Biodiversity of CS-proteoglycan Sulphation Motifs: Chemical Messenger Recognition Modules with Roles in Information Transfer, Control of Cellular Behaviour and Tissue Morphogenesis.

Anthony Hayes\textsuperscript{a}, Kazuyuki Sugahara\textsuperscript{b,c}, Brooke Farrugia\textsuperscript{d}, John M Whitelock\textsuperscript{d}, Bruce Caterson\textsuperscript{e}, James Melrose\textsuperscript{e,f}

\textsuperscript{a}Bioimaging Research Hub, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK. \textsuperscript{b}Graduate School of Life Science, Hokkaido University, Sapporo, Japan. \textsuperscript{c}Department of Pathobiochemistry, Faculty of Pharmacy, Meijo University, Nagoya, Japan. \textsuperscript{d}Graduate School of Biomedical Engineering, University of New South Wales, Sydney 2052, NSW, Australia. \textsuperscript{e}School of Biosciences, Cardiff University, Cardiff, CF10 1AX, Wales, UK. \textsuperscript{f}Raymond Purves Bone and Joint Research Laboratories, Kolling Institute of Medical Research, Royal North Shore Hospital and University of Sydney, St. Leonards, NSW 2065, Australia.

\¶\¶Address correspondence to: - Dr. J. Melrose, Raymond Purves Bone and Joint Research Laboratories, Level 10, Kolling Institute of Medical Research B6, The Royal North Shore Hospital, St. Leonards, NSW 2065, Australia. Ph +61 2 9926-4806, Fax +61 2 9926-5266 Email: james.melrose@sydney.edu.au
Abstract

Chondroitin sulphate glycosaminoglycan chains on cell and ECM proteoglycans can no longer be regarded as merely hydrodynamic space fillers. Overwhelming evidence over recent years indicates that sulphation motif sequences within the chondroitin sulphate chain structure are a source of significant biological information to cells and their surrounding environment. Chondroitin sulphate sulphation motifs have been shown to interact with a wide variety of bioactive molecules e.g. cytokines, growth factors, chemokines, morphogenetic proteins, enzymes and enzyme inhibitors, as well as structural components within the extracellular milieu. They are therefore capable of modulating a panoply of signalling pathways thus controlling diverse cellular behaviours including proliferation, differentiation, migration and matrix synthesis. Consequently, through these motifs, chondroitin sulphate proteoglycans play significant roles in the maintenance of tissue homeostasis, morphogenesis, development, growth and disease. Here we review (i) the biodiversity of chondroitin sulphate proteoglycans and their sulphation motif sequences and (ii) the current understanding of the signalling roles they play in regulating cellular behaviour during tissue development, growth, disease and repair.
1. Introduction

Chondroitin sulphate (CS) and its sulphation motifs on cell associated, pericellular and extracellular matrix (ECM) proteoglycans (PGs) represent a significant repository of information in tissues with the capacity to encode functional information rivalling that of RNA, DNA and proteins [1]. This information is realised when CS and its sulphation motifs interact with growth factors, cytokines, morphogenetic proteins, enzymes, inhibitors and pericellular matrix (PCM) and ECM stabilising glycoproteins. Such interactions have diverse effects on cellular metabolism, proliferation and differentiation, cell migration, matrix synthesis/stabilisation and tissue remodelling in development and are critical to the cellular control of tissue homeostasis. CS sulphation motifs on cell-associated, PCM and ECM proteoglycans also provide important molecular recognition and activity signals to stem/progenitor cell niches facilitating the sequestration of combinations of growth factors, cytokines and chemokine’s which maintain the niche microenvironment ensuring stem cell survival and their maintenance in a state of quiescent self-renewal within the niche environment. Perturbations in the signals which stem cells receive in this niche can also orchestrate stem/progenitor cell differentiation and pluripotency resulting in stem cell activation and proliferation into specific cell lineages with migratory properties facilitating their participation in tissue growth, development and repair processes.

Virtually every cell in the human body is surrounded by a dense glycocalyx of glycoconjugates consisting of mixtures of glycoproteins, proteoglycans and glycolipids which provide a protective and interactive barrier [2]. CS is a prominent glycosaminoglycan (GAG) component of many of these molecules and is the most abundant GAG in the human body [3]. The glycocalyx connects the cell to its external microenvironment and components in the glycocalyx act as biosensors through which cells perceive and respond to changes in the environment they live in [4, 5]. The endothelial glycocalyx also has mechanosensory shear and compression responsive functions, which regulate mechanotransductive effects on endothelial cell signalling and vascular permeability, which are important in the nutrition and development of tissues [2, 5-9]. In brain, the glycocalyx of microglia and oligodendrocytes contain cell surface sialic acid binding immunoglobulin like lectins (SIGLECS), which identify sialic acids in cell surface glycoconjugates in adjacent neurons facilitating cellular communication and signalling through an intracellular immunoreceptor tyrosine based inhibition motif (ITIM) which maintains a homeostatic balance in
neuronal cells [4]. Embryonic stem cells also assemble a glycocalyx containing cell surface epitopes which not only can be used to identify specific stages of stem cell differentiation but serve as interactive modules that can network with regulatory cues received from the ECM influencing stem cell differentiation [2]. The endothelial and epithelial glycocalyx also have important roles to play in inflammation and immunomodulation [10]. MUC1 (CD227), a high molecular weight (>400 kDa) widely distributed, multifunctional, type-1 membrane tethered epithelial glycoprotein, has roles in dendritic cells, monocytes, T and B cells in immune mediated inflammatory processes [11]. MUC1 in the mucosal lining also provides a protective lubricative barrier to microbial infection [12, 13]. The cerebrovascular glycocalyx also has important roles to play in neural protection; SIGLECS protect neurons from acute toxicity through interaction with glycolipids, which provide barrier functions [14, 15]. The glycocalyx displays brain specific functions through its participation in interactions with cell surface receptors, which undertake protein-phosphorylation mediated signalling by neurons and can also influence apoptosis and amyloid deposition [7]. GAG components in the glycocalyx have important roles in neuroprotection through interactions with CS-receptors and participation in cell signalling events which maintain cellular integrity and also preserve the tissue hydration provided by GAGs to the PCM and ECM.

2. The biodiversity of ECM and cell associated molecules decorated with CS GAG chains.

CS (Fig 1) is composed of β1-3 and β1-4 linked D-glucuronic acid and N-acetyl D-galactosamine repeat disaccharide units which can be O-sulphated at the 2, 4 and C6 position [16]. Furthermore, the D-glucuronic acid moiety may also be epimerised to α-L-Iduronic acid in the related GAG dermatan sulphate (DS) leading to a considerable degree of structural diversity in CS/DS and the ability to interact with a large range of cytokines, chemokines, morphogens and growth factors which regulate cellular proliferation and differentiation and tissue development [16-25]. CS also has indispensable roles to play in stem cell differentiation and the attainment of pluripotency [26].

CS (Fig 1) decorates a remarkably diverse collection of matrix and cell associated macromolecules (Fig 2-4). Their functional properties are summarised in Table 1 and their structural features shown diagrammatically in Figs 2-4. CS occurs as a number of isomeric forms. referred to as CS-A, CS-C, or CS-D based on the mono- or disulphate positions (Figure 1j). CS-B, also known
as dermatan sulphate, like CS-A is sulphated at the C4 position of N-acetyl D-galactosamine, but differs due to epimerization of D-glucuronic acid to L-Iduronic acid and this is sulphated at C2.

2.1 The CS-proteoglycans represent a bio-diverse group of molecules.

2.1.1 The Hy Alecants

The Hyalectans are a group of large HA interactive CS-proteoglycans [27-31] (Fig 2a). Aggrecan and versican form aggregate structures with HA which provide tissues with an ability to act as weight bearing and tension resisting structures while neurocan, brevican and aggrecan hyalectans form perineural-net structures through interaction with HA and tenascin-R (Fig 5a). Perineuronal nets have neuroprotective roles [32, 33] but also inhibit neuronal repair processes by inhibiting neurite outgrowth, both of these functions are due to the particular GAGs, which decorate these proteoglycans. Similar network structures between aggrecan, link protein and hyaluronan are also prominently featured in cartilaginous tissues where they have roles in weight bearing (Fig 5b).

2.1.2 The SLRPs

The small leucine rich proteoglycan (SLRP) family have well known functional roles in the regulation of collagen fibrillogenesis but have additional cell regulatory roles through their interactive properties with cytokines, growth factors and morphogens [34-38] Decorin and biglycan are two SLRPs which contain one or two CS chains and in specific contexts DS (Fig 2b).

2.1.3 SPACRCAN

SPACRCAN is a novel 400 kDa CS-proteoglycan of the inter-photoreceptor ECM providing an interface between the photoreceptors and the pigmented retinal epithelium in the fundus of the eye [39]. SPACRCAN contains 6-sulphated CS chains and a number of N- and O-linked oligosaccharides which collectively constitute ~60 % of its total mass (Fig 2c). SPACRCAN has two RHAMM like binding domains through which it interacts with HA to form an aggregate structure which organises the ECM and is also important in the hydration of this tissue [40].

2.1.4 Perlec an

Perlec an is a HS-proteoglycan in vascular tissues however chondrocytes and smooth muscle cells synthesise a hybrid form of perlec an where CS chains replace some of its HS chains [41-46]. Epithelial perlec an is also a hybrid perlec an and a unique proteoglycan containing HS, CS and KS chains [47]. The form of perlec an synthesised by foetal IVD progenitor cells contains 7-D-4 CS sulphation motifs [48]. The GAG side chains of perlec an in growth plate cartilage contain
embedded 4, 6-disulphated CS-E disaccharides that direct collagen fibrillogenesis[49] and are also found in the brain proteoglycan, appican interacting with neuroregulatory factors, which direct neuritogenesis [50-52].

2.1.5 Appican

Over sulphated disaccharides of CS-D and CS-E (Fig 1j) regulate neuronal adhesion, cell migration, and neurite outgrowth in the CNS. Several brain CS-proteoglycans including phosphacan (DSD-1) and bikunin contain embedded CS-D motifs within their CS side chains. Appican is the only brain proteoglycan identified, with embedded CS-E [51]. Appican is produced exclusively by astrocytes, which regulate neural cell adhesion and outgrowth. Appican also contains Alzheimer amyloid precursor protein (APP) as a core protein component, which is a Kunitz protease inhibitor/Protease nexin 2 domain [50] with voltage gated ion channel blocking properties relevant to neurite regulation [53]. The CS-E motif is essential for the interaction of the appican CS-chain with growth/differentiation factors, and the regulation of neuronal cell adhesion, migration and neurite outgrowth.

2.1.6 NG2/CSPG4

Chondroitin sulphate proteoglycan-4 (CSPG-4) (Fig 3a), also known as high molecular weight melanoma associated antigen in humans and nerve-glial antigen-2 (NG2) in rodents is a transmembrane CS-proteoglycan expressed by immature progenitor cells including oligodendrocyte, chondroblasts/osteoblasts, myofibroblasts, smooth muscle cells, pericytes, interfollicular epidermal and hair follicle cells [54, 55]. CSPG4 is a single pass type 1 transmembrane protein, occurring as a 250kDa glycoprotein, a 450kDa C4S-proteoglycan or can be non-glycanated [56, 57]. The CS side chain of CSPG-4 facilitates interactions with α4β1 integrin and fibronectin and has roles in the activation of proMMP2 by transmembrane MMPs. This ability to influence integrin and MMP activation implicates CSPG-4 in melanoma migration and invasion in skin [58, 59]. CSPG-4 may participate in cell signalling as a co-receptor or by association with cytoplasmic kinases such as FAK or ERK-1, 2. CSPG-4 binds FGF-1 and PDGF AA and presents these to their cognate receptors to influence cellular proliferation and differentiation [60]. The central non-globular domain of CSPG-4 binds to collagen V and VI [61, 62] facilitating cellular attachment and ECM stabilisation and may induce cytoskeletal reorganisation conducive to cell spreading and migration [63, 64]. In NG2 knockout mice the epidermis is very thin due to reduced
basal keratinocyte proliferation providing clues as to the likely role of this proteoglycan in skin development and homeostasis and insightful as to its possible roles in melanoma spread [54, 57].

2.1.7 Thrombomodulin

Thrombomodulin (TM-β, CD141) is a multifunctional 74-105 kDa cell-surface CS-proteoglycan mediator of endothelial anticoagulant activity, activator of Protein-C and a thrombin receptor (Fig 3b). The presence of CS on TM-β decreases the Kd for thrombin binding and significantly accelerates thrombin inhibition [65, 66]. The C-4-S chains on TM are relatively small (10-12 kDa) [67, 68] but essential for its anticoagulant activities [69]. TM acts as an anticoagulant protein through its actions on thrombin and by participating in the generation of activated protein C (APC) [66]. Once APC is formed it binds to protein-S on the cell surface and the APC-protein-S complex inactivates factors Va and VIIIa. [70-72]. TM’s domain structure and multi-component interactions with thrombin, Protein-C, Thrombin-Activatable Fibrinolysis Inhibitor, Complement, LewisX antigen, and HMGB1, a chromosomal protein which regulates transcriptional replication, facilitates TM’s physiologically significant anti-inflammatory, anti-coagulant, and anti-fibrinolytic properties [73, 74].

2.1.8 Phosphacan

Receptor-type protein tyrosine phosphatase beta (RPTP-β) is a transmembrane CS-proteoglycan expressed in the developing nervous system and contains an extracellular carbonic anhydrase (CAH) and fibronectin type III repeat domain, both of these domains foster protein-protein interactions. RPTP is expressed in 3 alternatively spliced forms RPTP-γ, RPTP-βζ, and a truncated form of RPTP-β with an 860 amino acid deletion (Fig 3c). Phosphacan is the proteolytically released ecto-domain of the transmembrane protein tyrosine phosphatase receptor-ζ of neurons and glial cells [75] and is a principal CNS proteoglycan promoting neuron-glial interactions, neuronal differentiation, myelination and axonal repair. The transient nature of cell signalling by phosphorylation requires specific phosphatases for regulatory control. Phosphorylation of tyrosine residues in cellular proteins plays an important role in the control of cell growth and differentiation in the brain [76-79]. The complexity of this regulatory system is evident in the spectrum and widespread distribution of spatially and temporally expressed protein tyrosine phosphatases. The CAH domain of RPTP-βζ promotes protein-protein
recognition, induces cell adhesion and neurite outgrowth of primary neurons, and differentiation of neuroblastoma cells. Interaction of phosphacan with contactin may generate unidirectional or bidirectional signals which direct neural development and axonal repair [80].

2.1.9 The Syndecan family

The GAG side chains of the syndecan proteoglycans provide subtle variation in their binding properties with ligands (Fig 3d). The core protein of the syndecans have a protease sensitive site close to the transmembrane attachment region, its cleavage results in the release of a soluble ecto-domain form of these proteoglycans. Although widely categorised as HS-proteoglycans, syndecan-1, 3 and 4 can also be substituted with CS chains [81]. HS chains have an invariant structure between syndecan family members however their CS chains may contain non-sulphated, 4-O-, 6-O-, and 4,6-O-disulfated N-acetylgalactosamine-CS-E. The CS chains of syndecan-4 generally display a greater overall sulphation level than the CS chains in syndecan-1 [82, 83]. The HS and CS chains of syndecan-1 and 4 bind FGF-2, midkine (MK) and pleiotrophin (PTN). The HS and CS side chains of syndecan-4 are found localised with integrins in focal adhesions in fibroblasts indicating that they have roles in cellular attachments and promote cellular migration [84] and may also influence cell signalling.

2.1.10 CD44

CD44 is the major HA receptor in the human body and is a ubiquitously distributed cell surface receptor (Fig 3e). CD44 can also occur as a part-time proteoglycan called epican, which is substituted with HS or CS chains. Epican is expressed by keratinocytes and mediates cell adhesive properties between keratinocytes in the epidermis [85, 86].

2.1.11 Bikunin

Bikunin is a 30-39 kDa serum proteinase inhibitor synthesized in the liver and is a member of the inter-α-trypsin inhibitor (220 kDa) (ITI) and pre-α-trypsin inhibitor (125 kDa) (Pre-α-TI) families [87, 88] (Fig 4a). A retrospective assessment of the Kunitz serine proteinase inhibitory proteins present in ovine articular cartilage, meniscus and intervertebral disc indicated that the 250, 120, 86, 58, 34-36 and 6-12 kDa SPIs in these tissues were related to ITI and pre-α-TI [89, 90]. Bikunin’s CS chains contain regions which are sulphated and non-sulphated, the sulphated region contain embedded CS-D disaccharides [91]. The CS chain in bikunin is relatively small but heterogenous.
Bikunin inhibits trypsin, thrombin, chymotrypsin, kallikrein, plasmin, elastase, cathepsins, Factors IXa, Xa, Xla, XIIa inhibitory activity and contains two 6 kDa Kunitz inhibitory domains. Bikunin counters inflammatory processes during a number of physiological processes and also has anti-tumour and anti-viral and neuroregulatory activities.

2.1.12 Type IX collagen

Type IX collagen [94, 95] contains a CS chain attached to the α2-chain of the type IX NC3 domain [96-98] and is the PG-Lt proteoglycan isolated from chick embryonic tibia and femur [99] and chick embryo sternal cartilage [100, 101] (Fig 4b). CS-substituted type IX collagen has also been isolated from chondrosarcoma [102] but is present as a minor glycanated form in articular cartilage [103]. The Type IX collagen of chick vitreous humour contains an extraordinarily large CS chain of 350kDa in size [104-107]. The related type XII [108, 109] and XIV collagen [110] which are basement membrane components, also bear CS chains and homology to type IX collagen.

2.1.13 Testican

The SPOCK gene encodes the protein core of a seminal plasma testican proteoglycan containing CS- and HS chains (Fig 4c). This protein's function is unknown, although similarity to thyropin-type cysteine protease-inhibitors suggests its function may be related to protease inhibition. Testican-1 inhibits cathepsin-L [111]. Testican-2, 3 also regulates MMP activation at the cell surface abrogating MT1 MMP activity and proMMP-2 processing [112, 113]. Testican is produced by endothelial cells [114] and has a widespread distribution, the brain is a particularly rich source of testican [115].

2.1.14 Serglycin

Serglycin is the only intracellular proteoglycan so far identified. Serglycin localizes to the α-secretory granules of platelets and mast cells, where it binds and regulates the activity of platelet factor-4 in platelets or tryptase and chymase in mast cells [116-119]. Serglycin is decorated with CS chains in the secretory granules of circulating basophils, but with heparin in resident tissue mast cells [116, 120]. Mast cell serglycin displays a 2-B-6 (-) epitope on the CS chains which decorate this proteoglycan [121]. Trypstatin, the Kunitz protease inhibitor domain 2 of bikunin/ITI is localized complexed with serine proteases in the α–granules of mast cells.
2.1.15 Colony stimulating factor

Human monocytes secrete two CS-proteoglycan forms of colony stimulating factor (CSF) containing two CS chains attached at the C-terminus of CSF-1 [122, 123].

2.1.16 Leprecan

Leprecan is a basement membrane CS-proteoglycan; it contains an N-terminal leucine and proline rich domain, a C-terminal globular domain containing two CS chains and a 2-oxoglutarate-Fe dependant dioxygenase and prolyl-3-hydroxylase enzymatic activity [124]. Prolyl hydroxylases 1-3 (PHD1-3) are oxygen-sensing enzymes which catalyse the hydroxylation of conserved prolyl residues in the HIF-1α sub-unit in normoxia targeting it for proteasomal degradation. HIF-1α and NF-κB are stabilised in hypoxia regulating a diverse range of ~200 genes in erythropoiesis, angiogenesis, cardiovascular function, inflammation, apoptosis and cellular metabolism [125-128].

2.1.17 Identification of neuroendocrine Pro-hormones as CS-Proteoglycans.

A number of CS-DS pro-hormones have been identified using a proteomics screen involving isolation by anion exchange chromatography, pre-digestion of the isolated anionic proteins with chondroitinase ABC and identification of the CS linkage tetrasaccharide by mass spectrometry. Many of these pro-hormones are stored in intracellular granules in neuroendocrine cells. Granule proteins such as Chromogranin-A are processed into hormone peptides such as secretogranin-1, 2, 3, cholecystokinin or neuropeptide W.

3. Aggregated proteoglycan structures in cartilage and brain.

Members of the hyalectan proteoglycan family including aggrecan, brevican, neurocan and versican form massive supramolecular perineural net structures in the CNS (Fig 5a) through interactions with HA and tenascin-R. Perineural nets protect neurons from oxidative stress and mechanical damage but also provide inhibitory signals preventing neural outgrowth. Aggrecan and versican also form massive supramolecular aggregate structures by interaction with HA and link protein in cartilage and fibrocartilaginous tissues (Fig 5b). These aggregated structures have impressive water regain properties equipping cartilage and IVD with the ability to withstand compressive forces.

4. CS sulphation motifs as molecular markers of tissue morphogenesis
During tissue morphogenesis several proteoglycans contain native 4-C-3, 7-D-4, 3-B-3[-] and 6-C-3 sulphation motifs (see Fig 1 for explanation). These native CS sulphation motifs are expressed in the surface regions of developing articular cartilages in the knee joint (Fig 6 a, b) perichondrial growth plate, and in vascular ingrowth and stromal vascular niches of transitional tissues associated with diarthrodial joint and IVD development.[129-132]. Confocal colocalisation of the aforementioned CS sulphation motifs with aggrecan, versican and perlecan in neonatal cartilages has demonstrated that aggrecan and perlecan in these tissues bear these sulphation motifs while versican does not [133]. Confocal studies in the human foetal elbow also demonstrated perlecan associated with perichondrial stem cell niches (Fig 7a) and with progenitor cell populations in the perichondrium (Fig 7c, g) and surface regions of the developing elbow joint cartilages (Fig 7c, h, i).

4.1 Cartilage proteoglycans containing CS sulphation motifs with roles in tissue development.

Aggrecan is the major CS-proteoglycan of cartilaginous tissues, with well-known space-filling and water imbibing properties that equip these tissues with resilience to compressive loading. Correct sulphation of CS-proteoglycans is essential for proper Indian hedgehog signalling in the developing growth plate [134], perlecan, a hybrid CS-HS proteoglycan in cartilage is also responsible for the localization and activity of the related Sonic hedgehog protein [135]. Native CS sulphation motifs such as 7-D-4 on proteoglycans may serve to immobilise growth factors/morphogens actively involved in tissue development [17]. The unique distributions of native CS motifs such as 7-D-4 with surface zone progenitor cells in articular cartilage [132, 136, 137] and within the developmental intervertebral disc (IVD) [129] and human foetal elbow [130] is suggestive of an early stage of progenitor cell differentiation and indicates that native CS sulphation motifs have functional roles in chondrogenesis and in IVD development [129, 136, 137].

4.2 Focal expression of the 7-D-4 CS sulphation motif in human foetal paraspinal blood vessels.

Perlecan produced by endothelial and smooth muscle cells is a prominent component of capillaries and larger blood vessels (Fig 8a, b). The 7-D-4 CS sulphation motif displays a focal distribution in the lumenal surfaces of capillaries and between the endothelial cells lining human foetal paraspinal blood vessels (Fig 8b). Pericytes are contractile cells that wrap around the abluminal surface of endothelial cells that line the capillaries and venules throughout the body (Fig
Caplan proposed that all stem cells were pericytes emphasising their vascular origins [138-141]. Pericytes are embedded in basement membrane where they communicate with endothelial cells of the body's smallest blood vessels by means of both direct physical contact and paracrine signalling [142-145]. Blood flow generated shear forces are also important functional determinants of the differentiation of stem cells in the luminal surfaces of blood vessels [8, 146].

4.3 The tissue distribution and function of oversulphated CS isomers CS-D and CS-E

Developmental studies on the whole rat brain have correlated changes in the CS side chain structure of phosphacan with measurable changes in the binding affinity of PTN and functional consequences on the cell signalling response. Phosphacan isolated from whole rat brain from various developmental stages was examined using the CS antibodies MO225, CS56 and 2H6 in a plasmon resonance study [147]. P7-phosphacan strongly reacted with CS56 and 2H6 but not MO225. P12-phosphacan showed moderate reactivity with CS56 and 2H6 but no reactivity to MO225 contrasting with P20-phosphacan which was strongly reactive with MO225 but low reactivity with CS56 and 2H6. mAb 2H6 is sold as an anti-CS-A Ab due to its high reactivity with whale cartilage CS-A however its reactivity with phosphacan of a defined CS-A content does not correlate with this. P7 phosphacan with a CS-A content of 64% had the highest reactivity with mAb 2H6 while P20 phosphacan with a CS-A content of 86% had very low reactivity. This showed that the 2H6 epitope was not to a simple CS-A unit but to a more extended binding epitope. Subsequent studies have shown that mAb CS-56 and MO-225 specifically recognize octasaccharides containing an A-D tetrasaccharide sequence, whereas 2H6 preferred sequences with A- and C-units such as C-C-A-C for strong binding but no D-unit, mAb MO225 also recognised the CS-E disaccharide motif from squid cartilage in an extended E-E-E-E-C binding motif [148, 149]. The development of CS oligosaccharide libraries [150] of defined structure has further enhanced the precision of such structure-function studies. These show that the CS motifs are differentially regulated in brain development and modulation in CS structure occurs in a spatiotemporal manner.

5. Cell regulatory proteoglycans are involved in neural development and repair.

Oversulphated CS/DS promotes neural development with variation in sulphation profiles of proteoglycans regulating vertebrate CNS development. The disulphated disaccharide D-unit promotes neurite outgrowth through the DSD-1 epitope embedded in the CS chains of DSD-1-PG/phosphacan [150-155]. Oversulphated DS displays neurite outgrowth activity [156]. The short
isoform, non-proteoglycan variant form of phosphacan/receptor protein tyrosine phosphatase-β also interacts with neuronal receptors and promotes neurite outgrowth [80]. Bikunin is also expressed in brain tissue [157, 158] and accumulates in brain tumours [159]. Like phosphacan, bikunin contains disulphated embedded CS-D motifs within the repeating disaccharide region of its CS chain [91]. Such motifs promote neurite outgrowth, suggesting that bikunin may also have similar roles to play in neural development. Appican is another brain CS-proteoglycan [50, 160] produced by astrocytes [161], which direct neural development. CS-E motifs embedded within the CS chains of appican [51] interact with neuroregulatory factors [52] inducing morphological change in C6 glioma cells and directed adhesion of neural cells to the ECM [53]. CS-E motifs also promote chondrocytic differentiation of ATDC5 cells. ATDC5 cells produce monosulphated CS-A or disulphated disaccharides (CS-E) in their ECM proteoglycans. Exogenously added CS-E also affects chondrogenic differentiation of ATDC5 cells, promoting chondrogenic differentiation demonstrating the existence of cell surface receptors for CS-E [162]. Embedded CS-E in the CS side chains of growth plate perlecan also promote collagen fibrillogenesis [49].

NG2 proteoglycan, phosphacan and syndecan-1-4 have roles in cellular regulation and tissue development, which may be of application in tissue repair strategies. For example NG2 proteoglycan stimulates endothelial cell proliferation and promotes migration during micro-vascular morphogenesis. NG2 is also expressed by chondroblasts and chondrocytes and acts as a cell surface α2-VI collagen receptor conferring cellular motility and α4β1 integrin mediated cell spreading by activation of FAK and ERK1/ERK2 signalling cascades. A better understanding of the CS-sulphation motifs and their binding partners and how these regulate cellular processes in tissue remodelling and repair may allow the development of improved therapeutic procedures in repair biology.

5.1 The balance between stimulatory and inhibitory signals in neural development

SRPX2 (Sushi repeat protein, X linked 2) (Fig 2a) is a novel secreted CS-proteoglycan, which promotes synaptogenesis in the cerebral cortex and is found as an embedded domain in some members of the lectican proteoglycan family. The SPRX2 gene is a target of the foxhead box protein P2 transcription factor (FoxP2) that modulates synapse formation [163]. Mutations in SRPX2 causes Rolandic epilepsy and speech impairment (RESDX syndrome). Interactome/cell surface binding/plasmon resonance studies have identified SRPX2 as a ligand for uPAR, the urokinase type
plasminogen activator (uPA) receptor [164]. uPAR knockout mice exhibit an enhanced susceptibility to epileptic seizures and anomalous cortical organization consistent with altered neuronal migration during brain development [165, 166]. uPAR is a crucial component of the extracellular plasminogen-plasmin system, which remodels the ECM during brain development. Cathepsin B and ADAMTS4 are also SRPX2 ligands and also likely participants in developmental processes in the brain [167]. ADAMTS-4 has been localised to regions of the spinal cord undergoing repair. ADAMTS-4 degrades aggrecan and versican in the CNS thus removing the inhibitory signals provided by the CS side chains of these PGs [168, 169].

Cathepsin B is a well known activator of pro uPA thus SRPX2 and its ligands represent a network of proteins with critical roles in brain development and specifically in the centres of speech and cognitive learning. The Rolandic and Sylvian fissures bisect the human cerebral hemispheres and it is the adjacent areas of the brain, which are responsible for speech processing. Ordered neuronal migration is therefore essential for the correct development of these areas of the brain. SRPX2 protein expression occurs in neurons from birth and has central roles to play in developmental processes in the centres of speech and cognitive learning. Two mutations have been identified in SRPX2 in RESDX patients. One mutation (N327S) results in altered glycosylation while a second mutation (Y72S) affects the first sushi domain of SRPX2 [170], 3D modelling indicates that the Y72S mutation affects an area of the SRPX2 core protein normally involved in protein-protein interactions [171]. Cultured cells from RESDX patients display alterations in the intracellular processing of proteins and likely misfolding which may have functional consequences [172].

Specific CS sulphation motifs are involved in interactions between neurons and glial cells to regulate the development and regeneration of the CNS. Migrating neurons are guided by glial cells through ECM proteoglycans they assemble such as phosphacan, and the CS-lectican proteoglycan family. Phosphacan promotes neurite outgrowth whereas versican, neurocan, and brevican inhibit this process thus collectively these proteoglycans direct neurite growth. This is a function of their differing GAG CS sulphation motifs. CS-D motifs in phosphacan promote neurite outgrowth while lectican CS-A and CS-C motifs inhibits neuronal migration [152] and regulate neural tissue morphogenesis [153] [150, 173]. Glucuronyl transferase-1 knockout embryonic stem cells (ESCs) lack CS resulting in a significantly altered ability to differentiate and reduced ability to develop into pluripotent cell lineages [26]. HS maintains ESCs in a state primed for differentiation
however CS maintains ESC pluripotency and promotes ESC differentiation. Binding of CS-A and CS-E to E-cadherin to overcomes cell inhibitory inhibitory signals enhances ESC differentiation[26]. The highly charged CS-D and CS-E sulphate motifs can mimic HS in terms of growth factor and cytokine binding however the less highly charged CS-A and CS-C isomers should not be discounted in such interactions. Surface plasmon resonance studies have demonstrated CS-A and CS-C bind with significant affinity to midkine, pleiotrophin, HGF and stromal cell derived factor-1β but with a lower affinity than CS-D, CS-E and HS [174] regulating the growth, differentiation and migration of neural precursor cells [175]. Such lower affinity interactions may provide a more subtle control mechanism than the strong on-off signals supplied by HS. FGF-2 and EGF dependant proliferation of glial cells regulates neurogenesis during CNS development [176-178]. Chondroitin-6-sulphate synthesis is upregulated in the injured CNS, induced by injury related cytokines and enhanced in axon-growth inhibitory glia [179] and of relevance to nerve regeneration through glial scar formations [180].

6. CS sulphation motifs and pathological remodelling of connective tissues

Several years ago [181, 182] it was noted that mAbs 3-B-3 (-) and 7-D-4 identified chondrocyte “cell-clusters” in pathological (osteoarthritic) canine and human articular cartilage and at that time these were considered a classical feature of the onset of late stage degenerative joint disease. In these early publications a lack of knowledge of stem/progenitor cells in cartilage and expression of proteoglycans (aggrecan) with CS GAG chains recognised by mAbs 3-B-3 (-) and 7-D-4 were interpreted to indicate a failed, late-stage, attempt to repair cartilage and replacement of new proteoglycans in a matrix that had been extensively degraded by MMPs. An alternative hypothesis now is that these ‘chondrocyte clusters’ arose from adult stem/progenitor cells in these tissues [183]. Tesche and Miosge [184, 185] showed that adult stem cell clusters were surrounded by a pericellular matrix containing perlecan. This is also a feature of stem cell niches in foetal knee, hip, IVD and elbow cartilage [129-133, 186]. It is expected that in different connective tissues, the CS sulphation motifs will be present on different matrix and cell surface proteoglycans, if this is the case, an important feature of the stem/progenitor cell niche may be the sulphation of the GAGs rather than the core proteins to which they are attached. Expression of different levels of GAG sulphotransferases in stem/progenitor cells would therefore also contribute to tissue repair [187-189]. More recently, cell clusters within the superficial zone of healthy articular cartilage have been
shown positive for both Notch 1 and CD166 [190], cell surface markers that are synonymous with the stem cell niche environment [136, 191].

6.1 CS expression and tumour development

Neuroendocrine tumours with different degrees of histological differentiation have correlative alterations in associated CS but little change in HS. Normal stroma contains no staining with anti-CS Abs while staining in tumour is significantly elevated and highest in advanced tumour grades [192]. CS-proteoglycan levels are elevated in liver cancer [193], renal [194], hepatocellular [195, 196] and gastric carcinoma [197], pancreatic cancer [198] and in mammary tumours [199]. In gastric and pancreatic cancer non-sulphated and 6-sulphated CS predominate over other GAG isoforms and the GAG chains display a smaller hydrodynamic size than normal tissues. Elevated levels of 4- and 6-sulphated CS are found in renal cancer. Decorin and versican levels are elevated 7-27 fold in pancreatic cancer, and contain non-sulphated and 6-sulphated CS. This contrasts with the normal pancreas where DS is the predominant GAG decorating versican and decorin core proteins. Tumour proteoglycans have altered interactive properties further impairing the normal functional properties of tumour affected tissues. Proteoglycans and GAGs modulate cellular processes relevant to all stages of tumour progression, including cell proliferation, cell-matrix interaction, cell motility and invasive growth. HS, CS/DS and HA all have well documented roles in tumour pathobiology [200]. CS is abundantly present in the ECM in ovarian cancer. Alterations in the sulphation of CS also influences cancer development and its aggressive status. The CHST15 gene is responsible for the biosynthesis of highly sulphated CS-E [269]. The single chain phage Ab GD3A11 to highly sulphated CS facilitates identification of biomarkers in aggressive tumour development. The GD3A11 epitope is minimally expressed in normal tissues but is intensely expressed in a number of ovarian cancer sub-types but not in benign ovarian tumours [201]. Serum over-sulphated CS levels measured using mAb WF-6 are also elevated in ovarian cancer and may also be a useful biomarker [202]. Silencing of CHST15 in-vitro and in a xenograft model of pancreatic cancer down-regulates tumour invasion in pancreatic ductal adenocarcinoma (PDAC). CS-E is detected in both tumour and stromal cells in PDAC and is considered to have multistage involvement in its development [203]. A single intra-tumoural injection of CHST15 siRNA almost completely silenced tumour growth providing evidence of the direct involvement of CHST15 in the proliferation of pancreatic tumour cells identifying a novel therapeutic target. The phage display antibody GD3G7 also reacts with the rare CS-E and DS-E epitopes in normal tissues, where DS-E epitope represents IdoUA-GalNAc (4,6-O-disulphate) [368]. CS-E is strongly up-regulated in
ovarian adenocarcinomas. Thus Ab GD3G7 is useful in defining tumour tissue alterations [204]. Quantitation of GAGs in colorectal tumour tissue using electrospray ionization mass spectrometry showed that neoplastic tissues displayed greater levels of CS and DS than non-neoplastic tissue where HS was decreased [205]. NEDD9 (CAS-L, HEF-1) cells have key roles in the migration and proliferation of MDA-MB-231 breast cancer cells [206]. Microarray studies in breast cancer samples demonstrate elevated CD44 and serglycin and down-regulation of syndecan-1, syndecan-2 and versican whereas CHST11, CHST15 and CSGALNACT1 were all up-regulated in NEDD9 cells, an increase in CS-E attached to CD-44 was also evident in tumour cells. Removal of CS using chondroitinase ABC inhibited colony formation by NEDD9 cells whereas exogenous application of CS-E enhanced NEDD9 cell proliferation and tumour development clearly demonstrating roles for CS-E in tumourogenesis [206]. A number of tumour cells express GAGs with alterations in sulphation level. Altered expression of CS and HS on tumour cells has a key role to play in malignant transformation and tumour metastasis [207, 208]. Receptor for Advanced Glycation End products (RAGE) is a receptor for CS-E in Lewis lung carcinoma (LLC) cells [273, 278]. RAGE binds strongly to CS-E and HS and to LLC cells and has roles in tumour development. Serglycin is the major proteoglycan produced by multiple myeloma (MM) cells. Knockdown of serglycin dramatically attenuated MM tumour growth [209]. Tumours, which develop in serglycin knockdown animals, display lower levels of HGF and reduced blood vessel development indicating that serglycin has roles in angiogenesis. The CS chains on serglycin are at least partly responsible for cellular attachment to CD44. Serglycin was originally considered to be a product of haematopoietic cells, recent studies have shown that it is also synthesized by a number of non-haematopoietic cells [210]. Serglycin is expressed by tumour cells, promotes an aggressive phenotype and confers resistance to drugs and inactivation by the complement system. Serglycin promotes inflammatory conditions through inflammatory mediators, which are normally complexed in intracellular granules thereby contributing to tumour development. The CS-lectican proteoglycan versican accumulates in the tumour stroma and has key roles in malignant transformation and tumour progression. Elevated expression of versican in malignant tumours is associated with cancer relapse and poor clinical outcome in prostate, breast and many other cancer types. Versican (so named from ‘versatile’) regulates cell adhesion, proliferation, apoptosis, migration, angiogenesis, cell invasion and metastasis. These processes involve interactions mediated by the CS and DS chains of versican and its G1 and G3 globular domains. Versican therefore represents a logical therapeutic target in tumour pathobiology [211]. Versican G3 domain regulates neurite growth
CD44 regulates apoptosis in chronic lymphocytic leukaemia (CLL) and its expression is mediated by the tumour microenvironment. Interaction of CD44 with HA and CS protects CLL cells from apoptosis. Specific antibodies to CD44 (IM7, A3D8) impair the viability of CLL cells and represent a potential therapeutic target [216]. CS chains in the microenvironment of breast cancer cells have been suggested as appropriate molecular therapeutic targets given that they promote many aspects of carcinogenesis in-vitro [217]. Dramatically elevated CS levels have been observed in the stromal microenvironment of many solid tumours. Intra-tumoural injection of chondroitinase ABC was ineffective in promoting primary tumour regression but led to development of secondary tumours indicating that the CS chains associated with primary tumours had a metastasis inhibitory role exploitable in therapeutic interventions. Cell surface HS and CS chains have roles in the infective stages of viral mediated carcinomas such as Merkel cell carcinoma, a highly lethal but rare form of skin cancer. HS and CS chains act as cellular receptors in the infective stages of Merkel cell poliomavirus. Modulation or removal of such CS or HS entry points may provide an approach to combat viral attachment to cells during these initial infective stages [218]. CS on the surface of breast cancer cells function as P-selectin ligands. CHST11 and CSPG4 are highly expressed in aggressive breast cancer cell lines and correlate with P-selectin binding levels. The CS chains of CSPG4 facilitate binding of P-selectin to highly metastatic breast cancer cell lines. Targeting of CS and its biosynthesis represents an attractive approach in anti-metastatic therapy [219, 220]. Therapeutic targeting of CSPG4 has been used to specifically target myeloma tumour cells using mAb based therapies [58, 221]. Adoptive transfer of genetically modified T cells is emerging as a powerful anti-cancer biotherapeutic. CSPG4 is an attractive target molecule in this approach due to its high expression in cancer cells in several types of human malignancies but restricted distribution in normal tissues [222] and helps to minimise any potential toxic side effects using such approaches. T-cells expressing a CSPG4-specific chimeric antigen receptor offer the possibility of targeting a broad spectrum of solid tumours for which no curative treatment is currently available [222-224].

The treatment of rhabdomyosarcoma (RMS) remains particularly challenging, with metastatic and alveolar RMS offering a particularly poor prognosis. CSPG4 is specifically expressed on RMS cells. The immunotoxin αMCSP-ETA', specifically recognizes CSPG4 on the RMS cell lines RD, FL-OH1, TE-671 and Rh30 and is internalized rapidly, induces apoptosis and kills RMS cells.
selectively. Preliminary studies have demonstrated promising results with the specific binding of this immunotoxin to RMS primary tumours [225]. Deterioration of liver function in liver cancer is accompanied by an increase in the amount of CS-proteoglycans. This alteration in proteoglycan composition interferes with the physiologic function of the liver. Glypicans, agrin, and versican also play significant roles in the development of liver cancer [193]. CS-proteoglycans have essential roles to play in tissue morphogenesis and in cancer development involving interactions with growth factors, morphogens, cytokines, cell surface receptors, and a number of matrix proteins [226].

GAGs play vital roles in every step of tumour progression. Tumour samples with different degrees of histological differentiation demonstrate important alterations in the CS chains of a number of proteoglycans. Immunolocalisations conducted with anti-CS antibodies consistently showed normal stroma was negative whereas tumoural stroma were positive with elevated staining in the higher grade cancer samples while the tumour cells themselves were negative. Syndecan-2 levels were low or undetectable in normal tissues but significantly elevated in endocrine tumours. Glypican-5 was overexpressed in high-grade tumours with epithelial differentiation, but not in neuroendocrine tumours. Normal neuroendocrine cells displayed positive cytoplasmic and membrane staining for glypican-1 but elevated expression in low-grade tumours and reduced in high grade tumours [192]. Use of a therapeutic CSPG4 specific antibody (225.28) enhanced and prolonged the inhibitory response of PLX4032 (Vemurafenib) in combination therapy suggesting Ab 225.28 may be useful as a delivery system in the treatment of melanoma [227].

7. CS sulphation motifs regulate cell behaviour - can they be used to promote tissue repair?

This review has demonstrated pivotal roles for CS-proteoglycans in developmental processes in cell migration, cellular recognition and tissue morphogenesis [1-4]. Novel CS sulphation sequences also occur in the functionally distinct layers of skin [228]; are associated with the long bone growth plates in endochondral ossification and occur at important growth zones in the developing intervertebral disc, diarthrodial joints and tendon [229, 230]. During lymphopoiesis, CS chains are also differentially modified at sites of B-cell differentiation and maturation [229, 231] and in the brain CS sulphation plays an important role in neurite outgrowth, synaptic plasticity and neurological development [232]. Accumulated evidence therefore points to specific GAG sequences in CS having roles in cell interactions and developmental processes. A greater understanding of
these processes through sustained basic research could eventually lead to their use in advanced therapeutic applications in regenerative medicine. With the advances in oligosaccharide synthesis methodology now available, the CS sulphation motifs discussed in this review can be synthesised and CS oligosaccharide microarrays prepared to answer structure-function questions relevant in tissue repair strategies. A number of lipid-derivatized CS oligosaccharides with well-defined sulphation features have been synthesised and used in CS oligosaccharide microarrays to characterise the preferred binding sequences of the anti-CS mAbs 2H6, MO225, 473HD and LY111 [167] and to assess prospective binding partners (growth factors, cytokines) and many of the effects of these CS oligosaccharides on cellular behaviour have also been determined in-vitro. This supports their therapeutic application in tissues such as the brain and CNS, and may lead to the re-establishment of nerve function in glial scar formations.

7.1 Development of smart CS-bioscaffolds to improve tissue repair

The development of CS-bioscaffolds and their applications with stem cells in cartilage, bone, cornea, skin and nerve repair strategies represent a significant advance in bioscaffold design and performance in tissue repair strategies. CS has indispensable roles to play in stem cell differentiation and attainment of pluripotency [26]. Accumulated evidence points to CS sulphation motifs having critical roles in cell interactions, cell differentiation, proliferation and matrix assembly. A greater understanding of these glyco-code mediated processes could lead to improved repair biology therapeutics. Cartilage is a particularly difficult tissue to repair and many biomatrices have been developed in order to perfect an effective repair strategy [233], these have focussed on MSCs as a therapeutic cell type. CS-bioscaffolds promote proliferation of bone marrow stromal MSCs and their differentiation to a chondrogenic phenotype appropriate for cartilage repair. Combinations of CS, gelatin, chitosan, HA incorporated into polyvinyl alcohol (PVA), polylactic-co-glycolic acid (PLGA) hydrogels [233-237] have been developed. A thermoresponsive photopolymerizable CS hydrogel has been used to prepare a chondrocyte matrix suitable for 3D printing [234]. CS tethered on silk fibroin, silk-gelatin-CS-HA biocomposites or CS biomimetic scaffolds [238] have proved suitable for induction of a chondrogenic phenotype in MSCs [239, 240]. Porous CS-alginate foams and chitosan-gelatin-C6S-HA cryogels promote the chondrogenic differentiation of MSCs [241-243] as do CS-HA-Silk-lentiviral inserted TGF-β3 gene, HA-CS-Heparin-Collagen scaffolds, and multilayered 3D CS chitosan constructs [244-246]. Atellocollagen-CS, collagen-CS-HA 3D hydrogels, cross-linked type II collagen-CS scaffolds [247-
PLGA-gelatin-CS-HA-TGF-β3 and elastic copolymer-CS-TGF-β3 scaffolds provide superior induction of chondrogenic cells from seeded MSCs [250, 251].

Fibrous tissue formed in response to implanted materials has been shown to contain CS [252], with increased infiltration of inflammatory mast cells. The mast cell proteoglycans serglycin and perlecan display a 2-B-6 (-) epitope on their CS chains [121]. Like 3-B-3 (-), 2-B-6 (-) is not generated by chondroitinase ABC digestion. The presence of this 2-B-6 (-) epitope has previously been reported in osteoarthritic cartilage, however how the epitope is generated, or its function remain to be established [253]. Generation of this epitope is due to the action of a member of the hyaluronidase (HYAL) family, HYAL-1 or HYAL-4, depolymerise CS via a hydrolytic cleavage reaction at the β1→4 disaccharide glycosidic linkage [254, 255]. HYAL-1 or HYAL-4 may also generate the 3-B-3 (-) ‘native’ CS epitope. The 2-B-6 (-) epitope, and production of HYAL-4 by mast cells are both associated with tissue remodelling and repair in inflammatory conditions.

7.2 Cell surface CS-receptors and CS interactive molecules that control cellular behaviour

Only a few cell interactive oligosaccharide sequences in CS-DS have so far been identified due to inherent difficulties in decoding their complicated structures. CS-DS hexasaccharide and octasaccharide motifs, which facilitate interactions with heparin cofactor-II and pleiotrophin, have been determined [256-258]. A major difficulty in the identification of these interactive CS-DS modules is due to them not being a well defined saccharide sequence, but rather several heterogeneous modification patterns, the so called ‘wobble CS-DS motifs’ [25]. The nomenclature for CS isoforms CS-A, CS-C, CS-D, CS-E, and DS, is confusing and misleading in that naturally occurring CS-A, for example, is not a homogenous polymer composed of CS-A disaccharide units only, but may contain a mixture of A, C, and unsulphated chondroitin units and embedded CS-D or CS-E motifs within the repeating disaccharide regions of CS-A GAG chains or the D-glucuronic acid can be epimerized to L-iduronic acid [259, 260]. HS has historically been considered to play more important roles in GAG-mediated cellular regulation than CS due to its higher propensity to interact with growth factors, morphogens and ECM components [261]. Recent studies have now demonstrated essential roles for CS-DS in a number of biological processes, especially in events, which regulate CNS development, and in tumourogenesis/metastasis.
The soluble ecto-domain of PTPR β/ζ, phosphacan interacts with the cell surface receptor contactin-1 (Fig 9a). Further cell signalling membrane proteins include the syndecan proteoglycan family (SDC 1-4), CSPG-4, betaglycan or endoglycan, these interact with cell surface receptors such as neuropilin-1, leukocyte common antigen-related phosphatase (LAR) and the related receptor protein tyrosine phosphatase β/ζ (Fig 9b, c). The neuronal Nogo axonal guidance receptor family consist of three GPI-anchored receptors (NgR1, R2, R3) (Fig 9d, e) [262], the semaphorins, and neuropilins are further receptors which interact with CS-ligands (Fig 9f, g) P75 is a transmembrane co-receptor which interacts with a number of receptors including the Nogo-1 receptor (NgR1) and acts as a signal transducer, converting signals initiated upon binding of myelin associated inhibitory proteins (MAIs) or CS-DS GAG to NgR1 converting this into intracellular signals via p75’s cytoplasmic domains and the Ras/MAPK and JNK pathways to inhibit neurite outgrowth.

7.4 E-Cadherin

The cadherins are calcium-dependent type-1 transmembrane proteins which form adherens junctions, binding cells tightly together within tissues and have essential roles to play during embryonic development and critical in the induction of stem cell pluripotency [263-266]. Cell adhesion is mediated by extracellular cadherin domains, intracellular cytoplasmic domains associate with a large number of adaptor and cytoskeletal signalling proteins which constitute the cadherin adhesome [267]. The cadherin membrane-spanning adherens junction proteins have crucial roles in cell–cell contact formation and are also connected to cytoplasmic proteins which regulate signalling pathways and relay information regarding cell interactions to the nucleus [268-272]. E-cadherin and LRP 5/6 interact cooperatively with cell surface CSPGs and frizzled to regulate intracellular signalling through effects mediated by the catenin system which affects actin polymerisation/depolymerisation regulating ERK phosphorylation and cell signalling (Fig 10 a-b). CS-DS contributes to several signalling pathways and biological events [273]. A CS-E isoform binds strongly to Wingless/int-3a (Wnt-3a) and to a number of growth factors, neurotrophic factors, and cytokines in-vitro [274-276] (Fig 10c). Wnt signalling controls a number of developmental processes, tissue renewal and regeneration, and the development of several diseases, particularly cancers [274-276]. Specific arrangement of sulphation motifs on CS-DS chains modulate Wnt signalling and diffusion. Early stages of embryonic stem cell differentiation are promoted or repressed, by CS-E, but not by CS-A through Wnt/β-catenin signalling pathways [275]. The migration of breast cancer cells in-vitro is reduced by CS-E, but not CS-DS [276]. CS-E regulates
type I collagen fibrillogenesis and expression, and is a positive regulator of breast carcinoma, through Wnt signalling [276]. Collectively, these findings provide insights into how cancer development is mediated through CS-E and Wnt/β-catenin signalling. However, it is still unclear what specific sulphation pattern(s) or length of CS-E saccharide is required to activate and regulate such development processes.

7.3 CS-E interactions with RAGE involved in tumour metastasis

The biosynthesis of stromal CS-DS proteoglycans is up-regulated in many tumours causing their accumulation in stromal tissues with attendant effects on tumour progression [277, 278]. The proportion of CS-E disaccharide units in CS-DS chains is elevated in ovarian and pancreatic cancers [212, 275], resulting in alterations in neoplastic growth and cell motility. Tumour cell signalling is also controlled by VEGF and cleavage of CD44 [279, 280]. The stronger expression of disulphated CS-E disaccharide units on CS-DS chains on the surface of metastatic Lewis lung carcinoma (LLC) cells correlates with their invasive properties in lung tissue [281]. In the lung RAGE acts as a receptor for cell surface CS-DS chains containing CS-E units expressed by LLC cells [282] (Fig 11a, b). RAGE recognizes CS-E unit containing decasaccharides [282], these markedly inhibit the pulmonary metastasis of LLC cells [281], most probably by competitive inhibition. Binding of CS-DS ligands to RAGE leads to downline effects on intracellular proteins such as Rap1 and PKC which effect the activation of NFκB and CREB signalling and transcriptional regulation (Fig 11c).

8. Conclusions

While advances in detection methodologies continue to improve the characterization of an ever-expanding repertoire of complex glycans in small amounts of sample, certain unifying principles have emerged with regard to how these entities regulate cellular metabolism. The sulphate motifs within glycosaminoglycans represent a key information storage and transfer medium, which cells can interpret, to effect tissue homeostasis. The glyco-code contained in glycosaminoglycans is an IT system which nature has developed over many hundreds of millions of years of evolution. However, despite the complexity and biodiversity of glycosaminoglycan structures it is the sulphate motifs which are key cell-directive players in the glyco-code and the sophisticated structures to which they are attached may be viewed as molecular scaffolds whereby varied planar orientations or densities of the sulphate groups can be explored to achieve optimal interactions with their respective ligands. Significant inroads have been made in the sequencing of glycosaminoglycans
and encoded sequences linked with biological processes continue to be identified. A greater understanding of this glyco-code will undoubtedly continue to improve our understanding of the development and regulation of connective tissues and may lead to significant improvements in how this information is applied in advanced strategies in repair biology.

Acknowledgements
No funding was received for this study other than the infrastructure provided by The Institute of Bone and Joint Research, Kolling Institute of Medical Research, The University of Sydney.

Disclosures
The authors have no conflicts or financial disclosures to make.

Author Contributions
All authors contributed to the writing of this manuscript, JM co-ordinated the review comments and final content of the manuscript. All authors endorsed the final version of the manuscript.
<table>
<thead>
<tr>
<th>Protein alternative name</th>
<th>Gene</th>
<th>Distribution</th>
<th>Function</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggrecan</td>
<td>CSPG1 ACAN</td>
<td>Widespread in ECM, cartilage, tendon, IVD</td>
<td>Tissue hydration, space-filling, weight bearing proteoglycan in cartilage, IVD, forms protective perineural nets with HA and tenasin-C, has roles in heart development.</td>
<td>[283, 284]</td>
</tr>
<tr>
<td>Versican</td>
<td>CSPG2 VCAN</td>
<td>Widespread in ECM, CNS</td>
<td>So named as a “versatile” proteoglycan based on its ability to promote cell proliferation, differentiation, cell migration in tissue remodelling and in connective tissue morphogenesis</td>
<td>[285-287]</td>
</tr>
<tr>
<td>Neurocan</td>
<td>CSPG3 NCAN</td>
<td>CNS</td>
<td>Brain lecticans HA binding proteoglycans interactive with NCAM, Ng-CAM/L1. Modulate cell binding in CNS development and neurite outgrowth activity in CNS/PNS. Upregulated in glial scars, inhibits astrocyte and neuronal growth, may act antagonistically with other brain PGs to regulate neurogenesis. Primary astrocytes and neural cell lines bind brevican independently of HA, controlling infiltration of axons and dendrites into maturing glomeruli in brain development.</td>
<td>[288, 289]</td>
</tr>
<tr>
<td>Brevican</td>
<td>CSPG7 BCAN</td>
<td>CNS, One of the most abundant brain proteoglycans</td>
<td>Found on the surface of immature oligodendrocyte and chondroblastic progenitor cells. Roles in cell-PCM stabilisation, cellular proliferation, migration, inhibits neurite outgrowth during axonal regeneration. May sequester FGF-2/PDGF. CSPG4 is a collagen VI transduction receptor activating FAK/ERK1/ERK2. Widely expressed by tumour cells and is specifically targeted by therapeutic measures combatting tumourogenesis. up-regulated in spinal cord injury and in chondrogenesis.</td>
<td>[141][56]</td>
</tr>
<tr>
<td>Chondroitin sulphate-4 NG2</td>
<td>CSPG4</td>
<td>Widespread distribution in ECM and CNS with roles in development. Integral transmembrane proteoglycan.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroglycan-C</td>
<td>CSPG5</td>
<td>CNS</td>
<td>Transmembrane CS-proteoglycan bearing an EGF ECM domain, acts as an active growth factor and ligand for ErbB3, sixth member (neuregulin-6) of the neuregulin family. Contains CS-E, CS-C.</td>
<td>[290][291]</td>
</tr>
<tr>
<td>Syndecan-1</td>
<td>SDC1</td>
<td>CNS</td>
<td>CS-E chains found in Syndecan-1. 3. Widely distributed cell surface CS and HS substituted PGs also containing CS-E and CS-C. Midkine interacts with CS-E motif and participates in neural development and repair but interacts weakly with CS-A and CS-C.</td>
<td>[82, 292, 293]</td>
</tr>
<tr>
<td>Syndecan-4</td>
<td>SDC4</td>
<td>Widespread distribution in vascular, epithelial and weight bearing connective tissues and brain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphacan</td>
<td>PTPRZ1</td>
<td>CS-KS-HNK-1 proteoglycan, known as receptor-type tyrosine-protein phosphatase zeta (PTPR-ζ) single pass type I membrane protein with cytoplasmic tyrosine protein phosphatase, carbonic anhydride and fibronectin type III domains. Alternative splice forms exist. DSD-1 is the mouse homologue. Roles in embryonic spinal cord development/neurogenesis. Contains CS-D which promotes embryonic axonal growth in CNS in mice. 473HD CS epitope in phosphacan has roles in neural precursor cell proliferation</td>
<td>[147, 152, 294, 295]</td>
<td></td>
</tr>
<tr>
<td>Chondrocyte and SMC perlecan</td>
<td>HSPG2</td>
<td>CS/HS hybrid proteoglycan found in cartilage, IVD,</td>
<td>CS replaces some HS chains in perlecan domain 1 in articular and growth plate cartilage, 4,6-disulphated CS found in growth plate perlecan regulates collagen</td>
<td>[43, 44]</td>
</tr>
<tr>
<td>Protein</td>
<td>Description</td>
<td>Function/Role</td>
<td>Reference(s)</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Leprecan PRH1</td>
<td>Basement membrane</td>
<td>100kDa Leu-Pro enriched CS-proteoglycan of basement membrane of cardiac and skeletal muscle, central nervous system (cerebral cortex and cerebellum), intestinal tract, trachea, ear, skin, liver, and kidney. Localises to the vascular basement membrane/smooth muscle in each organ. Also expressed in the notochord during embryonic chordate development. May also have roles in the secretory pathway and as a growth suppressor.</td>
<td>[296-298]</td>
<td></td>
</tr>
<tr>
<td>Thrombomodulin</td>
<td>Endothelial cell membrane</td>
<td>Endothelial cofactor proteoglycan with roles in the thrombin induced activation of Protein C anticoagulant pathway.</td>
<td>[299, 300]</td>
<td></td>
</tr>
<tr>
<td>Decorin DCN</td>
<td>CS substituted decorin and biglycan are found in bone, but DS forms are normally found in cartilage and skin</td>
<td>SLRPs have roles in ECM organisation/stabilisation/collagen fibrillogenesis. Facilitates cell signalling through interaction with inflammatory cytokines (IL-1, TNF-α) and growth factors (BMPs, WISP-1) and their receptors (EGF-R, IGFIR) affecting cell proliferation, survival, adhesion, migration, matrix synthesis. Controls the bioavailability of TGF-β regulates tissue fibrosis. Biglycan interacts with complement system and TLR4 in innate immune regulation.</td>
<td>[38, 301-303]</td>
<td></td>
</tr>
<tr>
<td>Biglycan BGN</td>
<td>Susceptibility gene in OA</td>
<td>ASPN is unique among the SLRPs in not having a GAG at its N terminus but contains an Aspartic acid repeat region which binds TGF-β BMP-2 to negatively regulate chondrogenesis and osteogenesis.</td>
<td>[304]</td>
<td></td>
</tr>
<tr>
<td>Asporin ASPN</td>
<td>Liver serum serine proteinase inhibitor, also synthesised by IVD cells, chondrocytes and meniscal cells</td>
<td>Epiphycan contains 7 LRRs instead of 10-11 LRRs like other SLRP members, related to osteoglycin</td>
<td>[305]</td>
<td></td>
</tr>
<tr>
<td>Epiphycan EPYC</td>
<td>DS SLRP found in epiphyseal cartilage</td>
<td>Stabilises the condensed HA layer in growth plate hypertrophic region, and around oocytes, proteinase inhibitory activity, anti-bacterial, antiviral, anti-metastatic, immune-modulatory and growth promoting properties</td>
<td>[91, 306, 307]</td>
<td></td>
</tr>
<tr>
<td>Bikunin/ITI AMBP</td>
<td>Brain</td>
<td>CS-A and CS-E brain proteoglycan (Amyloid precursor protein)</td>
<td>[50-53, 161]</td>
<td></td>
</tr>
<tr>
<td>Sushi repeat protein X-linked 2 SRPX2</td>
<td>CS-proteoglycan which is overexpressed in gastrointestinal cancer and has roles in synaptogenesis</td>
<td>SRPX2 is significantly upregulated in colon cancer and its expression levels correlates with tumour aggressiveness. SRPX2 siRNA markedly down regulates β-catenin, MMP-2 and -9 expression reducing tumour cell proliferation, adhesion and migration via the Wnt/β-catenin pathway. SRPX2 promotes synaptogenesis in the cerebral cortex. Mutations in SRPX2 result in Roland epilepsy and speech impairment (RESDX syndrome). Cathepsin B, ADAMTS-4 and uPAR - binding partners of SPRX2 in neural tissues.</td>
<td>[308, 309]</td>
<td></td>
</tr>
<tr>
<td>interphotoreceptor matrix proteoglycan 2 (SPACRSCAN) IMPG2</td>
<td>Eye interphotoreceptor matrix</td>
<td>IMPG2, interphotoreceptor matrix proteoglycan-2</td>
<td>[39]</td>
<td></td>
</tr>
</tbody>
</table>

[120, 121]
| **Serglycin**  
| **SRGN** | Mast cells, platelets, macrophages, T-lymphocytes, leucocytes | Mast cell serglycin is substituted with heparin side chains, macrophage serglycin has CS (CS-A, CS-E) side chains |
| **Endoglycan**  
| **PODXL2** | CD-34 sialomucin transmembrane proteoglycan family member | Contains extensive substitution with sialic acid and N- and O-linked glycan |
| **CD44**  
| **CD44** | CD44 V3 splice variants bearing CS chains have reduced affinity for HA | CD44 binds Ezrin, fibrin/fibrinogen, fibronectin, HA, osteopontin, Selectins-P, -E, -L. Ubiquitous HA receptor |
| **Miscellaneous** | Neurons and endocrine cells | Anion exchange, Chondroitinase ABC MS proteomics screen used to identify intracellular CS-DS proteins. Chromogranin-A, Secretogranin-1, 2, 3. Dermcidin, Neuropeptide W, Cholecystokinin, granule bone marrow cell CS-PGs and collagen and calcium-binding EGF domain-containing protein-1. |

**Abbreviations:** CNS, central nervous system; NCAM, neural cell adhesion molecule; PCM, pericellular matrix; FAK, focal adhesion kinase; ERK, extracellular regulated kinase; LRR, leucine rich repeat; SLRP, small leucine repeat proteoglycan; TLR4, Toll-like receptor-4, MS, mass spectrometry
References


Melrose, J., Smith, SM, Hughes, CE, Little, CB, Catersob, B, Hayes, A J. (2016) The 7D4, 4C3 and 3B3 (-) Chondroitin Sulphation Motifs are expressed at Sites of Cartilage and Bone Morphogenesis during Foetal Human Knee Joint Development. Journal of Glycobiology. 5, 1


165 Bruneau, N. and Szepetowski, P. (2011) The role of the urokinase receptor in epilepsy, in disorders of language, cognition, communication and behavior, and in the central nervous system. Curr Pharm Des. 17, 1914-1923


167 Archinti, M., Britto, M., Eden, G., Furlan, F., Murphy, R. and Degryse, B. (2011) The urokinase receptor in the central nervous system. CNS Neurol Disord Drug Targets. 10, 271-294


Garcia-Suarez, O., Garcia, B., Fernandez-Vega, I., Astudillo, A. and Quiros, L. M. (2014) Neuroendocrine tumors show altered expression of chondroitin sulfate, glypican 1, glypican 5, and syndecan 2 depending on their differentiation grade. Front Oncol. 4, 15


221 Amoury, M., Mladenov, R., Nachreiner, T., Pham, A. T., Hristodorov, D., Di Fiore, S., Helfrich, W., Pardo, A., Fey, G., Schwenkert, M., Thepen, T., Kiessling, F., Hussain, A. F., Fischer,


Gupta, V., Tenny, K. M., Barragan, M., Berkland, C. J. and Detamore, M. S. (2016) Microsphere-based scaffolds encapsulating chondroitin sulfate or decellularized cartilage. J Biomater Appl. 31, 328-343

Xu, H., Yan, Y. and Li, S. (2011) PDLLA/chondroitin sulfate/chitosan/NGF conduits for peripheral nerve regeneration. Biomaterials. 32, 4506-4516


40


314 Dockray, G. J. (2012) Cholecystokinin. Curr Opin Endocrinol Diabetes Obes. 19, 8-12
Legends to Figures

Figure 1. Organisation of GAG saccharides in full length, partially depolymerised, terminal (a) and stub epitopes (b) of the CS side chains of proteoglycans and in the GAG disaccharides in CS types A, B, C, D, E (c) and simplified diagrams of the sulphate presentations in a typical CS side chain (e-g). In the example shown the chain is terminated in a 3-B-3(-) epitope and contains internal 7-D-4 and 4-C-3 epitopes as shown. A 3-B-3(+) stub epitope attached to the linkage tetrasaccharide is also shown, this epitope is generated by chondroitinase ABC. The CS side chains can also be terminated in an alternate 2-B-6(-) epitope and have a 2-B-6(+) stub epitope.

Figure 2. Diagrammatic representation of the sub-domain structural organisation of SRPX2 (Sushi repeat protein, X linked 2) and members of the hyalectan family (a), Decorin and biglycan SLRP members (b) and the inter photoreceptor ECM proteoglycan SPACRCAN (c). SUSHI complement control protein modular data was obtained from the public SMART database (http://smart.embl-heidelberg.de/).

Figure 3. Structural organisation of cell associated CS-proteoglycans. CSPG4 (a), thrombomodulin (b), RPTP/Phosphacan (c), syndecan family (d), and a CS substituted variant of the HA receptor CD44 (e). Abbreviations not covered in key, ED, extracellular domain; TMD, transmembrane domain; CD, cytoplasmic domain.

Figure 4. Organisation of the Kunitz protease inhibitor CS-proteoglycan, bikunin (a), type IX collagen (b), and testican, seminal plasma CS-proteoglycan (c).

Figure 5. Diagrammatic CS-proteoglycans assembled into protective perineural nets in brain tissue (a) and ternary link-protein stabilised macro-aggregate structures with hyaluronan in articular cartilage which convey important hydrodynamic weight bearing and self lubricative properties to this tissue.

Figure 6. Confocal immunolocalisation of the 4C3 (a) and 7D4 (b) native CS sulphation motifs and perlecan using anti-domain IV antibody A7L6 (c) in human foetal (14 week gestational age) knee tibial cartilage. Cell nuclei are stained red with propidium iodide in (a) and (b) and with DAPI in (c). The primary antibody localisations were stained green using FITC conjugated anti mouse or rat IgG. Perlecan identifies stem cell niches in the surface regions of the developing cartilage. Figure modified from [130].

Figure 7. Perlecan and 4C3/7D4 immunolocate the perichondrial stem cell niche and activated progenitor cells involved in foetal elbow joint development. Immunolocalisation of perlecan and the CS sulphation motifs 4C3 and 7D4 using indirect fluorescent confocal microscopy of human foetal elbow (14 week gestational age). Perlecan is immunolocalised to the outer layers of the perichondrium (a, b) while the 7D4 (c-e) and 4C3 CS sulphation motifs (f-i) are located on cell associated proteoglycans deeper in the elbow cartilage rudiment and in the surface regions of the interzone cartilage of the developing elbow joint (f-i).

Figure 8. Focal expression of perlecan (a) and 7-D-4 CS sulphation motif (b) in foetal human paraspinal blood vessels. Perlecan is a well known vascular HS-proteoglycan, produced by endothelial cells. The 7-D-4 CS sulphation motif is focally expressed in the luminal surfaces of these small blood vessels (b) and may provide evidence of a vascular progenitor cell population directed by signals from pericytes on the abluminal surfaces of these vessels (c). Diagram of a small capillary showing the relationship of the endothelial cells and pericytes (d, e). Type IV collagen delineates the bld vessel the pericyte resides on (f) while NG2 proteoglycan is a pericyte marker (g).

Figure 9. Neuronal cell CS-interactive surface molecules with regulatory roles in neuronal development. Contactin-1 (a), leukocyte common antigen-related (LAR) tyrosine phosphatase receptor (b), protein tyrosine phosphatase receptors (PTP-σ and PTP-ζ) (c), o-α-receptor-1 and 3 (NgR1 and NgR3) (d, e), semaphorin-5A (Sem 5A) (f) and neuropilin-1 (NRP-1) (g).

Figure 10. CS-Interactive properties of E-cadherin (a), syndecan proteoglycan family (b) and the frizzled LRP-5/6 co-receptor complex (c) and resultant effects on cytoskeletal re-organisation and gene regulation.
Figure 11. Interactive properties of CS-DS GAG chains of proteoglycans (a), HMBG1, AGEs, S100 proteins and amyloid peptides with the extracellular domains of the RAGE receptor and effects on cytoskeletal proteins (b), activation of NFκB and CREB transcriptional regulation (c).