Biochem. J. (1992) 286, 761–769 (Printed in Great Britain)

Hyaluronan-binding region of aggrecan from pig laryngeal cartilage

Amino acid sequence, analysis of N-linked oligosaccharides and location of the keratan sulphate

Frank P. BARRY,* Janette U. GAW, Carmen N. YOUNG and Peter J. NEAME
Shriners' Hospital for Crippled Children and Department of Biochemistry and Molecular Biology, University of South Florida College of Medicine, Tampa, FL, U.S.A.

The hyaluronan-binding region (HABR) was prepared from pig laryngeal cartilage aggrecan and the amino acid sequence was determined. The HABR had two *N*-termini: one *N*-terminal sequence was Val-Glu-Val-Ser-Glu-Pro (367 amino acids in total), and a second *N*-terminal sequence (Ala-Ile-Ser-Val-Glu-Val; 370 amino acids in total) was found to arise due to alternate cleavage by the signal peptidase. The N-linked oligosaccharides were analysed by examining their reactivity with a series of lectins. It was found that the N-linked oligosaccharide on loop A was of the mannose type, while that on loop B was of the complex type. No reactivity was detected between the N-linked oligosaccharide on loop B′ and any of the lectins. The location of keratan sulphate (KS) in the HABR was determined by Edman degradation of the immobilized KS-containing peptide. The released amino acid derivatives were collected and tested for the presence of epitope to antibody 5-D-4. On the basis of 5-D-4 reactivity and sequencing yields, the KS chains are attached to threonine residues 352 and 357. There is no KS at threonine-355. This site is not in fact in G1, but about 16 amino acid residues into the interglobular domain. Comparison of the structure of the KS chain from the HABR and from the KS domain of pig laryngeal cartilage aggrecan was made by separation on polyacrylamide gels of the oligosaccharides arising from digestion with keratanase. Comparison of the oligosaccharide maps suggests that the KS chains from both parts of the aggrecan molecule have the same structure.

INTRODUCTION

The extracellular matrix of cartilage contains a number of proteoglycans whose structure and function have been characterized in some detail. These include the small proteoglycans decorin and biglycan (Fisher *et al.*, 1989), which are substituted with chondroitin sulphate (CS) or dermatan sulphate, and fibromodulin, which contains keratan sulphate (KS) (Oldberg *et al.*, 1989). The largest proteoglycan is aggrecan, which has a polypeptide chain of 2000–2500 amino acids and is richly substituted with KS and CS. Aggrecan is arranged in a series of domains, as was shown originally by electron microscopy after rotary shadowing (Paulsson *et al.*, 1987), and subsequently by sequence analysis (Doege *et al.*, 1987, 1991; Baldwin *et al.*, 1989). The three globular domains are referred to as G1, G2 and G3. G1 is at the *N*-terminus and is separated from G2 by 136 amino acids. G3 is at the *C*-terminus. The majority of the glycosaminoglycan chains occur in the extended region between G2 and G3. The KS attachment sites occur mainly in a region adjacent to the G2 domain and extend for about 230 amino acids in bovine aggrecan (Heinegård & Axelsson, 1977). The remaining part of the extended region is substituted with CS chains, in which the sequence Ser-Gly occurs frequently (Doege *et al.*, 1987). In the KS-rich region, the glycosaminoglycan chains are attached to serine or threonine and the amino acid sequence occurs as a series of tandemly repeated hexapeptides (Antonsson *et al.*, 1989). KS chains are also found in other parts of the molecule (Heinegård & Axelsson, 1977), especially on the G2 domain and interspersed among the CS chains in the CS domain.

The role of the G1 domain in cartilage aggrecan has been well characterized: it binds strongly and specifically to hyaluronan (HA), thereby mediating the formation of large multimolecular aggregates (reviewed by Hardingham, 1981; Kjellén & Lindahl, 1991). These aggregates, which are enmeshed in a network of collagen fibres, contribute a large negative charge density to the extracellular matrix of the tissue, thereby allowing it to absorb large quantities of water. This arrangement confers on cartilage its particular biomechanical properties, and provides a surface of low friction.

The multimolecular complex of aggrecan, HA and link protein is essential to the maintenance of cartilage function. Information on the precise structure of the G1 domain leads to a further understanding of the mechanism of assembly. The primary structure of G1 from rat chondrosarcoma (Neame *et al.*, 1987; Doege *et al.*, 1987), human (Doege *et al.*, 1991) and chicken (Chandresakaran & Tanzer, 1991) aggrecan has been determined. The structure consists of three disulphide-bonded loops, referred to as A, B and B′. Bonnet *et al.* (1986) showed that loop A has a weak but significant sequence similarity with members of the immunoglobulin family. This was further studied by Perkins *et al.* (1989), who used secondary structure predictions to confirm that, despite the modest sequence similarity, loop A can be identified structurally as a member of the immunoglobulin superfamily. Loops B and B′, which appear to be unique to those proteins that bind HA, are referred to as the proteoglycan tandem repeats (PTRs), and are similar in sequence. The domain structure of link protein (Neame *et al.*, 1986) is like that of G1 and the loops are also homologous. Similar HA-binding activity is associated with a number of other cell-surface molecules, such as the lymphocyte homing receptor, gp90^{Hermes} (Goldstein *et al.*,

Abbreviations used: CS, chondroitin sulphate; DSA, *Datura stramonium* lectin; f.a.b.-m.s., fast-atom bombardment m.s.; GNA, *Galanthus nivalis* lectin; HA, hyaluronan; HABR, hyaluronan-binding region; KS, keratan sulphate; MAA, *Maackia amurensis* lectin; PNA, peanut lectin; PTR, proteoglycan tandem repeat; SNA, *Sambucus nigra* lectin.

* To whom correspondence should be addressed at: Shriners' Hospital for Crippled Children, 12502 North Pine Drive, Tampa, FL 33612-9499, U.S.A.

Table 1. Sequences of HABR fragments arising from digestion with proteinases and with cyanogen bromide

See the text for details. Abbreviation: Chy, chymotrypsin.

Peptide	Superose 12 fraction no.	Retention time on reverse-phase h.p.l.c. (min)	Sequence
(*a*) *N*-terminus			VEVSEPDNSLSVSIPQP
			AISVEVSEPDNSLSVSIPQP
(*b*) Trypsin			
T1	22	27.6	VEVSEPDNSLSVSIPQPS
T2	21	27.5	VLLGTSLTIPCYFIDPMHPVXTAPXTAPL
T3	26	18.2	IKWSR
T4	21	24.8	VNSAYQDRVTLPNYPAIPSDATLEIQNLR
T5	21	18.7	SNDSGIYRCEVMH
T6	23	19.2	AISXRYTLDFDR
T7	21	26.3	ACLQNSAIIATPEQLQAAYEDGFHQCDAGWLA
T8	26	17.1	EGCYGDKDEFPGVR
T9	19	29.9	TYGIRDTNETYDVYCFAEEMEGEVFYA
T10	22	22.1	LATTGQLYLAWRGGM
T11	21	16.9	TVYLHANQTGYPDPSSR
T12	16	20.6	YDAICYTGEDFVDIPENFFGVGGEEDITIQTVT
(*c*) Glu-C			
E1	18	41.8	VEVSEPDNSLSVSIPQP
E2	25	21.7	VVLLVATE
E3	21	31.7	GQVRVNSAYQDRVTLPNYPAIPSDATLE
E4	21	23.1	IQNLRSNDSGIYRCE
E5	22	26.2	VMHGIEDSEATLE
E6	22	20.1	FDRAQRACLQ
E7	19	33.1	AGWLADQTVRYPIHT
E8	16	31.5	NFFGVGGEEDITIQTVTWPDVELPLP
(*d*) Asp-N			
D1	24	25.5	EVVVKGIVFHYRAIS
D2	23	19.0	DQTVRYPIHTPREGCY
D3	24	21.1	DEFPGVRTYGIR
D4	26	23.6	DMCSAGWLA
(*e*) CNBr			
CN1	19	23.2	MEGEVFYATSPEKFTFQEAANECRRLGARLAT
CN2	15+16	14.6	XXAGWLADRSVRYPISKARPNCGGNLLGVRT
(*f*) Glu-C/Asp-N			
E1/D1	–	34.9	DNSLSVSIPQPSPLRVLLG
E1/D2	–	20.1	DPMHPVXTAPXTAPLAPRI
(*g*) Glu-C/Chy			
E1/C1	–	12.0	XRVSKEK

1989). This protein has a sequence that is significantly similar to the PTR loops on aggrecan G1. The fibroblast proteoglycan versican has an *N*-terminal domain that binds to HA and whose sequence is also like the link protein and G1 sequences (Zimmermann & Ruoslahti, 1989).

The G1 domain of aggrecan is substituted with a number of N- and O-linked oligosaccharides which may play an important role in directing the folding of the polypeptide chain or in assembly of the ternary complex. The structure of the N-linked oligosaccharides on aggrecan has been investigated in detail by Nilsson *et al.* (1982). They demonstrated that the majority of these have a typical complex-type structure. In addition, Bonnet *et al.* (1985) showed that the HA-binding region (HABR) prepared by tryptic digestion contains KS. In this paper we describe the amino acid sequence of the pig aggrecan HABR and we describe the characterization of two of the N-linked oligosaccharides. We also report the precise location of the KS and compare its chain structure to the KS chains located in the KS domain.

EXPERIMENTAL

Materials

Cartilage was dissected from 2-year-old pig larynxes obtained from a local slaughterhouse. Sepharose CL-6B and Superose 12 columns were from Pharmacia. Enzymes used for sequence analysis were bovine pancreatic trypsin (EC 3.4.21.4), aspartate-*N* proteinase from *Pseudomonas fragi*, glutamate-*C* proteinase (EC 3.4.21.19) and chymotrypsin (EC 3.4.21.1) obtained from Boehringer Mannheim. Cyanogen bromide was from Aldrich. Chondroitinase ABC (*Proteus vulgaris*; EC 4.2.2.4) and keratanase (*Pseudomonas* species; EC 3.2.1.103) were from Seikagaku. Monoclonal antibody 5-D-4 was a gift from Dr. Bruce Caterson, University of North Carolina at Chapel Hill. Goat anti-mouse IgG (peroxidase-conjugated) was from Boehringer Mannheim.

Purification of G1

HABR was prepared from chondroitinase ABC/trypsin-treated A1 fraction as described by Bonnet *et al.* (1985). The ternary complex of HA, HABR and link protein was separated from the rest of the digestion products on Sepharose CL-6B. The HABR and link protein were purified on a Superose 12 column, eluting with 4 M-guanidinium hydrochloride/20 mM-Tris/HCl, pH 6.5. In all the experiments described below, the HABR was reduced and carboxymethylated as described by Hirs (1967).

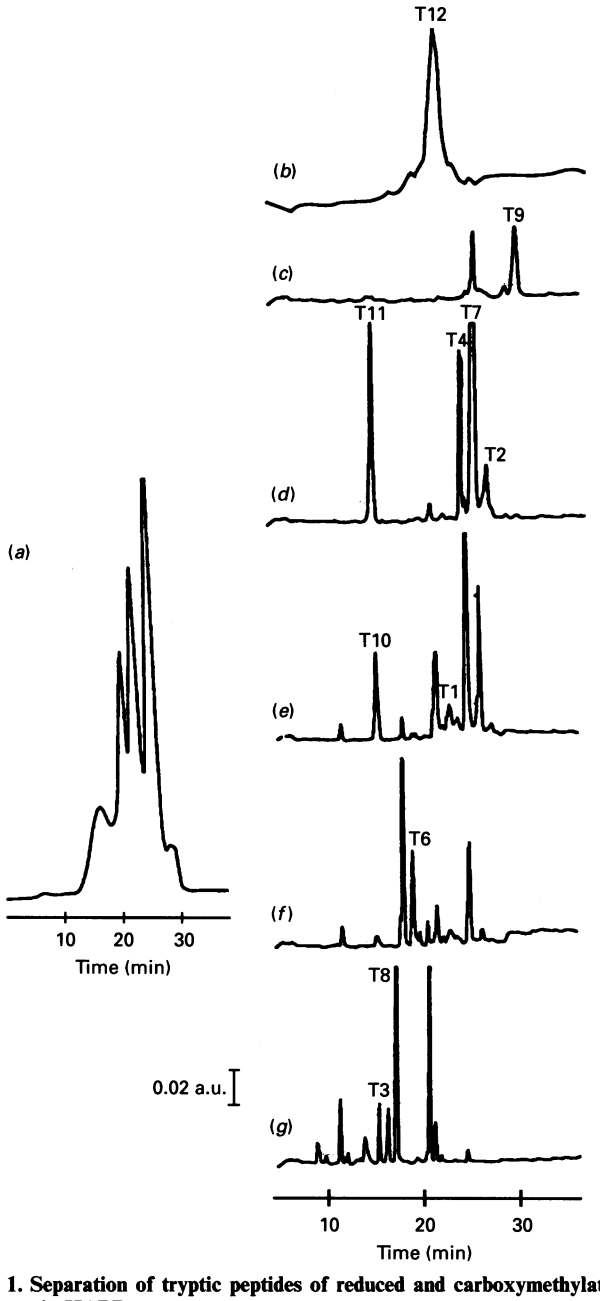

Fig. 1. Separation of tryptic peptides of reduced and carboxymethylated pig HABR

Chromatography was carried out on Superose 12 (chromatogram a) followed by reverse-phase h.p.l.c. (chromatograms b–g). Peptides were eluted from the Superose 12 column in 4 M-guanidinium hydrochloride/20 mM-Tris/HCl, pH 6.5. Collected fractions were applied directly to the h.p.l.c. The profiles shown in (*b*)–(*g*) are from Superose fractions 16, 19, 21, 22, 23 and 26 respectively. The eluate was monitored at 280 nm for the Superose chromatography and at 220 nm for the reverse-phase h.p.l.c. Fraction 16, containing the high-molecular-mass KS-peptide (*b*), was run on a Vydac C4 column; all the other fractions were run on a Vydac C18 column. a.u., absorbance units.

Peptide purification

HABR (100–500 μg) was digested with trypsin, aspartate-*N* proteinase/glutamate-*C* proteinase or cyanogen bromide. In each case the digest mixture was chromatographed on Superose 12 in 4 M-guanidinium hydrochloride/20 mM-Tris/HCl, pH 6.5, and 0.5 ml fractions were collected. These were applied to either a Vydac C18 (4.6 mm × 250 mm) column or a Brownlee RP300 (2.1 mm × 30 mm) column. For the separation of cyanogen bromide fragments, a Brownlee RP-300 (C8) column was used. H.p.l.c. separations were carried out on a Perkin–Elmer Series 410 LC Bio Pump with an LC-325 array detector (Vydac column), or an Applied Biosystems 120A Analyzer (Brownlee C8 and microbore C18 columns).

Sequence analysis

Sequence analysis was carried out on an Applied Biosystems 477A or 473A sequencer. The *N*-terminal sequence was obtained on the intact domain and also on a number of proteolytic fragments. Solid-phase sequence analysis was used for a few peptides. In these cases the peptides were covalently attached to Sequelon membranes by carbodi-imide coupling using the Sequelon-AA covalent attachment kit (Milligen). A modified sequencing cycle was used to collect KS chains released during Edman degradation. After cleavage with trifluoroacetic acid, the released amino acid was eluted with water (from bottle X1 on this instrument), transferred to the conversion flask, and then collected in the fraction collector of the 477 sequencer. An extended drying step (20 min) was then included to prepare the immobilized glycopeptide for the next round of Edman degradation. The glycosylated derivative was then assayed by the e.l.i.s.a. described below.

Amino acid composition analysis

Analyses were carried out on an Applied Biosystems 420A Derivatizer with online h.p.l.c. (Model 130A). Samples were hydrolysed for 45 min at 150 °C and analysed after pre-column derivatization with phenylisothiocyanate. A 5 μg sample of HABR was used for each analysis.

KS analysis

KS was analysed by e.l.i.s.a. using monoclonal antibody 5-D-4 (Caterson *et al.*, 1983). Digestion with keratanase was carried out in 0.1 M-Tris/acetate, pH 7.3. The KS concentration was 1 mg/ml [dimethyl Methylene Blue assay (Farndale *et al.*, 1982)] and 0.5 unit of enzyme was added. Digests were generally carried out in a volume of 10 μl for 3 h at 37 °C. PAGE on 24–30% gradient gels and staining for KS oligosaccharides with Azure A/silver was by the method of Lyon & Gallagher (1990). SDS/PAGE was by the method of Laemmli (1970).

N-linked oligosaccharides

Three peptides were used for analysis of the N-linked oligosaccharides. These were peptides T9, T11 and E4 (see Table 1). Sequencing of these peptides confirmed that they carried N-linked substituents and that they were in pure form. The Glycan Differentiaton Kit (Boehringer Mannheim) was used to determine the type of oligosaccharide on these peptides. This system relies on a series of digoxigenin-labelled lectins and an alkaline phosphatase-conjugated anti-digoxigenin antibody. The lectins employed were GNA (from *Galanthus nivalis*), SNA (*Sambucus nigra*), MAA (*Maackia amurensis*), PNA (peanut) and DSA (*Datura stramonium*). Blotting experiments were carried out using a Bio-Rad Bio-Dot apparatus with 20 pmol of peptide/well. For mass spectrometric analysis a tryptic fragment from bovine HABR was used. This had the sequence Ser-Asn-Asp-Ser-Gly-Ile-Tyr-Arg (SNDSGIYR) and therefore corresponded to the same site of N-glycosylation as peptide E4. Mass spectrometric analysis was carried out by M-Scan Inc. (Brandywine Business Center, West Chester, PA 19380, U.S.A.). Positive-ion fast-atom bombardment (f.a.b.)-m.s. was carried out using a VG Analytical high-field mass spectrometer operating at $V_{acc.}$ = 8 kV. Caesium iodide was used to calibrate the instrument and the spectrum was

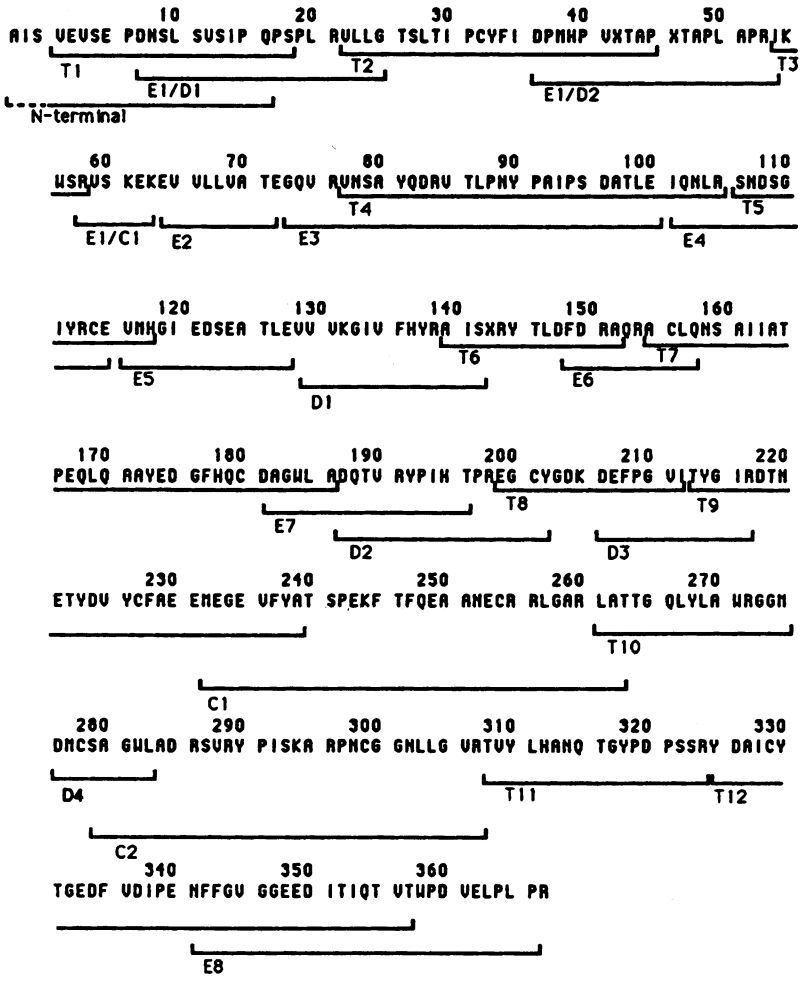

Fig. 2. Amino acid sequence of pig HABR

The peptides which were sequenced and which are shown in Table 1 are underlined to show overlaps. Peptides were prepared by digestion of reduced and carboxymethylated pig HABR with trypsin (T1–T12), glutamate-C proteinase (E1–E8), aspartate-N proteinase (D1–D4), cyanogen bromide (CN1 and CN2), and subdigestion of fragment E1 with aspartate-N proteinase (E1/D1 and E1/D2) and with chymotrypsin (E1/C1).

generated with an M-Scan FAB gun using xenon gas and operating at 10 kV/15 μA. Underderivatized sample was dissolved in 5% (v/v) acetic acid, and permethylated sample was dissolved in methanol.

RESULTS

Sequence analysis

The *N*-terminal sequence of reduced and carboxymethylated HABR was determined to 16 residues using 400 pmol of protein. The result obtained gave two sequences that overlapped by three residues (Table 1*a*). These data showed that the major *N*-terminus of HABR was Val-Glu-Val-Ser-Glu (VEVSE), but a portion of the sample had the *N*-terminus Ala-Ile-Ser-Val-Glu-Val-Ser-Glu (AISVEVSE). The initial yield of AISVEVSE was about 5% of the initial yield of VEVSE. This overlapping *N*-terminal sequence arises from alternative cleavage by the signal peptidase. Reduced and carboxymethylated HABR was digested with trypsin, aspartate-N proteinase, glutamate-C proteinase, chymotrypsin and cyanogen bromide. In each case peptides were separated first by size-exclusion chromatography on Superose 12 in 4 M-guanidinium hydrochloride/20 mM-Tris/HCl, pH 6.5, and then by reverse-phase h.p.l.c. A large proportion of the sequence was determined from tryptic fragments. The other digests were carried out to provide overlaps and to fill in some gaps. Fig. 1 shows the tryptic profiles obtained. Preliminary separation of the peptides and estimation of molecular mass was achieved by chromatography on Superose 12 (Fig. 1*a*). Final purification was achieved by applying each fraction from the Superose 12 column to a reverse-phase h.p.l.c. column. The sequences of 12 tryptic peptides are listed in Table 1(*b*). These peptides are named T1–T12. Digestion with glutamate-C proteinase gave rise to a number of fragments which were purified by the same method. The sequences of eight of these peptides (referred to as E1–E8), which provided suitable overlaps with the tryptic fragments, are shown in Table 1(*c*). Two large gaps in the sequence (residues 240–260 and 283–307; see Fig. 2) were filled by sequencing four aspartate-N fragments (D1–D4; Table 1*d*) and two cyanogen bromide fragments (referred to as CN1 and CN2; Table 1*e*).

The approaches described above did not provide sufficient data to allow the sequence of the first 72 residues of the pig HABR to be determined with a great degree of certainty. Three tryptic fragments (T1, T2 and T3) and one glutamate-C fragment (E2) constituted most of the first 72 residues, but without overlaps. To overcome this the *N*-terminal glutamate-C fragment E1 was subdigested in separate experiments with aspartate-N proteinase and with chymotrypsin. The resulting peptides were separated by reverse-phase h.p.l.c. as before. The former treatment gave rise to two fragments (E1/D1 and E1/D2; Fig. 3),

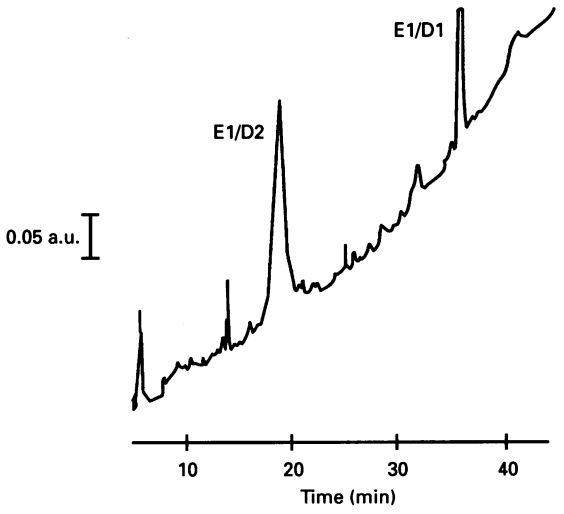

Fig. 3. Separation by reverse-phase h.p.l.c. on a Vydac C18 column of the fragments arising from subdigestion of glutamate-C fragment E1 from pig HABR with aspartate-N proteinase

Fragment E1, arising from digestion of pig HABR with glutamate-C proteinase, was purified by h.p.l.c. and then subdigested with aspartate-N proteinase. The resulting subfragments were purified, again by h.p.l.c. on a Vydac C18 column (4.6 mm × 250 mm). The absorbance of the eluate was monitored at 220 nm. The identities of the peaks are given in Table 1. a.u., absorbance units.

Table 2. Amino acid composition of HABR from pig laryngeal cartilage aggrecan

The composition determined experimentally and the composition deduced from the sequence are given.

	Residues/molecule	
Amino acid	From sequence	Found
Asp	35	35.8
Glu	41	41.9
Cys	10	10.2
Ser	22	22.5
Gly	25	25.6
His	6	8.2
Arg	23	22.5
Thr	27	23.5
Ala	29	29.7
Pro	26	24.6
Tyr	18	19.4
Val	28	23.5
Met	5	5.1
Ile	20	20.5
Leu	27	26.6
Phe	12	12.3
Lys	7	9.2
Trp	5	–
Total	366	361.1

while the latter treatment gave a peptide (E1/C1) with the sequence Xaa-Arg-Val-Ser-Lys-Glu-Lys. Peptide E1/D1 (Asp-Asn-Ser-Leu-Ser) could not be sequenced in good enough yield to provide an overlap with peptide T2 (Val-Leu-Leu-Gly-Thr) by the standard gas-phase protocol. However, when peptide E1/D1 was coupled to Sequelon using carbodi-imide, the yield improved substantially and the sequence could be read to the *C*-terminus. Presumably the hydrophobic *C*-terminal region of E1/D1 caused it to be washed out from the sequencer. No attempt was made to determine the disulphide linkages; these are assumed to be the same as in the rat HABR (Neame *et al.*, 1987). Three residues remained unidentified despite several attempts at sequencing; these were residues 42, 46 and 143. In rat and human aggrecan, for which the cDNA sequences are available (Doege *et al.*, 1987, 1991), these are either serine or threonine. It is possible that the residues are glycosylated or otherwise modified and therefore not detectable. Mass spectrometry of T6 from the bovine HABR indicated that the residue at position 143 had the same mass as threonine. However, derivatives of this amino acid behaved anomalously in both Edman degradation and amino acid analysis.

The amino acid composition of the HABR was determined. The results are given in Table 2 and are compared with the composition deduced from the sequence. There is strong agreement between the two sets of data, suggesting that all of the HABR was sequenced.

N-Linked oligosaccharides

Peptides containing N-linked oligosaccharides were purified by reverse-phase h.p.l.c. and their identity was confirmed by *N*-terminal sequence analysis. HABR has three sites of potential N-glycosylation: Asn-107, Asn-220 and Asn-314. In an attempt to determine the type of oligosaccharide at each of these positions, we used lectin probes with each purified N-oligopeptide immobilized on nitrocellulose. The peptides used were T9, T11 and E4 (Table 1). T9 and T11 arise from loops B and B′ respectively, and peptide E4 from loop A. Approximately 20 pmol of each peptide was used for each lectin using a dot-blot apparatus and the result is shown in Fig. 4. Peptide E4 reacted strongly and specifically with lectin GNA indicating that it contains non-reducing terminal mannose (Shibuya *et al.*, 1988). The strongest binding to GNA occurs with α1-3-linked mannose, but oligosaccharides with mannose in other α-linkages also bind (Shibuya *et al.*, 1988). E4 did not react with any of the other lectins used. Peptide T9 reacted weakly with GNA and very strongly with DSA and MAA (Fig. 4). DSA binds to oligosaccharides that contain Gal(β1-4)GlcNAc (Crowley *et al.*, 1984), and MAA recognizes terminal sialic acid in α2-3 linkage (Wang & Cummings, 1988). Neither of the peptides T9 or E4 reacted with SNA or PNA. These lectins recognize α2-6-linked sialic acid (Shibuya *et al.*, 1987) and the core structure Gal(β1-3)GalNAc (Goldstein & Hayes, 1987) respectively. Peptide T11 did not react with any of the lectins used. The type of N-linked oligosaccharide on loop A of the bovine HABR was also investigated by mass spectrometry. In this case the tryptic fragment carrying the oligosaccharide was purified by the methods described above. The positive ion f.a.b.-m.s. spectrum (not shown) gave a major $(M+H)^+$ pseudomolecular signal at m/z 2289, with sodium adduct ions at 2312 and 1334. These molecular ions are consistent with the presence of the following glycopeptide structure as the major component:

(Ser-Asn-Asp-Ser-Gly-Ile-Tyr-Arg)
|
($Hexose_6$ $HexNAc_2$)

Additional pseudomolecular ion signals and their possible assignments are given in Table 3. It is not possible by f.a.b.-m.s. to identify the precise monosaccharide stereochemistry, but it is highly likely that the hexose residues are mannose and the hexosamine residues are glucosamine.

KS analysis

Bonnet *et al.* (1985) observed a decrease in the molecular mass of pig laryngeal HABR of some 6 kDa after treatment with

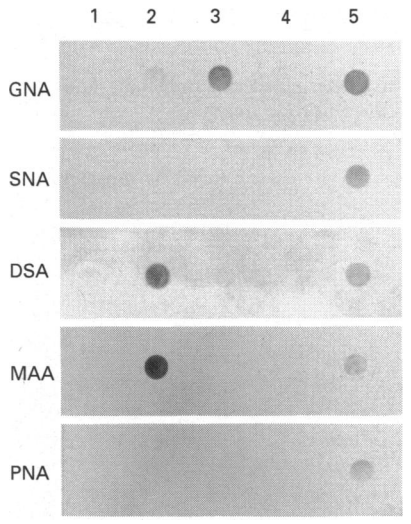

Fig. 4. Analysis by dot-blot with lectin probes of peptides containing N-linked oligosaccharides from pig HABR

Analysis, by dot-blot with lectin probes, of peptides T11 (lane 1), T9 (lane ,2) and E4 (lane 3) from pig HABR. Lane 4 contains ribonuclease as a negative control, and lane 5 contains positive controls carboxypeptidase Y (GNA), fetuin (SNA and MAA) and asialofetuin (DSA and PNA); 0.1 μg of each control glycoprotein/well was used except for carboxypeptidase Y, which was present at 1 μg/well.

Table 3. Mass spectrometric analysis of the N-glycosylated tryptic peptide from loop A of bovine HABR

Additional signals and possible assignments are given.

m/z	Possible assignment
2451	SNDSGIYR-(Hexose$_7$HexNAc$_2$)
2473	Sodium adduct of 2451
2495	Disodium adduct of 2451
2613	SNDSGIYR-(Hexose$_8$HexNAc$_2$)
2636	Sodium adduct of 2613
2127	SNDSGIYR-(Hexose$_5$HexNAc$_2$)
2149	Sodium adduct of 2127
2171	Disodium adduct of 2127
2169	(Hexose$_8$HexNAc$_2$+H)$^+$; pseudomolecular ion
2137	(Hexose$_8$HexNAc$_2$)$^+$; A-type fragment ion
1965	(Hexose$_7$HexNAc$_2$+H)$^+$; pseudomolecular ion
1933	(Hexose$_7$HexNAc$_2$)$^+$; A-type fragment ion
1761	(Hexose$_6$HexNAc$_2$+H)$^+$; pseudomolecular ion
1729	(Hexose$_6$HexNAc$_2$)$^+$; A-type fragment ion
1557	(Hexose$_5$HexNAc$_2$+H)$^+$; pseudomolecular ion
1525	(Hexose$_5$HexNAc$_2$)$^+$; a-type fragment ion
1484	(Hexose$_6$HexNAc$_1$)$^+$; pseudomolecular ion
1321	(Hexose$_6$HexNAc$_2$+H)$^+$; A-type fragment ion
1280	(Hexose$_5$HexNAc$_1$)$^+$; A-type fragment ion
219	(Hexose$_1$)$^+$; A-type fragment ion

keratanase. They also observed that the content of galactose and glucosamine was decreased. A Western blot of reduced and carboxymethylated pig HABR using antibody 5-D-4 as probe confirms this (Fig. 5). It is noteworthy that the anti-KS antibody reacts most strongly with the higher-molecular-mass forms of HABR, consistent with a heterogeneous population of glycopeptides, the larger having more KS. We were interested in determining the precise site(s) of attachment of the KS. The approach we adopted was to monitor the release of KS chains during Edman degradation of a peptide covalently attached to a

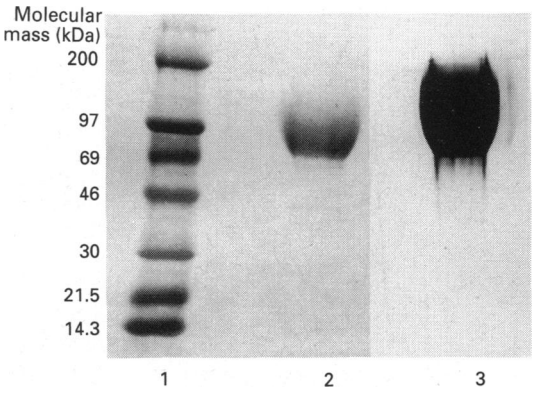

Fig. 5. Electrophoresis of pig HABR on a 5–20% gradient polyacrylamide gel in the presence of SDS

Lane 1, molecular mass markers; lane 2, pig HABR (20 μg) stained with Coomassie Blue; lane 3, an immunoblot of HABR (20 μg) probed with antibody 5-D-4.

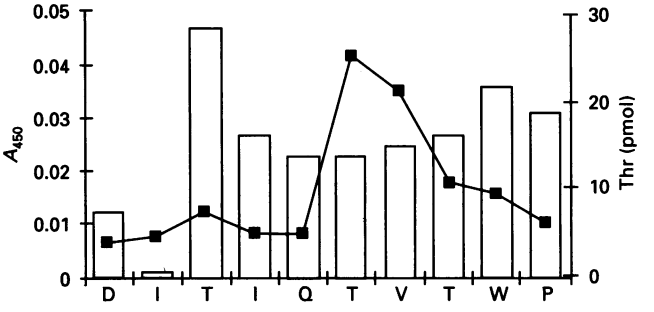

Fig. 6. Release of epitope to antibody 5-D-4 after Edman degradation of peptide E8 covalently attached to a Sequelon membrane

Peptide E8 was attached to Sequelon by carbodi-imide coupling and was subjected to Edman degradation in the 477A sequencer using a modified reaction cycle. Fractions were collected corresponding to each cleaved Edman product and were tested in an e.l.i.s.a. using antibody 5-D-4 and a peroxidase-conjugated second antibody. The bars show the absorbance at 450 nm. A second sample of the same peptide was sequenced by conventional methods and the yield of threonine at each cycle is given by the line.

solid support. The advantage of this procedure is that the KS-containing anilinothiazolinone can be eluted from the membrane with water. Under these conditions the remaining peptide stays attached to the support. Peptide E8 (Asn-Phe-Phe-Gly-Val) was sequenced after immobilization on Sequelon using carbodi-imide. After six normal sequencing cycles the parameters were changed to include a water elution step after each trifluoroacetic acid cleavage, and the released anilinothiazolinone was collected in a fraction collector. These fractions were dried and their reactivity to antibody 5-D-4 was measured by an e.l.i.s.a. (Fig. 6). There was an increase in the amount of epitope in fraction 12, which corresponds to threonine-352, the first residue in the sequence Thr-Ile-Gln-Thr-Val-Thr. These data suggest that threonine-352 is substituted with KS. It is not possible to deduce from these data whether the other threonines in this sequence are similarly substituted. However, the increased yield of threonine-355 compared with threonines-352 and -357 strongly suggests that it is not substituted, while the other two carry KS chains (Fig. 6).

Digestion of HABR with keratanase gave rise to a series of oligosaccharides that could be separated on 24–30% gradient polyacrylamide gels and visualized by staining with azure A/silver

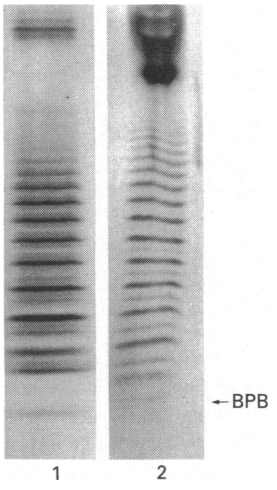

Fig. 7. Electrophoresis on a 24–30% gradient polyacrylamide gel of the oligosaccharide products arising from digestion with keratanase

Pig aggrecan KS domain (lane 1) and pig HABR (lane 2) were digested with keratanase and the resulting oligosaccharides were separated by electrophoresis on polyacrylamide gels with a 24–30% gradient. Staining was with azure A/silver. BPB, Bromophenol Blue.

(Fig. 7). The purpose of this experiment was to explore the structure of the KS chains at this position and to compare their structure with the KS chains that occur in the KS domain. The KS domain was therefore prepared from pig laryngeal A1 fraction by digestion with chondroitinase ABC and trypsin. The resulting fragments were separated by chromatography on a column of Sepharose CL-6B in 0.5 M-sodium acetate, pH 6.8. The Dimethyl Methylene Blue-positive fractions which eluted from the column were pooled and the *N*-terminal sequence and amino acid composition determined. These results (not shown) indicated that the fragment consisted of the entire KS domain of pig aggrecan and that the KS chains were still attached to the polypeptide core. The results also showed that the KS-rich fragment was pure. This fragment was digested with keratanase in the same manner and the oligosaccharide products were also separated on a 24–30% gradient polyacrylamide gel. The pattern of bands was compared with those obtained from the HABR KS chains; the distribution of bands was essentially similar, suggesting that the distribution of glycosidic bonds which are susceptible to digestion by keratanase is the same in KS chains from HABR and from the KS domain.

DISCUSSION

Sequence analysis

The amino acid sequence of HABR was determined from a series of proteolytic and cyanogen bromide fragments that provided sufficient overlap for most of the sequence to be deduced without ambiguity (Fig. 2). The only weak point is residues 64–72. Fragment E2 had the sequence Val-Val-Leu-Leu-Val-Ala-Thr-Glu. The assignment of Glu to position 64 was made on the basis of identity with other sequences and because fragment E2 was generated by digestion with glutamate-C proteinase, which cleaves peptide bonds *C*-terminal to Glu.

Two *N*-termini were observed during sequencing of intact HABR. The major *N*-terminus was Val-Glu-Val-Ser-Glu, and the minor sequence was Ala-Ile-Ser-Val-Glu. The latter sequence presumably arises because of alternate cleavage by the signal peptidase. The mature *N*-terminus in rat HABR is known from the protein sequence (Neame *et al.*, 1987) and the complete sequence of human aggrecan was determined from cDNA clones (Doege *et al.*, 1991). Alignment of these with the pig sequence (Fig. 8) shows that the signal peptidase acts with similar specificity in human and pig aggrecan, cleaving at Thr-Val and Ser-Val respectively. In rat aggrecan cleavage is at a Ser-Glu bond. Secondary cleavage in pig aggrecan occurs at an Xaa-Ala bond; in both rat and human aggrecan the residue that aligns with Xaa is Ala. If this is also the case with pig aggrecan then the secondary cleavage has a different specificity about the residue *N*-terminal to the cleaved bond.

Alignment of the pig HABR sequence with those from rat and human is shown in Fig. 9. There is a very high degree of similarity between the three sequences. Loop B is more strongly conserved than loops A and B′; this is also the case in the link protein sequences known to date (Neame *et al.*, 1986; Deák *et al.*, 1986; Dudhia & Hardingham, 1989). In addition, the gp90^{Hermes} lymphocyte homing receptor contains a domain that is more closely similar to the B loop of aggrecan and link protein than to the B′ loop. These loops are involved in binding to HA (Périn *et al.*, 1987) through a mechanism that is as yet unclear, but which most likely involves the positively charged amino acids which are clustered in loops B and B′. In the three sequences shown in Fig. 9, and in bovine G1 (Neame, P. J. & Rosenberg, L. C., unpublished work) and chicken G1 (Chandresakaran & Tanzer, 1991) most of these basic residues are conserved. Interaction between G1 and HA therefore probably involves

Fig. 8. Amino acid sequences around sites of cleavage by signal peptidase in rat, human and pig aggrecan

Arrows indicate cleavage sites.

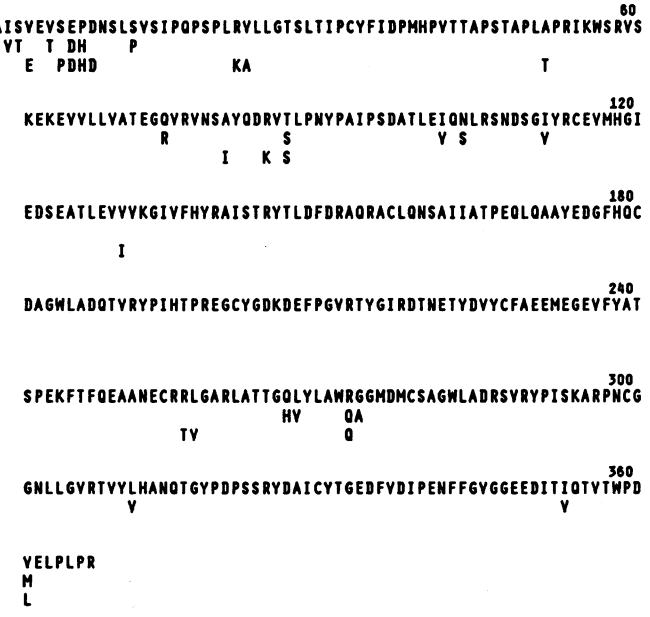

Fig. 9. Sequence of pig aggrecan HABR (top) aligned with the sequences of HABR from human (centre) and rat (bottom) aggrecans

The pig sequence is given in full (top). For the rat and human sequences, only those residues which differ are shown.

electrostatic interactions. This idea is supported by the data of Hardingham *et al.* (1976), who found that chemical modification of lysine and arginine residues resulted in complete loss of binding.

N-Linked oligosaccharides

The structure of the N-linked oligosaccharide on each of the three loops of pig HABR was investigated by lectin blotting (Fig. 4). The purpose of this experiment was to determine if there are differences in the type of N-linked substituent between loops A, B and B′. The N-oligopeptide from loop A was purified after digestion with glutamate-C proteinase (peptide E4). It reacted strongly with GNA but not with any of the other lectins. This suggests that this oligosaccharide is of the high-mannose type (Kornfeld & Kornfeld, 1985). This conclusion is supported by the results of m.s. analysis of an analogous tryptic peptide from bovine HABR (Table 3). The N-linked oligopeptide from loop B was purified as a tryptic fragment (peptide T9). This oligopeptide reacted with lectins DSA and MAA and very weakly with GNA. The reactivity with DSA indicates the presence of Galβ1-4GlcNAc or lactosamine, and reactivity with MAA indicates that it contains terminal sialic acid α2-3-linked to galactose. Both of these observations suggest that this oligosaccharide has a structure typical of complex-type N-oligosaccharides (Nilsson *et al.*, 1982; Kornfeld & Kornfeld, 1985). Peptide T11, from the B′ loop, did not show a reaction with any of the lectins used. This may be due to the fact that this N-linked oligosaccharide has a structure that is unlike those on loops A and B.

It is of course possible that the reactivity of peptides T9 and E4 with the lectins is due to O-linked oligosaccharides. However, the presence of such structures would cause a decreased yield or absence of serine and/or threonine during sequence analysis. This was not observed. In addition, m.s. of the tryptic fragment from bovine HABR did not reveal the presence of any additional oligosaccharide structures.

Nilsson *et al.* (1982) have shown that about 70% of the N-linked oligosaccharides on aggrecan are of the complex type. The fact that loop A has a different type of N-substituent to loops B and B′ is interesting. Loop A is a member of the immunoglobulin family on the basis of modest sequence similarity (Bonnet *et al.*, 1986). Secondary structure predictions (Perkins *et al.*, 1989) confirm that the sequence of loop A can in fact be identified with crystal structures for the variable Ig fold. Loops B and B′ appear to be unique to proteins that bind HA. There may be some significance in the fact that the hydroxyamino acid which follows the glycosylated asparagine in loop A is serine, while in loops B and B′ it is threonine. Bause & Legler (1984) have show that, in synthetic peptides, the rate of N-glycosylation is higher for peptides that contain threonine. Since the synthesis of mannose-type oligosaccharides precedes the synthesis of complex types, perhaps the structure of the N-linked oligosaccharide on loop A arises because of slower glycosylation. The N-linked oligosaccharides on loops B and B′ would be synthesized more quickly, and thus would be available for more extensive processing.

Lohmander *et al.* (1983) reported that proteoglycan monomer and link protein synthesized by rat chondrosarcoma chondrocytes in culture in the presence of tunicamycin, which inhibits N-oligosaccharide synthesis, retained the ability to aggregate. Therefore the N-oligosaccharides do not apparently play a role in directing the binding of the HABR to HA or link protein. Furthermore, the sequence of chicken G1 (Chandresakaran & Tanzer, 1991) shows a somewhat different distribution of potential N-glycosylation sites compared with mammalian G1. It is possible that the N-linked oligosaccharides on HABR confer protection against denaturing conditions or against degradation by intra- or extra-cellular proteinases. It is also possible that they play a role in cell–matrix or cell–cell signalling.

KS chains were located on threonines residue 352 and 357 by measuring the recovery of epitope to antibody 5-D-4 after sequential Edman degradation of peptide E8. This, in combination with yield measurements (Fig. 6), suggested that threonines 352 and 357, but not 355, were substituted with KS. This site is not within the HABR domain but extends some 23 residues into the interglobular domain which, in rat and human aggrecan, is about 136 residues long. The KS oligosaccharides produced by digestion of HABR with keratanase were analysed by PAGE (Fig. 7). KS oligosaccharides from the KS-rich domain of pig aggrecan were digested and analysed in the same way. The oligosaccharide map was similar in both cases, suggesting that there are no major structural differences between the KS chains. There is no similarity, however, between the peptide sequence where KS substitution occurs in the interglobular domain and the consensus hexapeptide sequence that occurs in the bovine KS domain (Antonsson *et al.*, 1989) and in the pig KS domain (F. P. Barry and P. J. Neame, unpublished work). Fosang *et al.* (1991) have shown that there is a site on the interglobular domain that is sensitive to digestion by the metalloproteinase stromelysin. This cleavage occurred at the Asn^{341}-Phe^{342} bond, which is 11 residues *C*-terminal to the site of KS attachment. It is conceivable that the adjacent KS chains provide protection against attack by this or other proteinases.

We have demonstrated the presence of mannose-type oligosaccharides on loop A of the HABR, which has not been reported previously. The structure is different from the complex-type N-linked oligosaccharide on loop B. The KS chains reported by Bonnet *et al.* (1985) occur on the interglobular domain close to the B′ loop, and the amino acid sequence around this site shows no similarity with KS attachment sites within the KS domain. It will be interesting to examine the role which these post-translational modifications play in aggrecan function.

This work was funded by National Institutes of Health Grant AR35322, and by the Shriners of North America.

REFERENCES

Antonsson, P., Heinegård, D. & Oldberg, A. (1989) J. Biol. Chem. **264**, 16170–16173
Baldwin, C. T., Reginato, A. M. & Prockop, D. J. (1989) J. Biol. Chem. **264**, 15747–15750
Bause, E. & Legler, G. (1984) Biochem. J. **195**, 639–644
Bonnet, F., Dunham, D. & Hardingham, T. E. (1985) Biochem. J. **228**, 77–85
Bonnet, F., Périn, J.-P., Lorenzo, F., Jollès, J. & Jollès, P. (1986) Biochim. Biophys. Acta **873**, 152–155
Caterson, B., Christner, J. E. & Baker, J. R. (1983) J. Biol. Chem. **258**, 8848–8854
Chandresakaran, L. & Tanzer, M. L. (1991) J. Cell Biol. **115**, 124A
Crowley, J. F., Goldstein, I. J., Arnarp, J. & Lönngren, J. (1984) Arch. Biochem. Biophys. **231**, 524–533
Deák, F., Kiss, I., Sparks, K. J., Argraves, W. S., Hampikian, G. & Goetinck, P. F. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3766–3770
Doege, K., Sasaki, M., Horigan, E., Hassell, J. R. & Yamada, Y. (1987) J. Biol. Chem. **262**, 17757–17767
Doege, K. J., Sasaki, M., Kimura, T. & Yamada, Y. (1991) J. Biol. Chem. **266**, 894–902
Dudhia, J. & Hardingham, T. E. (1989) J. Mol. Biol. **206**, 249–253
Farndale, R. W., Sayers, C. A. & Barrett, A. (1982) Connect. Tissue Res. **9**, 247–248
Fisher, L. W., Termine, J. D. & Young, M. F. (1989) J. Biol. Chem. **264**, 4571–4576
Fosang, A. J., Neame, P. J., Hardingham, T. E., Murphy, G. & Hamilton, J. A. (1991) J. Biol. Chem. **266**, 15579–15582
Goldstein, I. J. & Hayes, C. E. (1987) Adv. Carbohydrate Chem. Biochem. **35**, 127–340

Goldstein, L. A., Zhou, D. F. H., Picker, L. J., Minty, C. N., Bargatze, R. F., Ding, J. F. & Butcher, E. C. (1989) Cell **56**, 1063–1072
Hardingham, T. (1981) Biochem. Soc. Trans. **9**, 489–497
Hardingham, T. E., Ewins, R. J. F. & Muir, H. (1976) Biochem. J. **157**, 127–143
Heinegård, D. & Axelsson, I. (1977) J. Biol. Chem. **252**, 1971–1979
Hirs, C. H. W (1967) Methods Enzymol. **11**, 59–62
Kjellén, L. & Lindahl, U. (1991) Annu. Rev. Biochem. **60**, 443–475
Kornfeld, R. & Kornfeld, S. (1985) Annu. Rev. Biochem. **54**, 631–664
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Lohmander, L. S., Fellini, S. A., Kimura, J. H., Stevens, R. L. & Hascall, V. C. (1983) J. Biol. Chem. **258**, 12280–12286
Lyon, M. & Gallagher, J. T. (1990) Anal. Biochem. **185**, 63–70
Neame, P. J., Christner, J. E. & Baker, J. R. (1986) J. Biol. Chem. **261**, 3519–3535
Neame, P. J., Christner, J. E. & Baker, J. R. (1987) J. Biol. Chem. **262**, 17768–17778
Nilsson, B., De Luca, S., Lohmander, S. & Hascall, V. C. (1982) J. Biol. Chem. **257**, 10920–10927
Oldberg, A., Antonsson, P., Lindblom, K. & Heinegård, D. (1989) EMBO J. **8**, 2601–2604
Paulsson, M., Mörgelin, M., Wiedemann, H., Beardmore-Gray, M., Dunham, D., Hardingham, T., Heinegård, D., Timpl, R. & Engel, J. (1987) Biochem. J. **245**, 763–772
Périn, J.-P., Bonnet, F., Thurieau, C. & Jollès, P. (1987) J. Biol. Chem. **262**, 13269–13272
Perkins, S. J., Nealis, A. S., Dudhia, J. & Hardingham, T. E. (1989) J. Mol. Biol. **206**, 737–753
Shibuya, N., Goldstein, I. J., Broekart, W. F., Nsimba-Lubakis, M., Peeters, B. & Peumans, W. J. (1987) J. Biol. Chem. **262**, 1596–1601
Shibuya, N., Goldstein, I. J., Van Damme, E. J. M. & Peumans, W. J. (1988) J. Biol. Chem. **263**, 728–734
Wang, W.-C. & Cummings, R. D. (1988) J. Biol. Chem. **263**, 4576–4585
Zimmermann, D. R. & Ruoslahti, E. (1989) EMBO J. **8**, 2975–2981

Received 15 January 1992/18 March 1992; accepted 26 march 1992