exhibiting least abnormality in SDS-PAGE, permitting the determination of M_r using only a single gel concentration (Weber & Osborn, 1975). The PDS(ABC)core and PDS(BES)core both migrated with apparent M_r of 53,000 in 5% SDS-PAGE. Though these species have more residual carbohydrate than the PDS(HF)core, this value is considered to be an over-estimate because of the value of only 45,000 determined for PDS(ABC)core in a Ferguson plot. Nakamura *et al.* (1983) suggested the presence of three O-linked oligosaccharides (sialic acid - galactose - N-acetylgalactosamine) and some N-linked oligosaccharide(s) in a proteodermatan sulphate from calf skin. Glössl *et al.* (1984) reported two or three N-linked oligosaccharides and no O-linked oligosaccharides in a PDS from cultured human fibroblasts. Evidence from this project shows no galactosamine other than at the site of attachment of the dermatan sulphate chain, although there is a considerable amount of glucosamine associated with several of the tryptic and V8 peptides from the carboxyterminal region of the molecule. This does not support the structures proposed by Nakamura *et al.* but suggests similarity with the molecule studied by Glössl *et al.*. Neutral sugar analyses of this molecule are also compatible with the Glössl model in that the galactose/mannose ratio of 1:1 better supports typical N-linked oligosaccharide composition.

The single glycosaminoglycan chain of PDS has a molecular weight of 16KDa (Pearson & Gibson, 1982). This in addition to the 42KDa core and a contribution of 4-7KDa from oligosaccharide side chains would give a molecular weight of 60-65 KDa for the intact PDS. The M_r determined, from SDS-PAGE on 5% gels, for intact PDS averaged 87,000, the over-estimation being severe due to the high carbohydrate content of the molecule.

Amino acid sequencing of PDS produced only a few residues of sequence. The acid stages of the sequencing procedure caused a charring of the carbohydrate components of the proteoglycan which resulted in breakdown of the thin sample layer, sufficient to prevent the effective interaction of the sequencer reagents and extraction solvents with the sample. The

deglycosylation of PDS eliminated the problem of disruption of the sample film in the sequencer reaction cup. The initial deglycosylation method using chondroitinase ABC was straightforward and produced, in good yield, a core which was judged to be homogeneous by SDS-PAGE. The use of an ultrafiltration device instead of dialysis, to remove the enzymatic digestion products of the glycosaminoglycan (largely disaccharides), was undertaken in order to minimise the membrane surface available for non-specific binding of the protein core and to reduce the time required for this step. The procedure was successful, as shown by hexosamine analyses.

The failure of attempts to identify residue 4 in sequence analyses of the PDS(ABC)core suggested that this residue may have been a cysteine involved in a disulphide bridge. However, when a sample of reduced and alkylated PDS(ABC)core was sequenced no identification could be made at residue 4.

The posibility that residue 4 was a GAG bearing serine was suggested by the presence of glycine at residue 5. The only known biosynthetic requirement for the attachment of a chondroitin sulphate (the biosynthetic precursor of dermatan sulphate - Malmström *et al.*, 1975) side chain was the sequence serine-glycine (Isemura *et al.*, 1981). The failure to identify residue 4 would be consistent with the presence of a GAG-bearing serine at that position. Poor solubility of glycosylated residues in butyl chloride has been demonstrated (Thomas *et al.*, 1981) so that very little of the glycosylated ATZ-amino acid would be extracted from the sequencer cup. Additionally, no standard was available for such a PTH derivative so that anything extracted would not be identified

The use of ultrafiltration to remove the carbohydrate degradation products following deglycosylation was again successful, as judged by hexosamine data, in the preparation of the PDS(BES)core and PDS(HF)core. The yield at residue 4 of each of these cores, of cysteic acid and serine respectively, was quite low. The ATZ-cysteic acid is not readily soluble in butyl chloride (similar results were obtained for the cysteic acid at residue 7 in the oxidised insulin B chain sample) and serine breakdown was a common problem of the conversion method in use

at that time (Edman & Henschen, 1975). Sequencing of residue 4 of each member of this series of differently deglycosylated core preparations supported the presence of an O-linked substituent at residue 4. A similar strategy will probably be of value in the characterisation of other glycosylation sites.

With a knowledge of the sequence surrounding this residue the work of Chopra *et al.* (1985) proved that this O-substituent was the only dermatan sulphate side-chain on PDS.

The identification of the same amino acid sequence for PDS and each of the cores was sufficient evidence to prove that proteolysis of the amino-terminal end of the molecule had not occurred during the deglycosylation reactions. Although precautions were taken throughout the extraction and purification of the PDS, it cannot be assumed that no such degradation of the molecule occurred during these procedures and therefore it cannot be categorically stated that this is the true amino-terminus of the molecule as it exists *in vivo*.

4.2.1 Isolation and Characterisation of Cyanogen Bromide Peptides of PDS(ABC)Core

The isolation of at least analytical amounts of each of the CNBr peptides was achieved in chromatographic systems using only volatile eluents (TFA, HFBA, acetonitrile). This saved a great deal of time in that lyophilisation was the only step necessary, prior to analysis by either SDS-PAGE or amino acid analysis. As SDS-PAGE was used, especially during the development of the peptide isolation protocol, as a screening method to ascertain the utility of the various chromatographic steps, the speed with which fractions from a chromatogram could be prepared for analysis was very useful.

The initial group separation achieved on Sephacryl S-300 proved to be essential as it reduced the complexity of the samples submitted to subsequent fractionation steps. The S-300 step was omitted from the protocol, on a trial basis, but was reinstated as it was found that subsequent steps could not produce the required resolution of the more complex mixtures.

The resolving power of the Vydac RP-201-TP (C_{18}) column, which was to be so valuable later in the separation of the tryptic and V8 peptides, was recognised during the

separation of the two small CNBr peptides, CB-7 and CB-8. The problem encountered in the detection of CB-8 in gel electrophoretic systems was probably due to its composition and small size. Detection of CB-8 relied on the sensitivity of detection at 214nm made possible by the use of salt-free, high purity eluents, which have low absorbance at this wavelength, in all chromatography.

Peptide CB-6a was never isolated, its sequence was determined from CB-6, which is a product of a partial cleavage comprising peptides CB-6a and CB-7. Peptide CB-2, which has not been isolated, was not fully characterised. It has the same amino-terminal sequence as CB-7 (seen during the sequencing of CB-group B), but is shorter than CB-1, hence the unknown carboxy-terminal detail shown in Fig. 21.

4.2.2 Isolation and Characterisation of Tryptic and Staphylococcus aureus V8 Protease Derived Peptides of CB-Group A

It was anticipated from the high number of lysine and arginine residues in the composition of CB-5, that a large number of small tryptic peptides would be obtained from CB-group A. Cleavage with trypsin occurred at all lysines and arginines. Amino acid analysis indicated a lesser number of glutamic acid cleavage sites for *Staph. aureus* V8 protease. Half of the glutamic acid in this analysis proved to be glutamine. Cleavage occurred at only one aspartic acid (of five in this substrate), supporting the restriction of the specificity of this enzyme to, principally, glutamic acid when an ammonium bicarbonate buffer is used (Drapeau, 1977).

All the tryptic and V8 peptides necessary for the determination of the sequence of CB-5 and its linkage to CB-6 were separated by a single chromatographic method. A very long shallow gradient was employed on a Vydac RP-201-TP (C_{18}) column. The thin layer electrophoresis/thin layer chromatography screening system determined that all the peptides that were sequenced were sufficiently homogeneous to provide unambiguous data. Amino acid analyses of the peptides also showed stoichiometry so close to integer values that little

contamination was anticipated. The fact that most of the peptides were sequenced throughout was very useful in the assembly of the peptide sequences.

4.2.3 Carboxy Terminal Sequencing

The use of the enzyme carboxypeptidase Y offers a considerable improvement over the other carboxypeptidases because of its broader specificity (Martin *et al.*, 1977). The disadvantage of enzymatic sequencing is the lack of the two-stage control inherent in the Edman degradation (coupling and cleavage steps are discrete), which permits the stepwise cleavage of a single residue at a time. The amino acids released from intact PDS and the peptide CB-1 (assumed to be the carboxy-terminal peptide because it lacked homoserine) were the same. However, the apparent order of release varied somewhat between the two substrates. Though a strict interpretation of these data would suggest the assignment of a different sequence to each peptide, there is sufficient similarity to suggest that these two peptides have the same carboxy-terminal sequence and that the differences are due solely to the inconsistency of the method. Another analytical problem which could interfere with the interpretation of the data is the fact that asparagine and glutamine, amino acids not found in an acid hydrolysate, would be released as such by the enzyme and co-elute, in our amino acid analysis system, with serine.

4.2.4 Ordering of Peptides and Assembly of Sequence

The ordering of the epitopes, recognised by the various MAbs, within the molecule proved to be very useful in the arrangement of the sequence data obtained from peptides from other digests. Sufficient data were available to place all the CNBr peptides in order within the molecule. The sequence presented above was proven to be continuous throughout, except for the linkage between peptides CB-7 and CB-1. No evidence has been found to suggest that these peptides are not connected, but definitive proof awaits the analysis of the peptides from the carboxy-terminal region of the molecule.

Some of the sites of carbohydrate attachment have been characterised. Residue 4 is the site of attachment of the single dermatan sulphate chain. Residue 160, a potential site for the attachment of a glycosaminoglycan chain (Johnson & Baker, 1973; Isemura *et al.*, 1981), is not glycosylated. A serine residue was identified at this position after deglycosylation with only chondroitinase ABC. Evidence from the cathepsin C experiments (Chopra *et al.*, 1985) suggested a single glycosaminoglycan chain at residue 4, and no galactosamine was detected in CB-6 or CB-7, both of which include residue 160. Residue 182 is probably an asparagine residue with an attached oligosaccharide chain as no identification was made at this position in sequencing CB-1, an extra aspartate occurs in the composition of peptide B-V-13 and there is a substantial amount of glucosamine associated with this peptide.

4.3 Some Characteristics of the Sequence: Predicted Secondary Structure and Hydropathy Profile

Utilising a comparison of X-ray crystallographic and sequence data from 15 proteins, Chou & Fasman (1974) determined the probability of any amino acid occuring within a region of α-helix or β-sheet. Similar parameters were determined for the probability of a residue occuring within a β-turn and at each of the four positions in a β-turn (Chou & Fasman, 1977). The probability values were revised using a set of 29 proteins, and rules were established for the interpretation of the calculated probabilities (for review see Chou & Fasman, 1978). Utilising these parameters repetitively averaged over a four residue segment throughout the length of available sequence, it is possible to make a prediction of the secondary structure which is adopted by the molecule. The method of Garnier *et al.* (1978) uses a set of similarly derived parameters and an algorithm which assigns a conformational state to each residue.

Circular dichroism spectra, interpreted using the computerised method of Provencher & Glöckner (1981), indicate that PDS comprises 46% β-sheet, 2% helix, 14% β-turn and 39% random coil conformations (Scott *et al.*, 1986). Fig. 27 and Table 10 show the results of predictions of secondary structure made by the methods of Chou & Fasman (1978) and

Garnier *et al.* (1978). The above circular dichroism data were used to select "decision constants" for the modified version of the method of Garnier *et al.*.

The limited agreement between the three methods demonstrates the problems inherent in their use. Attempts have been made to use various methods in tandem; however, the predictive accuracy is not improved (Nishikawa, 1983). The methods agree on the prediction of β-turns involving residue 4, the glycosaminoglycan bearing serine, and residue 182, which is probably the site of attachment of an oligosaccharide side chain. Oligosaccharide and glycosaminoglycan side chains are usually attached to amino acids within β-turns (Aubert *et al.*, 1976; Walton *et al.*, 1979). The inclusion of decision constants based on the circular dichroism data prevented the prediction of the later turn.

A region of β-sheet is predicted, between residues 20-40, by all three methods. This region is believed to be restrained by two disulphide bridges, which would doubtless influence the folding in a manner that cannot be incorporated into these predictive methods. Another prominent feature predicted by all three methods is a helical region at residues 120-140. The strength of the helix forming potential in this region is great as it is predicted even in the modified Garnier *et al.* system.

These, and several other published methods, provide only predictions of secondary structure. Nishikawa (1983) assessed the efficiency of three of these methods by applying them to a number of proteins which lie outside the database used in the derivation of the predictive parameters. This work scored the predictive accuracy of each method at less than 45%.

The fact that globular proteins have a hydrophobic core suggests the infolding of hydrophobic regions and the concentration of hydrophilic regions at the surface of the molecule (Kauzmann, 1959). The method of Hopp & Woods (1981) is designed to predict protein antigenic determinants by analysing amino acid sequences in order to find the point of greatest local hydrophilicity. Numerical hydrophilicity values, assigned to the amino acids, are repetitively averaged along the peptide chain. The method of Krystek *et al.* (1985b) also

Fig. 27. Predictions of secondary structure from the amino acid sequence of PDS using the methods of Chou & Fasman (1974) and Garnier *et al* (1978). Garnier *et al*: method 1 is an unmodified analysis. Method 2 has beemn modified by the inclusion of circular-dichroism parameters derived from analysis of PDS.
"H" indicates the prediction of α-helix, "E"- extended structure or β-sheet, "T"- β-turn and "C"- random coil.

```
        10        20        30        40        50        60
DEASGIGPEEHFPEVPEIEPMGPVCPFRCQCHLRVVQCSDLGLEKVPKDLPPDTALLDLQ
CCTTTTHHHHHHCHHHHHCTTTTCTTTTEEEEEEEEETTTTHHHHHTTTTCCHHHHHHTT
HHTCCCCCCCCCCHHHHCTTCTTETTTTTTTTEEEEEETCHHHEETTTCCTTHHHHHHH
EETCCCCCCCCCCCCTEETTCTTETTEETTEEEEEEEECTEEEEETECCEEEEEEEEC

        70        80        90        100       110       120
NNKITEIKDGDFKNLKNLHTLILINNKISKISPGAFAPLVKLERLYLSKNQLKELPEKMP
TTHHHHTTTTCCCCCEEEEEEEEECCCCCTTTTCEEEHHHHEEEETTTTHHHHCCCTT
HCCEEEEEHHHHHHHHHHHEEEEHHTTTTTEECTCHHHHHHHHHHHHHHHHHHHHHHH
CCEEEEEETTTTTCCTCEEEEEEEETTEEEEETEEEEEEEEEECTTCCEEEECTHCC

        130       140       150       160       170       180
KTLQELRVHENEITKVRKKPFNGLNQMIVVELGTNPLKSSGIENGAFQGMKKLSYIRIAD
TTHHHHHHHHHHHHHHHHHTTTTEEEEEEECCTTTTCCTTTTCCHHHHHHHEEEEEEEET
HHHHHHHHHHHHHHHHHTTCCTTCCCEEEEEETCCCCCCCCCCCHHHHHHHCCEEEEET
HHHHEEEHCHHHHHEEEEECTTCCCEEEEEEETCCCCECEEEEEEEHEEEEEEEEEEEE

        190       200       210
[redacted]NITTIPQGLPPSLTELHLDGNKITKVDAASLKG
TTTCCCTTTTTTTTHHHHTTTTEEEHHHHHH
TTTEEEETCCCCCEEEHHHHHTHEEEHHHHHHHT
EEEEEEETCCCCCEEEEECTTTEEEEEEEEEETT
```

Format:

SEQUENCE
CHOU & FASMAN
GARNIER *et al* 1
GARNIER *et al* 2

TABLE 10
SECONDARY STRUCTURE PREDICTIONS

Method	Helix	β-sheet	β-turn	Coil
Circular Dichroism	2	46	14	39
Chou & Fasman	32	20	33	15
Garnier *et al.* 1[1]	43	20	17	20
Garnier *et al.* 2[2]	6	58	15	21

[1]Unmodified
[2]Modified to incorporate decision constants based on circular dichroism data.

considers differences in the amino acid sequence between the antigen and the homologous protein of the immunised animal. The monoclonal antibodies available in this department were raised against bovine PDS in mice. As the sequence of the mouse PDS is not known, the method of Krystek *et al.* is not applicable in this case.

The striking feature of the hydropathy plot (Fig. 28) is the predominance of regions predicted to be hydrophilic. The regions within which the epitopes, corresponding to the various MAbs available in this department, have been localised may be identified on Fig. 28. The epitope corresponding to the MAb 7B1 is in the region 10-21 (Pringle, 1985). This region is predicted to have two hydrophilic patches which could be involved in epitopes. Similarly, the region containing the epitope for 5D1 (120-147) is predicted to be highly hydrophilic. The epitope for 3B3 however falls in a region (residues 148-170) with only a single predicted zone of moderate hydrophilicity. The epitope for 6D6, on peptide CB-1, falls outside the sequence reported here and is therefore beyond the region shown in this hydropathy plot.

The region containing the two disulphide bridges (22-40) is predicted to be hydrophobic, though not strongly so, suggesting that this region is folded and interiorised.

It is suggested that protein-protein interaction sites tend to be modestly hydrophilic, but also contain residues that could be involved in hydrophobic interactions (Krystek *et al.*, 1985a). This analysis could therefore be of assistance in the determination of the site of interaction between PDS and collagen.

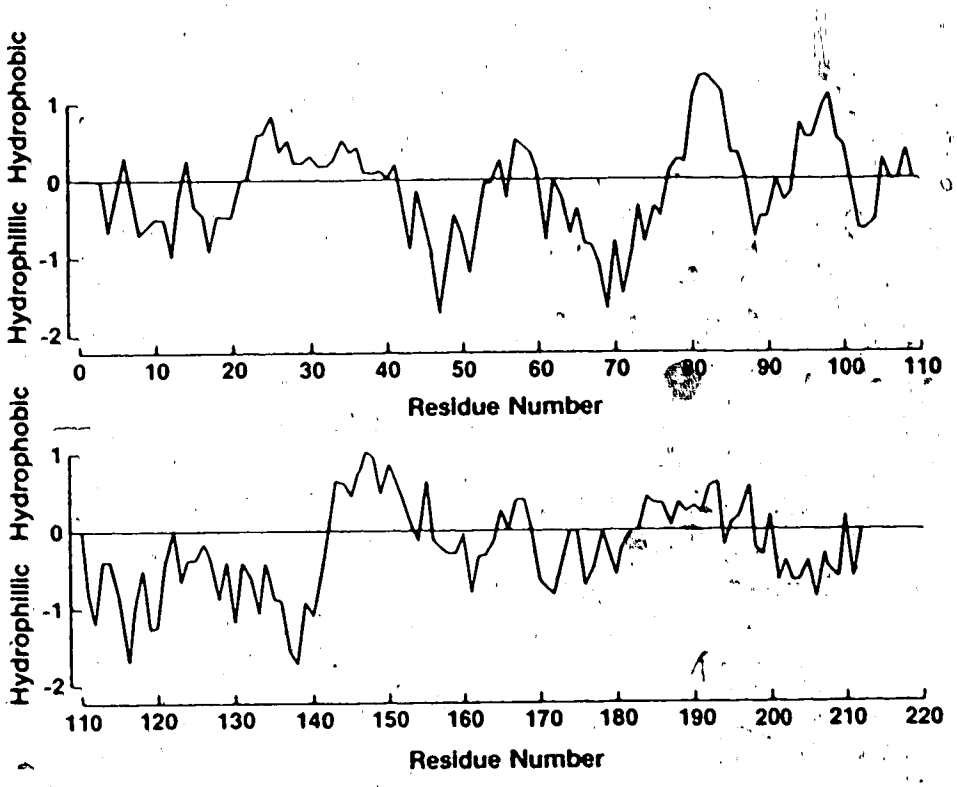


Fig. 28. Hydropathy plot for the first 214 residues of the protein core of PDS. Method of Hopp & Woods (1981). The peptides which are believed to include the epitopes recognised by the MAbs used during this project, are shown.

The consideration of hydropathy in the prediction of β-turns, which usually lie close to the surface of a molecule, has been undertaken by Rose (1978). Most of the turns predicted by all three methods above, correspond to a hydrophilic region as predicted here.

4.4 Tests of Sequence Homology

The available sequence was examined for homology with the sequences of other proteins, previously determined. One programme (MicroGenie) failed to identify any homology. A second programme (IFind), set to search the data-base using less stringent parameters, identified very short homologous sequences in five proteins. Three of the five

homologies were to regions of rabbit cardiac and skeletal muscle myosin chains which showed greater homology to each other than to the PDS sequence. These three homologous regions scored nine residues identical within a region of thirteen, six within a region of seven and eight within a region of thirteen. The other two homologies were to regions of modification methylase (EC 2.1.1.72) from *Streptococcus pneumoniae* (score 6/7) and calcium transporting ATPase (EC 3.6.1.38) from rabbit slow twitch skeletal muscle (score 6/10). The short length and low scores of these few homologies support the unique nature of this sequence among proteins sequenced to date.

A search of the sequence for regions of internal homology identified a region scoring 11 residues identical within a region of 22 as follows:

```
57 LDLQNNKITEIKDGDFKNLKNL 78
81 LILINNKISKISPGAFAPLVKL 102
     * *   ** ** * ** **
```

possibly suggesting the involvement of gene duplication in the evolution of this gene.

Reducing the stringency of the search parameters identified several other short regions (5-10 residues) which demonstrated homologies of about 50%.

Krusius & Ruoslahti (1986) have recently published the complete sequence (329 residues predicted from cDNA sequence) of a small proteoglycan from cultured human fibroblasts, which is very similar to the sequence presented here (see Fig. 29). The initial report from that group (Brennan *et al.*, 1984) made a comparison of their early amino acid sequencing results (23 residues of a proteoglycan from human fetal membranes) to the initial report of sequence data obtained from this project (24 residues: Pearson *et al.*, 1983). The proteoglycans from the two human tissues yielded the same sequence for the first 23 residues. The amino-terminal sequence of the proteoglycan from human tissue was, however, identical to the sequence determined for the bovine skin PDS only for the first nine residues. The rest of

```
1         10         20         30         40         50         60
DEASGIGPEEHFPEVPEIEPMGPVCPFRCQCHLRVVQCSDLGLEKVPKDLPPDTALLDLQ
          VPDDRDFEPSL*                                D          T

          70         80         90         100        110        120
NNKITEIKDGDFKNLKNLHTLILINNKISKISPGAFAPLVKLERLYLSKNQLKELPEKMP
                     A    V      V      T

          130        140        150        160        170        180
KTLQELRVHENEITKVRKKPFNGLNQMIVVELGTNPLKSSGIENGAFQGMKKLSYIRIAD
          A                  NL     I

          190        200        210
TNITTIPQGLPPSLTELHLDGNKITKVDAASLKG
     S                    SR
```

Format:

SEQUENCE Bovine Skin PDS
SEQUENCE Human Fibroblast PDS (Krusius & Ruoslahti, 1986)
Shown only where different.

Fig. 29. A comparison of the first 214 residues of the amino acid sequences of the PDS molecules from bovine skin and human lung fibroblasts.
The asterisk (*) indicates that one residue is omitted from the sequence derived from the human tissue with respect to that derived from bovine.

the sequence was different, although residues 21-24 of the bovine sequence were the same as residues 20-23 of the human sequence. The publication of the full sequence of the PDS from human tissue showed that the amino terminal region apparently contains the only major difference between these molecules and that, with the exception of residues 10-21 and 144-145,

all the differences identified between the two sequences could have resulted from single base mutations in the gene coding for the molecule. Many of these differences are neutral substitutions, in which an amino acid is replaced by one which is functionally very similar. The two molecules differ in 10% of the residues so far identified. Though differences in natural selection pressures will cause differences in the rates of mutation of different proteins, the above value is of the same order as the difference (11%) between the α-haemoglobin chains of bovines and humans (Chai, 1976).

Residues 10-21 were radically different in the two sequences and the omission of one residue has occurred in the human PDS sequence with respect to the bovine. The two regions of the protein core which are obviously structurally specialised are the GAG-bearing region and the region which is conformationally restrained by a pair of disulphide bridges. These two regions are separated by a region of sequence which, to the extent of current knowledge, is the least well conserved between the molecules from human and bovine tissues. This variation suggests that this region is not functionally significant.

With such conservation of the majority of the sequence between the two species, it is likely that the rest of the sequence of the bovine PDS will be virtually identical to that from the human tissue. This suggests that there are two more potential N-glycosylation sites, of the form N-X-S/T (Marshall, 1972), to investigate and that the bovine molecule will be 330 residues long. As no free sulphydryl groups were found in the bovine PDS (Scott *et al.*, 1986) it is likely that the two cysteines at residues 283 and 316 in the human sequence are linked by a disulphide bridge. However, until the sequence of the whole molecule from bovine skin has been completely and independently determined, the possibility of the occurrence of other differences such as seen at residues 10-21 cannot be discounted.

4.5 Conclusions

As a major component of the characterisation of the molecule, the amino acid sequence is very useful in comparing similar molecules from different tissues and species. As the primary gene product, the protein core is the feature by which proteoglycan families should be identified. The post-translational modifications, which may vary within a family based on a single core, can be used to subdivide the family, as the significance of such differences becomes better understood. Many small proteoglycans have been identified which appear to be similar to the molecule studied here. As sequence data becomes available these molecules can be compared to PDS, allowing any structural relationships to be confirmed. Small dermatan sulphate proteoglycans have recently been extracted from tendon (Vogel & Heinegård, 1985), bone (Fisher *et al.*, 1983), lung fibroblast cultures and placental membrane (Brennan *et al.*, 1984) and cornea (Nakazawa *et al.*, 1983). The proteoglycans from tendon and skin have both been shown to have an effect on collagen fibrillogenesis *in vitro* (Vogel *et al.*, 1984; Scott *et al.*, 1986) and it has been shown that they also have the same amino-terminal amino acid sequence (Hassell, Kimura & Hascall, 1986). An antiserum to the tendon PG-II recognises cartilage PG-II. Antisera raised against bone PG-II cross-react with tendon and cartilage PG-II. The cornea PG-II is also related to other members of this group both physically and immunologically (Heingård *et al.*, 1985). Amino acid sequence analyses indicate that the DSPG-II from cartilage (Rosenberg *et al.*, 1985) has the same amino-terminal sequence as the small dermatan sulphate proteoglycan from skin (Hassell, Kimura & Hascall, 1986).

Of the amino acid sequence available to date for the large cartilage proteoglycan, no homology has been identified with the bovine skin PDS sequence. (Walton *et al.*, 1979; Bonnet *et al.*, 1983; Périn *et al.*, 1984; Sai *et al.*, 1986; Doege *et al.*, 1986; Périn *et al.*, 1986). Bourdon *et al.* (1986) have identified a region of homology between the small CSPG from rat yolk sac carcinoma cell line L2 and the prepro core protein sequence of a fibroblast PDS, though there is no homology between the mature proteins.

The virtual identities of the sequences of bovine skin PDS and the fibroblast PDS deduced by Krusius & Ruoslahti (1986) was discussed above. PDS molecules from other bovine tissues, for which some sequence data is available, show identity with the sequence presented here (mature articular cartilage - Rosenberg *et al.*, 1985; tendon - Hassell *et al.*, 1986) suggesting, to the limits of known sequence data, identity of these molecules from different tissues of the same species, and supporting the suggestion that such differences as were seen betweeen bovine skin PDS and the human lung fibroblast PDS were species differences.

4.6 Suggestions for Further Study

The completion of the sequencing project is obviously a priority. Along with the elucidation of the remainder of the primary structure, the identification of the sites of disulphide bridges, the location of the other glycosylation sites, and the characterisation of the oligosaccharides at these sites, will complete this phase of the project. The oligosaccharide sites will probably be identified in the same manner as were the previously determined sites. The enzyme N-glycanase (peptide: N-glycosidase F from *Flavobacterium meningosepticum* which cleaves the N-glycan linkage between asparagine and glucosamine thus releasing, intact, the N-linked oligosaccharide; Tarentino *et al.*, 1985) could also be utilised to remove the carbohydrate moieties allowing positive identification of the substituted amino acids at such sites. The complete elucidation of the disulphide bridges could be approached by the study of a set of peptides derived from peptide A-V-24. Deamidation of the glutamine residue at position 30, in dilute acid, would allow *Staph. aureus* V8 protease cleavage between the two cysteines at residues 29 and 31. The comparison of the peptides thus produced with those produced by tryptic cleavage of the same region will provide sufficient information for the full characterisation of the two disulphide bridges.

The two protein cores recognised for this molecule (Pringle, 1985) and for other similar molecules (fibroblast culture: Glössl *et al.*, 1984; cartilage: Rosenberg *et al.*, 1985; bone: Hassell *et al.*, 1986), under certain conditions, have been suggested to differ only in the number of N-glycosidically linked oligosaccharides (Glössl *et al.*, 1984). The separation of these two species, followed by the study of their relative glycosylation using HPLC-peptide mapping, and sequencing would be very useful in the confirmation that differential glycosylation is the only difference. The determination of the variability in the glycosylation with regard to the size of the individual oligosaccharide chains and which of the glycosylation sites are involved in the variability could also be determined from these experiments.

Pursuing the physico-chemical characterisation of the molecule, X-ray crystallographic determination of the full tertiary structure would seem to be a logical next step. The work

already in progress in Dr. Paul Scott's laboratory involving secondary structure analysis and the application of such structural data to an understanding of the biological interactions of the molecule which seem to be involved in its *in vivo* role, forms another obvious extension of the work on PDS. Using a turbidimetric assay, to monitor the fibrillogenesis of collagen, and other assays under development, Dr. Scott is studying the binding of PDS to collagen and the possible role of PDS in the regulation of fibrillogenesis. The peptides produced and characterised in this project could be put to use in the study of the interaction of PDS with collagen and other molecules, in the study of the glycosylation of the molecule and in the determination of the exact sites of the epitopes recognised by the Mabs, produced by Dr. G.A. Pringle, which were used in this project.

Once the entire sequence of the functional form of this molecule is known, studies can be undertaken to examine its biosynthesis including any precursor forms, post-translational processing, and its breakdown during turnover and in disease or other abnormal states.

REFERENCES

Aaron, B.B. & Gosline, J.M. (1981) Elastin as a random network elastomer. *Biopolymers* 20, 1247-1260.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. (1983) Molecular biology of the cell, Chap. 12. Pub: Garland, New York.

Anderson, J.C. (1975) Isolation of a glycoprotein and proteodermatan sulphate frpm bovine achilles tendon by affinity chromatography on concanavalin A-sepharose. *Biochim. Biophys. Acta.* 379, 444-455.

Antonopoulos, C.A., Axelsson, I., Heinegård, D. & Gardell, S. (1974) Extraction and purification of proteoglycans from various types of connective tissue. *Biochim. Biophys. Acta.* 338, 108-119.

Arcadi, J.A. (1952) *Bull. Johns Hopkins Hosp.* 90, 334.

Aubert, J.P., Biserte, G. & Loucheux-Lefebvre, M.H. (1976) Carbohydrate-peptide linkage in glycoproteins. *Arch. Biochem. Biophys.* 175, 410-418.

Axelsson, I. (1984) Heterogeneity, polydispersity and physiologic role of corneal proteoglycans. *Acta. Ophthalmol.* 62, 25-38.

Bonnet, F., Le Glédic, S., Périn, J.-P., Jollès, J., & Jollès, P. (1983) Identification of cyanogen bromide fragments of the protein core of bovine nasal cartilage proteoglycan monomer. *Biochim. Biophys. Acta.* 743, 82-90.

Bonnet, F., Périn, J.-P. & Jollès, P. (1978) Isolation and chemical characterisation of two distinct link proteins from bovine nasal cartilage proteoglycan complex. *Biochim. et Biophys. Acta.* 532, 242-248.

Bornstein, P. & Traub, W. (1979) The chemistry & biology of collagen. In: The Proteins, Vol 4. Ed: Neurath, H. & Hill, R.L. Pub: Acad. Press. New York. pp411-632.

Bourdon, M.A., Oldberg, A., Pierschbacher, M. & Ruoslahti, E. (1985)Molecular cloning and sequence analysis of a chondroitin sulphate proteoglycan cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1321-1325.

Bourdon, M.A., Shiga, M. & Ruoslahti, E. (1986) Identification from cDNA of the precursor form of a chondroitin sulphate proteoglycan core protein. *J. Biol. Chem.* 261, 12534-12537.

Brauer, A.W., Oman, C.L. & Margolies, M.N. (1984) Use of o-Phthalaldehyde to reduce background during automated Edman degradation. *Analyt. Biochem.* 137, 134-142.

Brennan, M.J., Oldberg, A. Pierschbacher, M.D. & Ruoslahti, E. (1984) Chondroitin/Dermatan sulphate proteoglycan in human fetal membranes. *J.Biol.Chem.* 259, 13742-13750.

Bruckner, P., Vaughan, L. & Winterhalter, K.H. (1986) Type IX collagen identified as proteoglycan Lt. *Ann. NY Acad. Sci.* 460, 397-398.

Byers, P.H., Click, E.M., Harper, E. & Bornstein, P. (1975) Intrchain disulphide bonds in procollagen are located in a large nontriple-helical COOH-terminal domain. *Proc. Natl. Acad. Sci. U.S.A.* 72, 3009-3013.

Carlin, B., Jaffe, R., Bender, B.L. & Chung, A.E. (1981) Entactin: a novel basal lamina associated sulphated glycoprotein. *J. Biol. Chem.* 256, 5209-5214.

Carlstedt, I., Cöster, L., Malmström, A. & Fransson, L.-A. (1983) Proteoheparan sulphate from human skin fibroblasts. *J. Biol. Chem.* 258, 11629-11635.

Cassel, J.M. (1966) Collagen aggregation phenomena. *Biopolymers* 4, 989-997.

Chai, C.K. (1976) Genetic Evolution. Pub: University of Chicago Press, Chicago.

Chakrabarti, B. & Park, J.W. (1980) Glycosaminoglycans: Structure and interaction *CRC Crit. Rev. Biochem.* 8, 225-313.

Chang, J-Y. (1983) Manual micro-sequence analysis of polypeptides using dimethylaminoazobenzene isothiocyanate. In: *Methods in Enzymology,* vol 91, Ed: Hirs, C.H.W. & Timasheff, S.N. Pub: Acad. Press., New York. pp455-466.

Chen, W.T., Hasegawa, E., Hasegawa, T., Weinstock, C. & Yamada, K.M. (1985) Development of cell surface linkage complexes in cultured fibroblasts. *J. Cell. Biol.* 100, 1103-1114.

Chopra, R.K., Pearson, C.H., Pringle, G.A., Fackre, D.S. & Scott, P.G. (1985) Dermatan sulphate is located on serine-4 of bovine skin proteodermatan sulphate. *Biochem.J.* 232, 277-279.

Cöster, L. & Fransson, L.-A. (1981) Isolation and characterisation of dermatan sulphate proteoglycans from bovine sclera. *Biochem. J.* 193, 143-153.

Cöster, L., Fransson, L.-A., Sheehan, J., Nieduszynski, I.A. & Phelps, C.F. (1981) Self association of proteoglycans from bovine sclera. *Biochem. J.* 197, 483-490.

Cöster, L., Malmström, A., Carlstedt, I. & Fransson, L.-A. (1983) The core protein of fibroblast proteoheparan sulphate consists of disulphaide bonded subunits. *Biochem. J.* 215, 417-419.

Chou, P.Y. & Fasman, G.D. (1974) Prediction of protein conformation. *Biochemistry*, 13, 222-245.

Chou, P.Y. & Fasman, G.D. (1977) β-turns in proteins. *J. Mol. Biol.* 115, 135-175.

Chou, P.Y. & Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. In: *Advances in Enzymology*, vol 47. Ed: Meister, A. Pub: Wiley, pp45-148.

Cöster, L. & Fransson, L.A. (1981) Scleral proteodermatan sulphate. *Biochem. J.* 193, 143-153.

Cowan, P.M. & McGavin, S. (1955) Structure of poly-L-proline. *Nature (London)* 176, 501-503.

Damle, S.P., Cöster, L. & Gregory, J.D. (1981) Proteodermatan sulphate isolated from pig skin. *J. Biol. Chem.* 257, 5523-5527.

Davidson, J.M., McEneany, L.S.G. & Bornstein, P. (1977) Intermediates in the conversion of procollagen to collagen. *Eur. J. Biochem.* 81, 349-355.

Deak, F., Kiss, I., Sparks, K., Argraves, W.S., Hampikian, G. & Goetinck, P.F. (1986) Complete amino acid sequence of chicken cartilage link protein deduced from cDNA clones. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3766-3770.

Delphench, B. & Halavent, C. (1981) Characterisation and purification from human brain of a hyaluronic acid-binding glycoprotein, hyaluronectin. *J. Neurochem.* 36, 855-859.

Doege, K., Fernandez, P., Hassell, J.R., Sasaki, M. & Yamada, Y. (1986) Partial cDNA sequence encoding a globular domain at the C-terminus of the rat cartilage proteoglycan. *J. Biol. Chem.* 261, 8108-8111.

Drapeau, G.R. (1977) Cleavage at glutamic acid with staphylococcal protease. In: *Methods in Enzymology*, vol 47. Ed: Hirs, C.H.W. & Timasheff, S.N. Pub: Acad. Press., New York. pp189-191.

Drapeau, G.R., Boily, Y. & Hounard, J. (1972) Purification and properties of an extracellular protease of *S. aureus*. *J.Biol.Chem.* 247, 6720-6726.

Edge, A.S.B., Faltynek, C.R., Hof, L., Reichert, L.E.Jr. & Weber, P. (1981) Deglycosylation of glycoproteins by trifluoromethanesulphonic acid. *Analyt. Biochem.* 118, 131-137.

Edman, P. (1950) Method for determination of the amino acid sequence in peptides. *Acta. Chem. Scand.* 4, 283-293.

Edman, P. & Begg, G. (1967) A protein Sequenator. *Eur.J.Biochem.* 1, 80-91.

Edman, P. & Henschen, A. (1975) Sequence Determination. In: *Protien Sequence Determination.* Ed: Needleman, S.B. Pub: Springer Verlag, New York. pp232-279.

Eggert, F.M. & Jones, M. (1985) Measurement of neutral sugars in glycoproteins as dansyl derivatives by automated HPLC. *J. Chromatogr.* 333, 123-131.

Engel, J., Odermatt, A., Engel, J.A., Madri, H., Furthmayr, H. Rohde, H. & Timpl, R. (1981) Shape, domain organisation and flexibility of laminin and fibronectin, two multifunctional proteins of the extracellular matrix. *J. Mol. Biol.* 150, 97-120.

Fahrenbach, W.H., Sandberg, L.B. & Cleary, E.G. (1966) Ultrastructural studies on early elastogenesis. *Anat. Rec.* 155, 563-568.

Ferguson, K.A. (1964) Starch gel electrophoresis - application to the classification of pituitary proteins and polypeptides. *Metab., Clin. Exp.* 13, 985-1002.

Fisher, L.W., Termine, J.D., Dejter, S.W., Whitson, S.W., Yanagashita, M., Kimura, J.H.,

Hascall, V.C., Kleinmann, H.K. & Hassell, J.R. (1983) Proteoglycans of developing bone. *J. Biol. Chem.* 258, 6588-6594.

Fransson, L.-A., Coster, L., Malmström, A. & Sheehan, J.K. (1982) Self association of scleral proteodermatan sulphate. *J. Biol. Chem.* 257, 6333-6338.

Fujii, N. & Nagai, Y. (1981) Isolation and characterisation of a proteodermatan sulphate from calf skin. *J. Biochem. (Tokyo)* 90, 1249-1258.

Gallagher, J.T., Lyon, M & Steward, W.P. (1986) Structure and function of heparan sulphate proteoglycans. *Biochem. J.* 236, 313-325.

Gallop, P.M. & Paz, M.A. (1975) Posttranslational protein modifications, with special attention to collagen and elastin. *Physiol. Rev.* 55, 418-487.

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120, 97-120.

Gillard, G.C., Merrilees, M.J., Bell-Booth, P.G., Reilly, C.H. & Flint, M.H. (1977) The proteoglycan content and the axial periodicity of collagen in tendon. *Biochem. J.* 163, 145-151.

Glajch, J.L., Gluckman, J.C., Charikofsky, J.G., Minor, J.M., Kirkland, J.J. (1985) Simultaneous selectivity optimisation of mobile and stationary phases in Reversed Phase HPLC for isocratic separations of PTH-amino acids. *J. Chromatogr.* 318, 23-39.

Glössl, J., Beck, M. & Kresse, H. (1984) Biosynthesis of proteodermatan sulphate in cultured human fibroblasts. *J. Biol. Chem.* 259, 14144-14150.

Goldberg, B. & Rabinovitch, M. (1983) Connective tissue. In: Histology. Cell and tissue Biology. Ed: Weiss, L. Pub: Elsevier Biomedical, New York.

Gross, J. & Nagai, Y. (1965) Specific degradation of the collagen molecule by tadpole collagenolytic enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 54, 1197-1204.

Hardingham, T.E., Ewins, R.J.F. & Muir, H. (1976) Cartilage proteoglycans. Structure and heterogeneity of the protein core and the effects of specific protein modifications on the

binding to hyaluronate. *Biochem.J.* 157, 127-143.

Hascall, V.C. (1977) Interaction of cartilage proteoglycans with hyaluronic acid. *J. Supramolec. Struc.* 7, 101-120.

Hascall, V.C. & Sajdera, S.W. (1970) Physical properties and polydispersity of proteoglycan from bovine nasal cartilage. *J. Biol. Chem.* 245, 4920-4930.

Hassell, J.R., Kimura, J.H. & Hascall, V.C. (1986) Proteoglycan core protein families. *Ann. Rev. Biochem.* 55, 539-567.

Hayashi, R. (1977) Carboxypeptidase Y in sequence determination of peptides. In: *Methods in Enzymology,* vol 47, Ed: Hirs, C.H.W. & Timasheff, S.N. Pub: Acad. Press., New York. pp84-93.

Heinegård, D. & Axelsson, I. (1977) Distribution of keratan sulphate in cartilage proteoglycans. *J. Biol. Chem.* 252, 1971-1979.

Heinegård, D., Björne-Persson, A., Cöster, L., Franzén, A., Gardell, S., Malmström, A., Paulsson, M., Sandfalk, R. & Vogel, K. (1985) The core proteins of large and small interstitial proteoglycans from various connective tissues form distinct subgroups. *Biochem. J.* 230, 181-194.

Heinegård, D. & Hascall, V.C. (1979) Characteristics of the non-aggregating proteoglycans isolated from bovine nasal cartilage. *J. Biol. Chem.* 254, 927-934.

Heinegård, D., Paulsson, M., Inerot, S. & Carlström, C. (1981) A novel low molecular weight chondroitin sulphate proteoglycan isolated from cartilage. *Biochem. J.* 197, 355-366.

Heinegård, D. Weislander, J., Sheehan, J., Paulsson, M. & Sommarin, Y. (1985a) Separation and characterisation of two populations of aggregating proteoglycan from cartilage. *Biochem. J.* 225, 95-106.

Heinegård, D. & Paulsson, M. (1984) Structure and metabolism of proteoglycans. In: Extracellular matrix biochemistry. Ed: Piez, K.A. & Reddi, A.H. Pub: Elsevier, New York. pp 277-328.

Hewitt, A.T., Kleinmann, H.K., Pennypacker, J.P. & Martin, G.R. (1980) Identification of

an adhesive factor for chondrocytes. *Proc. Natl. Acad. Sci. U.S.A.* 77, 385-388.

Highberger, J.H., Corbett, C., Dixit, S.N. Yu, W., Seyer, J.M., Kang, A.H. & Gross, J. (1982) Amino acid sequence of chick skin α-1(I)-CB-8 and the complete primary structure of the helical portion of chick skin collagen α1(I) chain. *Biochemistry* 21, 2048-2055.

Hirano, H., Yamada, Y., Sullivan, M., de Crombrugghe, B., Pastan, I. & Yamada, K.M. (1983) Isolation of genomic-DNA clones spanning the entire fibronectin gene. *Proc. Natl. Acad. Sci. U.S.A.* 80, 46-50.

Hofman, H., Fietzek, P.P. & Kühn, K. (1978) The role of polar and hydrophobic interactions for the molecular packing of type-I collagen: A three dimensional evaluation of the amino acid sequence. *J. Mol. Biol.* 125, 137-165.

Hopp, T.P. & Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci.* 78, 3824-2828.

Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. & Woodhead-Galloway, J. (1973) Analysis of the primary structure of collagen for the origins of molecular packing. *J. Mol. Biol.* 79, 137-148.

Hynes, O. & Yamada, K.M. (1982) Fibronectins: Multifunctional modular glycoproteins. *J. Cell Biol.* 95, 369-377.

Isemura, M., Hanyu, T., Kosaka, H., Ono, T. & Ikenaka, T. (1981) Comparative study of carbohydrate complexes. III. Peptide structures of the linkage region in proteoglycans of human, porcine and shark cartilages. *J.Biochem.* 89, 1113-1119.

Johnson, A.H. & Baker, J.R. (1973) Amino acid sequences of peptides from the linkage region of chondroitin sulphate. *Bioch. Soc. Trans.* 1, 277-279

Kang, A.H., Nagai, Y., Piez, K.A. & Gross, J. (1966) Studies on the structure of collagen utilising a collagenolytic enzyme from tadpole. *Biochemistry* 5, 509-515.

Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. In: Advances in protein chemistry. Ed: Anfinsen, C.B.Jr., Anson, M.L., Bailey, K. & Edsall,

J.T. Pub: Acad. Press, New York. Vol 14, pp1-63.

Kivirikko, K.I. & Myllyla, R. (1984) Biosynthesis of collagens. In: *Extracellular matrix biochemistry*. Ed:Piez, K.A. & Reddi, A.H. Pub: Elsevier, New York. pp83-118.

Konat, G., Offner, H. & Mellah, J. (1984) Improved sensitivity for detection and quantitation of glycoproteins on polyacrylamide gels. *Experientia* 40, 303-304.

Krusius, T & Ruoslahti, E. (1986) Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. *Proc. Natl. Acad. Sci. USA*. 83, 7683-7687.

Krystek, S.R., Reichert, L.E. & Andersen, T.T. (1985a) Analysis of computer generated hydropathy profiles for human glycoprotein and lactogenic hormones *Endocrinology* 117, 1110-1124.

Krystek, S.R., Dias, J.A., Reichert, L.E. & Andersen, T.T. (1985b) Prediction of antigenic sites in follicle-stimulating hormones: difference profiles enhance antigenicity prediction methods *Endocrinology* 117, 1125-1131.

Kurkinen, M., Vartio, T. & Vaheri, A. (1980) Polypeptides of human plasma fibronectin are similar but not identical. *Biochim. et Biophys. Acta*. 624, 490-498.

Lai, C.Y. (1977) Regeneration of amino acids from anilinothiazolinones. In: *Methods in Enzymology*, vol 47, Ed: Hirs, C.H.W. & Timasheff, S.N. Pub: Acad. Press., New York. pp369-373.

Liau, Y.H., Galicki, N.I. & Horowitz, M.I. (1978) Heterogeneity of rat rib chondroitin sulphate and susceptibility to rat gastric chondrosulphatase. *Biochim. Biophys. Acta*. 539, 315-323.

Lohmander, L.S., De Luca, S., Nilsson, B., Hascall, V.C., Caputo, C.B., Kimura, J.H. & Heinegård, D. (1980) Oligosaccharides on proteoglycans from swarm rat chondrosarcoma. *J. Biol. Chem.* 255, 6084-6091.

Lowry, O.H., Rosenbrough, N.J., Farr, A.L. & Randall, R.J. (1951) Protein measurement with the Folin reagent. *J. Biol. Chem.* 193, 265-275.

Malmström, A. (1984) Biosynthesis of dermatan sulphate. II. Substrate specificity of the C-5

epimerase. *J. Biol. Chem.* 259, 161-165.

Malmström, A., Fransson, L.-A., Hook, M. & Lindahl, U. (1975) Biosynthesis of dermatan sulphate. I. Formation of L-iduronic acid residues. *J. Biol. Chem.* 250, 3419-3425.

Margolies, M.N., Brauer, A., Oman, C., Klapper, D.G. & Horn, M.J. (1982) Improved automatic conversion for use with a liquid-phase sequenator. In: *Methods in Protein sequence analysis.* Ed: Elzinga, M. Pub: Humana Press, Clifton, NJ. pp189-203.

Marshall, R.D. (1972) Glycoproteins *Annu. Rev. Biochem.*, 41, 673-702.

Martin, B., Svendsen, I. & Ottesen, M. (1977) Use of carboxypeptidase Y for carboxy-terminal sequence determination in proteins. *Carlsberg. Res. Commun.* 42, 99-102.

Mechanic, G. (1979) Collagen Biochemistry. In: *Skeletal Research: An Experimental Approach.* Ed: Simmons, D.J. & Kunin, A.S. Pub: Acad. Press, New York, pp227-241.

Miller, E.J. (1984) Chemistry of the collagens and their distribution. In: Extracellular matrix biochemistry. Ed: Piez, K.A. & Reddi, A.H. Pub: Elsevier; New York. pp41-81.

Miller, E.J. (1985) Recent information on the chemistry of the collagens. In: Chemistry and biology of mineralised tissues. Ed.: Butler, W.T. [Pub: Ebsco Media, Birmingham, Alabama, U.S.A.] pp80-93.

Monson, J.M., Click, E.M. & Bornstein, P. (1975) Further characterisation of procollagen. Purification and analysis of the proα1 chain of chick bone procollagen. *Biochemistry* 14, 4088-4092.

Moore, S. (1963) On the determination of cystine as cysteic acid. ~~*J. Biol. Chem.* 238, 235~~-237.

Mort, A.J. & Lamport, D.T.A. (1977) Anhydrous hydrogen fluoride deglycosylates glycoproteins. *Analyt. Biochem.* 82, 289-309.

Nakamura, T., Matsunaga, E. & Shinkai, H. (1983) Isolation and some structural analyses of a proteodermatan sulphate from calf skin. *Biochem. J.* 213, 289-296.

Nakazawa, K., Newsome, D.A., Nilsson, B., Hascall, V.C. & Hassell, J.R. (1983) Purification of keratan sulphate proteoglycan from monkey cornea. *J. Biol. Chem.* 258,

6051-6055.

Neame, P.J., Périn, J.-P., Bonnet, F., Christner, J.E., Jollès, P. & Baker, J.R. (1985) An amino acid sequence common to both cartilage proteoglycan and link protein. *J. Biol. Chem.* **260**, 12402-12404.

Neame, P.J., Christner, J.E. & Baker, J.R. (1986) The primary structure of link protein from rat chondrosarcoma proteoglycan aggregate. *J. Biol. Chem.* **261**, 3519-3535.

Nishikawa, K. (1983) Assessment of secondary structure prediction of proteins. Comparison of computerised Chou-Fasman method with others. *Biochim. Biophys. Acta.* **748**, 285-299.

Oike, Y., Kimata, K., Shinomura, T., Nakazawa, K. & Suzuki, S. (1980) Structural analysis of chick-embryo cartilage proteoglycan by selective degradation with chondroitin lyases (chondroitinases) and endo-β-D-galactosidase (keratanase). *Biochem.J.* **191**, 193-207.

Oldberg, A., Kjellén, L. & Hook, M. (1979) Cell surface heparan sulphate. *J. Biol. Chem.* **254**, 8505-8510.

Pearson, C.H. & Gibson, G.J. (1982) Proteoglycans of bovine periodontal ligament and skin. *Biochem.J.* **201**, 27-37.

Pearson, C.H., Winterbottom, N., Fackre, D.S., Scott, P.G. & Carpenter, M.R. (1983) The amino-terminal amino acid sequence of bovine skin proteodermatan sulphate. *J. Biol. Chem.* **258**, 15101-15104.

Périn, J.-P., Bonnet, F., Jollès, J., & Jollès, P. (1984) Sequence data concerning the protein core of the cartilage proteoglycan monomers. *FEBS Letts.* **176**, 37-42.

Périn, J.-P., Bonnet, F., & Jollès, P. (1986) Structural relationship between link proteins and proteoglycan monomers. *FEBS Letts.* **206**, 73-77.

Petruska, J.A. & Hodge, A.J. (1964) A subunit model for the tropocollagen macromolecule. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 871-876.

Pierschbacher, M. D. & Ruoslahti (1984) Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature.* **309**, 30-33.

Piez, K.A. (1968)Cross-linking of collagen and elastin. *Annu. Rev. Biochem.* 37, 547-567.

Piez, K.A. (1984) Molecular and aggregate structures of the collagens. In: Extracellular matrix biochemistry. Ed: Piez, K.A. & Reddi, A.H. Pub: Elsevier, New York. pp1-39.

Poole, A.R. (1986)Proteoglycans in health and disease: structures and functions. *Biochem. J.* 236, 1-14.

Pringle, G.A. (1985) Thesis, University of Alberta.

Pringle, G.A., Dodd, C.M., Osborn, J.W., Pearson, C.H. & Mosman, T.R. (1985) Production and characterisation of monoclonal antibodies to bovine skin proteodermatan sulphate. *Collagen Rel. Res.* 5, 23-39.

Provencher, S.W. & Glöckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20, 33-37.

Robert, L. & Moczar, M. (1982) Structural glycoproteins. In: Methods in enzymology. Vol. 82. Ed:Cunningham, L.W. & Frederiksen, D.W. Pub: Acad. Press, New York. pp839-852.

Robinson, H.C., Horner, A.A., Hook, M., Ogren, S. & Lindahl, U. (1978) A proteoglycan form of heparin and its degradation to single-chain molecules. *J. Biol. Chem.* 253, 6687-6693.

Rodén, L. & Horowitz, H.I. (1978) Structure and biosynthesis of connective tissue proteoglycans. In: The glycoconjugates. Vol 2. Ed: Horowitz & Pigman Pub: Acad Press, New York. pp3-71.

Rose, G.D. (1978) Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature* 272, 586-590.

Rosenberg, L.C., Choi, H.U., Tang, L.-H., Johnson, T.L., Pal, S., Webber, C., Reiner, A. & Poole A.R. (1985) Isolation of dermatan sulphate proteoglycans from mature bovine articular cartilages. *J. Biol. Chem.* 260, 6304-6313.

Rosenbloom, J. (1984) Biology of disease. Elastin: Relation of protein and gene structure to disease. *Lab. Invest.* 51, 605-624.

Rosenthal, H.G., Bentley, J.P. & Albin, E.E. (1976) A sensative automated procedure for the determination of glucuronic acid. *Conn.Tiss.Res.* 4, 155-161.

Ross, R. & Bornstein, P. (1969) The elastic fibre. I. The separation and partial characterisation of its macromolecular components. *J. Cell. Biol.* 40, 366-381.

Ruoslahti, E. Engvall, E. & Hayman, E.G. (1981) Fibronectin: current concepts of its structure and functions. *Coll. Res.* 1, 95-128.

Sage, H. & Gray, W.R. (1979) Studies on the evolution of elastin. I. Phylogenetic distribution. *Comp. Biochem. Physiol.* 64B, 313-327.

Sai, S., Tanaka, T., Kosher, R.A. & Tanzer, M.L. (1986) Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein. *Proc. Natl. Acad. Sci. USA.* 83, 5081-5085.

Sakashita, S. Engvall, E. & Ruoslahti, E. (1980) Basement membrane glycoprotein laminin binds to heparin. *FEBS Lett.* 116, 243-246.

Scott, J.E. (1980) Collagen-proteoglycan interactions. *Biochem. J.* 187, 887-891.

Scott, J.E. & Orford, C.R. (1981) Dermatan sulphate-rich proteoglycan associates with rat tail-tendon collagen at the d band in the gap region. *Biochem. J.* 197, 213-216.

Scott, J.E., Orford, C.R. & Hughes, E.W. (1981) Proteoglycan -collagen arrangements in developing rat tail tendon. *Biochem. J.* 195, 573-581.

Scott, J.E. & Haigh, M. (1985) Proteoglycan-type I collagen fibril interactions in calcifying and non-calcifying connective tissues. *Biochem. Soc. Trans.* 13, 933.

Scott, J.E. & Hughes, E.W. (1986) Proteoglycan-collagen relationships in developing chick and bovine tendons. Influence of the physiological environment. *Conn. Tiss. Res.* 14, 267-278.

Scott, P.G. & Pearson, C.H. (1978) Cathepsin D: Cleavage of soluble collagen and cross-linked peptides. *FEBS Lett.* 88, 41-45.

Scott, P.G. & Pearson, C.H. (1981) Cathepsin D: specificity of peptide bond cleavage in type I collagen and effects on type III collagen and procollagen *Eur. J. Biochem.* 114, 59-62.

Scott, P.G., Winterbottom, N., Dodd, C.M., Edwards, E. & Pearson, C.H. (1986) A role for disulphide bridges in the protein core in the interaction of proteodermatan sulphate and collagen. *Bioch. Biop. Res. Commun.* 138, 1348-1354.

Segrest, J.P. & Jackson, R.L. (1972) Molecular weight determination of glycoproteins by polyacrylamide gel electrophoresis in SDS. In: *Methods in Enzymology* vol.28, Ed: Ginsburg, V., Pub: Acad. Press., New York, pp54-63.

Siegel, R.C. (1979) Lysyl oxidase In: *Int. Rev. Conn. Tiss. Res.* 8, 73-118, Ed:Hall, D.A. & Jackson, D.S. Pub: Acad. Press, New York.

Simpson, R.J., Neuberger, M.R. & Liu, T.-Y. (1976) Complete amino acid analysis of proteins from a single hydrolysate. *J.Biol.Chem.* 251, 1936-1940.

Smith, D.W., Weissman, N. & Carnes, W.H. (1968) Cardiovascular studies on copper-deficient swine. XII. Partial purification of a soluble protein resembling elastin. *Biochem. Biophys. Res. Commun.* 31, 309-315.

Smithies, O., Gibson, D., Fanning, E.M., Goodfleish, R.M. & Gilman, J.G. (1971) Quantitative procedures for use with the Edman-Begg sequenator. Partial sequences of two unusual immunoglobulin light chains, Rzf and Sac. *Biochem.*, 10, 4912-4921.

Somack, R. (1980) Complete phenylthiohydantoin amino acid analysis by HPLC on ULTRASPHERE-octadecyltrimethyloxysilane. *Analyt. Biochem.* 104, 464-468.

Stow, J.L., Glasgow, E.F., Handley, C.J. & Hascall, V.C. (1983) Biosynthesis of proteoglycans by isolated rabbit glomeruli. *Arch. Biochem. Biophys.* 225, 950-957.

Suzuki, S. (1960) Isolation of novel disaccharides from chondroitin sulphates. *J. Biol. Chem.* 235, 3580-3588.

Swank, R.T. & Munkres, K.D. (1971) molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with SDS. *Analyt. Biochem.* 39, 462-477.

Tarentino, A.L., Gomez, C.M. & Plummer, T.H.Jr. (1985) Deglycosylation of asparagine-linked glycans by peptide: N-glycosidase F. *Biochem.* 24, 4665-4671.

Tarr, G.E. (1977) Improved manual sequencing methods. In: *Methods in Enzymology*, vol 47,

Ed: Hirs, C.H.W. & Timasheff, S.N. Pub: Acad. Press., New York. pp335-357.

Tarr, G.E., Beecher, J.F., Bell, M. & McKean, D.J. (1978) Polyquaternary amines prevent peptide loss from sequenators. *Anal. Biochem.* 84,622-627.

Tarr, G.E. (1981) Rapid separation of amino acid phenylthiohydantoins by isocratic HPLC. *Analyt. Biochem.* **111**, 27-32.

Thomas, K.A., Baglan, N.C. & Bradshaw, R.A. (1981) The amino acid sequence of the γ-subunit of mouse submaxillary gland 7S nerve growth factor. *J. Biol. Chem.* 256, 9156-9166.

Timpl, R., Rohde, H., Gehron-Robey, P., Rennard, S.I., Foidart, J.M. & Martin, G.R. (1979) Laminin, a glycoprotein from basement membranes. *J. Biol. Chem.* 254, 9933-9937.

Toole, B.P. & Lowther, D.A. (1965) The isolation of a dermatan sulphate-protein complex from bovine heart valves. *Biochim. Biophys. Acta.* **101**, 364-366.

Toole, B.P. & Lowther, D.A. (1968) Dermatan sulphate-protein: Isolation from and interaction with collagen. *Arch. Biochem. Biophys.* **128**, 567-578.

Uldbjerg, N., Malmström, A., Ekman, G., Sheehan, J., Ulmsten, U. & Wingerup, L. (1983) Isolation and characterisation of dermatan sulphate proteoglycan from human uterine cervix. *Biochem. J.* 209, 497-503.

Vogel, K.G. & Heinegård, D. (1985) Characterisation of proteoglycans from adult bovine tendon. *J. Biol. Chem.* 260, 9298-9306.

Vogel, K.G., Paulsson, M. & Heinegård, D. (1984) Specific inhibition of type I and Type II collagenn fibrillogenesis by the small proteoglycan of tendon. *Biochem. J.* 223, 587-597.

Walton, A.G., Volger, H.G. & Jaynes, E.N. (1979) On the structure of proteoglycan core. *Int. J. Biolog. Macromolecules.* **1**, 89-92.

Weber, K. & Osborn, M. (1975) Proteins and sodium dodecyl sulphate: Molecular weight determination on polyacrylamide gels and related procedures. In: *The Proteins*, 3rd. Edition, Vol. 1, Ed: Neurath, H., Hill, R.L. & Boeder, C.L. Pub: Acad. Press, New

York. pp179-223

Zimmerman, C.L., Appella, E. & Pisano, J.J. (1977) Rapid analysis of amino acid thiohydantoins by HPLC, *Analyt. Biochem.* 77, 569-573.