Revisiting some O-Glycans

Table 1. Different types of O-linked glycans in humans.

Type of O-linked glycan	Structure and peptide linkage	*Glycoprotein*
Mucin-type	(R)-GalNAc1-Ser/Thr	*Secreted + plasma membrane*
GAG	(R)-GlcA1–3Gal1–3Gal1–4Xyl1-Ser	*Proteoglycans*
O-linked GlcNAc	GlcNAc1-Ser/Thr	Nuclear and cytoplasmic
O-linked Gal	Glc1–2Gal1-*O*-Lys	*Collagens*
O-linked Man	NeuAc2–3Gal1–4GlcNAc1–2Man1-Ser/Thr	α-Dystroglycan
O-linked Glc	Xyl1–3Xyl1–3Glc1-Ser	*EGF protein domains*
O-linked Fuc	NeuAc2–6Gal1–4GlcNAc1–3Fuc1-Ser/Thr	*EGF protein domains*
	Glc1–3Fuc1-Ser/Thr	*TSR repeats*

Wopereis et al., 2006, Clinical Chemistry 52:574-600

Table 2. Diversity of mucin-type O-linked glycans.

Core	Structure	Human tissue
1	Gal1–3GalNAc	Most cells and secreted proteins
2	Gal1–3 (GlcNAc1–6)GalNAc	All blood cells
3	GlcNAc1–3GalNAc	Colon and saliva
4	GlcNAc1–3 (GlcNAc1–6)GalNAc	Mucin-secreting cell types
5	GalNAc1–3GalNAc	Meconium
6	GlcNAc1–6GalNAc	Ovarian tissue
7	GlcNAc1–6GalNAc	
8	Gal1–3GalNAc	Bronchia

Wopereis et al., 2006, Clinical Chemistry 52:574-600

Glycosaminoglycans and Proteoglycans

Glycosaminoglycans: anionic polysaccharides made up of repeating disaccharide units

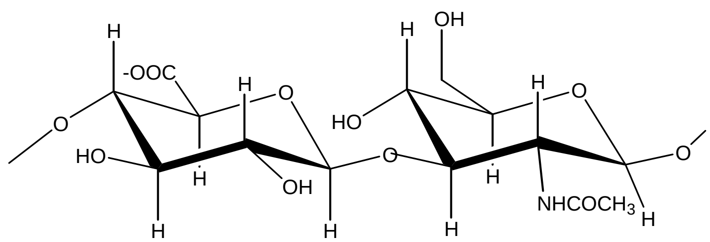
*** one of the sugars is an amino sugar: *N*-acetyl glucosamine (GlcNAc) or *N*-acetyl galactosamine (GalNAc)**

*** at least one of the sugars is negatively charged due to a carboxyl group or a sulfate**

Wopereis et al., 2006, Clinical Chemistry 52:574-600

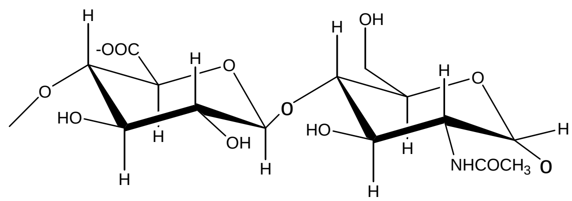
Examples:

hyaluronic acid (hyaluronan)
[D-GlcA-β1,3-D-GlcNAc-β1,4-]n

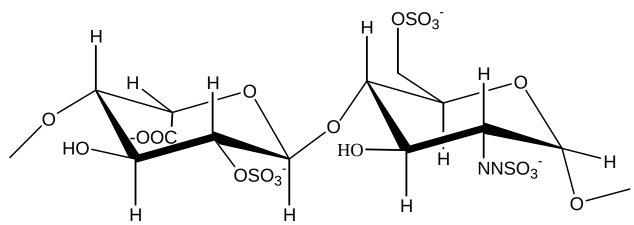


heparan sulfate / heparin

[D-GlcA β1,4-D-GlcNAc- α1,4-]n

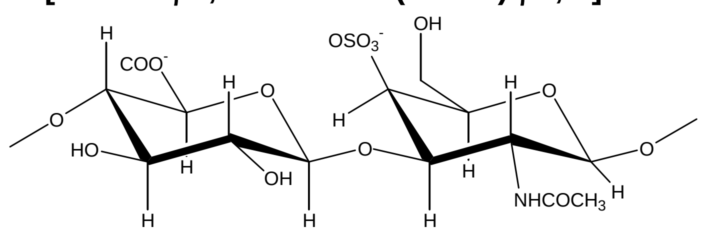


or [L-IdoA(2-OSO3)-α1,4-D-GlcNSO3(6-OSO3)-α1,4-]n



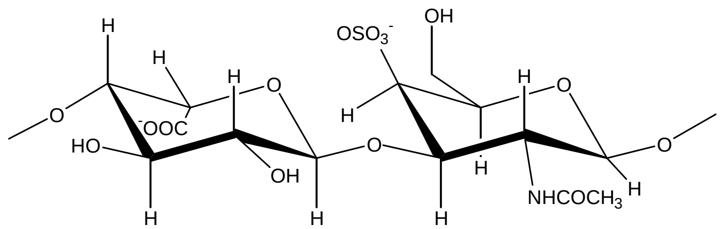
chondroitin sulfate (CS)

[D-GlcA β1,3-D-GalNAc(4-SO3)-β1,4-]n



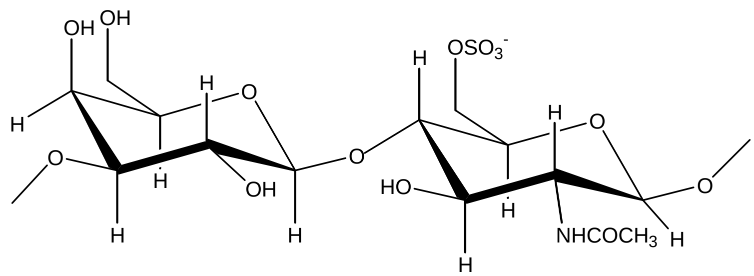
dermatan sulfate (DS),

[L-IdoU α1,3-D-GalNAc(4-SO3)-β1,4-]n



keratan sulfate

[D-Gal β1,4-D-GlcNAc(6-SO3)-β1,3-]n



Most GAGs are synthesized onto proteins and exit as glyconjugates known as proteoglycans

The only GAG known to be synthesized independent of a protein is hyaluronic acid (HA, hyaluronan). HA is synthesized at the plasma membrane while proteoglycans are synthesized in the ER/Golgi.

HA is an essential polysaccharide in animals and is found in pathogenic bacteria and viruses. HA is a structural element in the vitreous humor of eye, synovial fluid and skin of vertebrates. HA interacts with proteins (e.g. CD44, RHAMM, fibrinogen) & influences angiogenesis, cancer, cell motility, wound healing and cell adhesion.

HA is a polymer made up of the repeating disaccharide

[-GlcA-β1,3-GlcNAc-β1,4-]n

n can exceed 30,000 (i.e. Mr >10^7)

HA is synthesized by HA synthase: an integral membrane, Mn^{++} or Mg^{++} requiring, protein that has both β1,3GlcAT and β1,4GlcNAcT activity. HA synthases range from 419-588 amino acids (~49-69 kD). The vertebrate enzymes (DG42, HAS1, HAS2, HAS3), the streptococcal HasA, and the Paramecium bursaria chloreall virus A98R share several regions of sequence similarity.

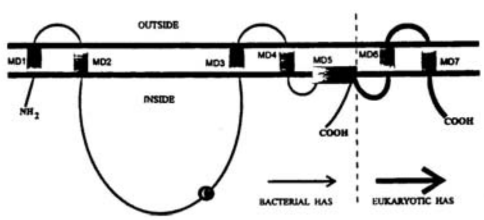


FIG. 3. **Proposed membrane topology for the HAS family.** Very similar hydropathy plots and primary structure (28–71% identity) among all the HAS isozymes suggest that they are similarly organized within the membrane. The scheme depicts the N and C termini and the large central domain, between MD2 and MD3, inside the cell. The larger eukaryotic HASs (*thick line*) have additional amino acids in all regions (see Fig. 2) compared with the bacterial HASs (*thin line*), except for the highly conserved carboxyl 178 residues of the central domain and MD1–MD5. In particular, the carboxyl ~25% of the eukaryotic HASs has two additional predicted membrane domains (MD6 and MD7), missing in the bacterial proteins. The conserved Cys is indicated by the *circled C*. MD5 can be modeled as an amphipathic helix, which would orient the C terminus of all HAS members inside.

Expression of an active recombinant HAS from *Streptococci pyogenes* in *E. coli* showed that a single purified gene product synthesized high Mr HA when incubated *in vitro* with UDP-GlcA and UDP-GlcNAc (DeAngelis & Weigel, P.H., 1994, Biochemistry 33:9033-9039)

Group A Streptococci has an HA synthesis operon (*has*) composed of 3 genes: *has*A (hyaluronate synthase); *has*B (UDP-Glc dehydrogenase); *has*C (UDP-Glc pyrophosphorylase)

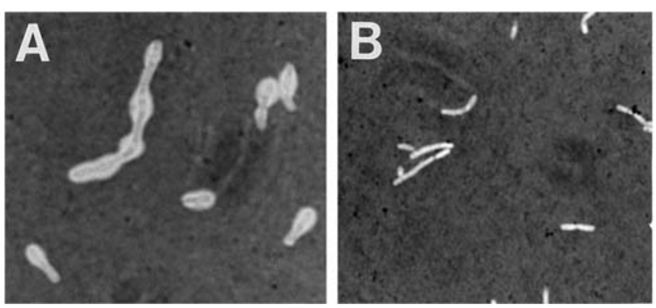


Fig. 4. Photomicrographs of recombinant E. coli.
Panel A, India ink staining (1,000× magnification) reveals that E. coli K5 cells with pPmHAS produce a substantial capsule that appears as a white halo around the cells.
Panel B, the capsular material could be removed from the E. coli K5(pPmHAS) cells by brief treatment with Streptomyces HA lyase. PmHAS directs polymerization of the HA polysaccharide. *Pasteurella multocida* is a prevalent gram negative animal pathogen. It produces **hyaluronan to avoid host defenses**. *Pasteurella* HAS is a class II HAS. (DeAngelis *et al*., 1998, J.Biol.Chem. 273:8454)

Proteoglycans: proteins that contain one or more covalently bound glycosaminoglycans (GAGs)

Proteoglycan = protein core + ≥ 1 GAG

To date > 30 genes for protein core of different proteglycans identified

Secreted proteoglycans

SLRPS

(small leucine-rich proteoglycans)

Modular proteoglycans

***Basement Membrane Proteoglycans*:**

perlecan, agrin, bamacan

***Hyalectans*:**

versican, aggrecan neurocan, brevican

SLRP

Modular

TABLE 1. *Structure and properties of secreted pericellular proteoglycans*

General features	Designation: Gene product	Designation: Gene	Protein core[a]: Size (kDa)	Glycosaminoglycan: Type [Number]	Chromosomal location: Human	Chromosomal location: Mouse	Tissue distribution
SLRP	Decorin	DCN	36	CS/DS [1]	12q21.3-q23	10	Ubiquitous, collagenous matrices, bone, teeth, mesothelia, floor plate
Small, ubiquitous PCs enriched in leucine, with 24 amino acid tandem repeats flanked by cysteine clusters.	Biglycan	BGN	38	CS/DS [1-2]	Xq28	X	Interstitium, and cell surfaces
	Fibromodulin	FMOD	42	KS [4]	1q32		Collagenous matrices
	Lumican	LUM	38	KS [2-3]	12q21.3-q22	10,distal	Cornea, intestine, liver, muscle, cartilage
	Epiphycan[b]		36	CS/DS [2]			Epiphyseal cartilage
Modular Multidomain proteoglycans with protein modules homologous to the Ig superfamily, selectin, EGF, laminin, LDL receptor, N-CAM and protease inhibitors.	Versican	CSPG2	265-370	CS/DS [10-30]	5q13.2	13	Blood vessels, brain, skin, cartilage
	Aggrecan	ACC1	220	CS [≈ 100]	15q26	7	Cartilage, brain, blood vessels
	Neurocan	MNC1	136	CS [3-7]		8	Brain, cartilage
	Brevican		100	CS [1-3]			Brain
	Perlecan	HSPG2	400-467	HS/CS [3-10]	1p36	4, distal	Basement membranes, cell surfaces, sinusoidal spaces, cartilage
	Agrin	ACRN	200	HS [3-6]	1p32-pter	4	Synaptic sites of neuromuscular junctions, renal basement membranes
	Testican		44	HS/CS [1-2]	21		Seminal fluid

[a] The size is based on the amino acid sequence deduced from cDNA cloning. In general, however, the size of the individual protein cores is larger when estimated by SDS-PAGE due to varying degrees of N- and O-linked glycosylations.
[b] This proteoglycan, originally named PG-Lb (6), has been recently renamed epiphycan (L. Rosenberg and M. Höök, personal communication) to reflect its typical tissue distribution in the epiphyseal cartilage.

Iozzo & Murdoch, 1996, FASEB 10:598

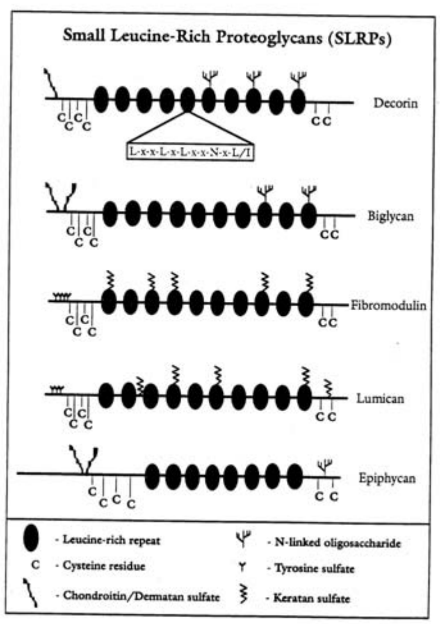


Figure 1. The family of the small leucine-rich proteoglycans. The consensus sequence for the LRR is shown in the rectangle. A key to the various structural components is given in the bottom panel.

SLRPs

Iozzo & Murdoch, 1996, FASEB 10:598

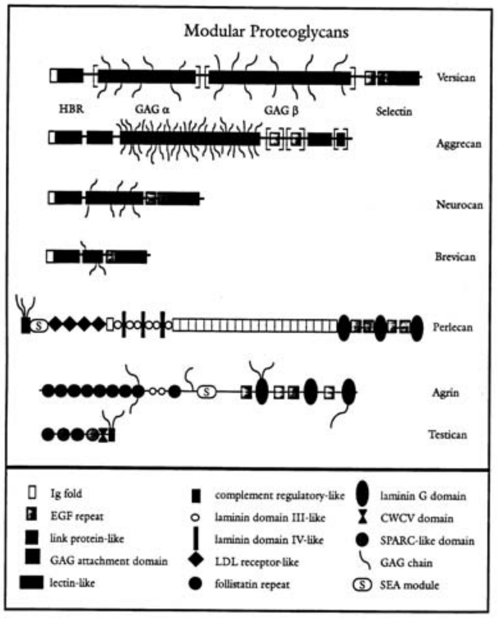


Figure 3. A schematic diagram of the modular proteoglycans. The hyalectans, the proteoglycans that harbor hyaluronan- and lectin-binding modules at opposite ends, comprise versican, aggrecan, neurocan, and brevican. The nonhyaluronan-binding modular proteoglycans comprise perlecan, agrin, and testican. A key to the various protein modules and structural motifs is provided in the bottom panel. Parentheses indicate the site of alternative splicing.

Iozzo & Murdoch, 1996, FASEB 10:598

Functions of proteoglycans:

- tissue organizers
- influence cell growth & maturation of specialized tissues
- biological filters
- modulate growth-factor activities
- regulate collagen fibrillogenesis & skin tensile strength
- affect tumor cell growth & invasion
- influence corneal transparency
- influence neurite outgrowth
- some proteoglycans are essential for growth, others may be redundant

Heparin and Heparan Sulfate (as examples)

Definitions

Heparanoids: glycoconjugates that contain heparin or heparan sulfate

* HSPG: heparan sulfate proteoglycan

FGF: fibroblast growth factor

bFGF: basic fibroblast growth factor

Location

* on cell surface and basement membrane of animals cells

* mast cells, mucosa, lung

Functions

* anticoagulant (heparin)

* binds FGF (heparin, heparan sulfate)

* free HS fragments bind bFGF and may extend half life of bFGF

* binds laminin, fibronectin

* cell surface phenomenon (heparan sulfate)

* bind to binding proteins on cell surface (heparin and heparan sulfate; binding to clusters of Arg & Lys on protein surface)

* many heparinoids must bind two proteins simultaneously to carry out their physiological role (e.g. antithrombin/thrombin interaction with heparin requires at least 18 consecutive glycosyl residues)

* bind selectins, protease inhibitors, lipoprotein lipase (HS)

* implicated in cell proliferation, differentiation, adhesion, migration, morphogenesis, blood coaggulation, viral & bacterial infection, tumor cell malignancy

Structure

HS is a complex mammalian GAGs whose molecular fine structure varies in a cell-type specific fashion.

Basic repeat of alternating GlcN and uronic acids that contain *N*- and *O*-linked sulfate. Heparin is a member of the HS family.

Heparin: higher N-and O-sulfate content than HS; greater proportion of the uronic acid as iduronic acid

HS: greater proportion of uronic acid is GlcA.

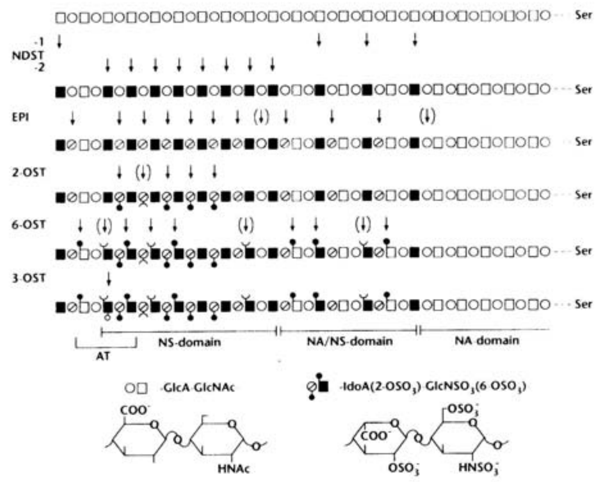
Generalized structures

Heparin: [L-iduronic acid(2S)-α1,4-GlcN(NS,6S)-α1,4-]n

HS: [D-GlcNAc-α1,4-D-GlcA-α1,4-D-GlcNAc-] or
[GlcN(NS)-GlcA-GlcNAc]

Heparan sulfate: block structures composed of highly sulfated, iduronic acid-rich sequences separated by unsulfated, GlcA-rich sequences (relative lengths of sulfated and unsulfated sequences vary significantly)

Minireview: Heparan Sulfate



NS: N-sulfated
NA: N-acetylated

hrough compositional chains containing polybasic cell retention sequences. By con-
ent bovine (10) or hu- trast, the binding to FGF-2 was affected to a much lesser

Biosynthesis

*** heparan sulfate proteoglycan isoforms and structures differ in different cell types and the heparan sulfate chains in each cell type are mixtures, likely representing different biosynthetic maturity**

Attachment sequence in core protein

HS & CS often found on Ser at Ser-Gly or Gly-Ser

Biosynthesis of protein linkage region

primer: core proteins

CS shares same linkage structure as HS

xylosyltransferase (same XylT for CS and HS)

```
protein +        UDP-Xyl  →         protein + UDP
  |                                   |
 Ser                                 Ser -Xyl
```

consensus sequence for the recognition signal of XylT

a-a-a-a-G-S-G-a-b-a, with a=E or D and b=G, E, or D

Subsequent reactions with UDP-Gal, UDP-GalA and UDP-GlcA and the enzymes galactosyl transferase I (same for CS and HS) and another GalT and a GlcAT

Linkage region

```
protein
  |
 Ser-Xyl-Gal-Gal-GlcA
```

GlcNAcT I transfers GlcNAc from UDP-GlcNAc onto the non-reducing terminal GlcA of linkage region to begin HS synthesis

GalNAcT I transfers GalNAc from UDP-GalNAc onto the non-reducing GlcA terminal of linkage region to initiate CS synthesis

Elongation of GAG chain

GlcNAcT II for HS

GalNAcT II for CS

GlcNAcT & GlcAT in single protein

co-purification (2000-fold), similar pI, similar migration by SDS-PAGE. Note: a single enzyme synthesizes N-acetylheparosan in E.coli K 5, composes of alternating GLcA and GlcNAc) (*Petit et al 1995*)

Revisiting some N-Glycans

There is heterogeneity in the glycosylation at a given glycosylation site in proteins.

Fetuin: major glycoprotein in fetal calf serum; Mr 48,000, contains 6 oligosaccharides: 3 O-linked; 3 N-linked

 Asn-linked Oligosaccharide Structures on Fetuin

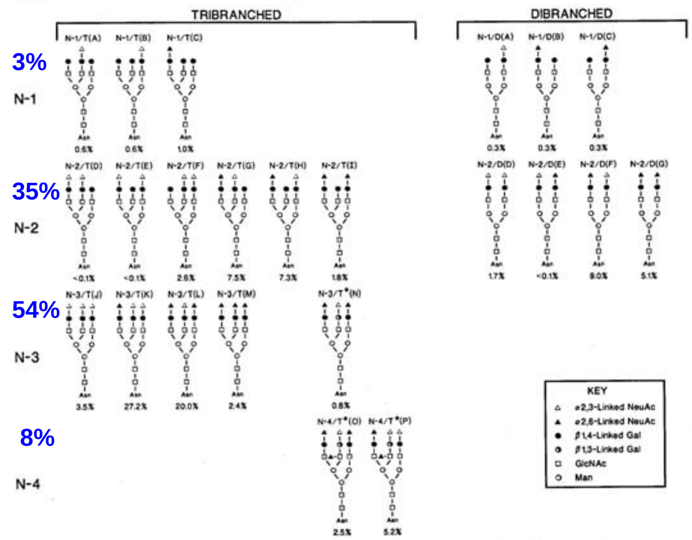


Fig. 7. **Quantitative relationships of the Asn-linked oligosaccharides on fetuin.** Structures of the 23 Asn-linked oligosaccharides isolated from fetuin are schematically illustrated along with the percentage of total Asn-linked oligosaccharides represented by each. Detailed structural assignments are provided in Fig. 6.

Green et al., 1988, JBC 263:18253