Chemical Editing of Proteoglycan Architecture

Timothy O'Leary,^{1,2} Meg Critcher,^{1,2} Tesia N. Stephenson,² Xueyi Yang,² Noah H. Bartfield,²
Richard Hawkins,² Mia L. Huang^{2,3*}

¹ Equal authorship

² Department of Molecular Medicine, Scripps Research, 120 Scripps Way, Jupiter, FL 33458-5284

³ Department of Chemistry, Scripps Research, 120 Scripps Way, Jupiter, FL 33458-5284

* Corresponding author email: miahuang@scripps.edu

Abstract

Proteoglycans are heterogeneous macromolecular glycoconjugates that orchestrate many important cellular processes. While much attention has focused on the poly-sulfated glycosaminoglycan chains that decorate proteoglycans, other important elements of proteoglycan architecture, such as their core proteins and cell surface localization, have garnered less emphasis. Hence, comprehensive structure-function relationships that consider the replete proteoglycan architecture as glycoconjugates are limited. Here, we present a comprehensive approach to study proteoglycan structure and biology by fabricating defined semi-synthetic modular proteoglycans that can be tailored for cell surface display. To do so, we integrate amber codon reassignment in the expression of sequence-fined proteoglycan core proteins, metabolic oligosaccharide engineering to produce functionalizable glycosaminoglycans, and bioorthogonal click chemistry to covalently tether the two components. These materials permit the methodical dissection of the parameters required for optimal binding and function of various proteoglycan-binding proteins, and they can be modularly displayed on the surface of any living cell. We demonstrate that these sophisticated materials can recapitulate the functions of native proteoglycans in mouse embryonic stem cell differentiation and cancer cell spreading, while permitting the identification of the most important contributing elements of proteoglycan architecture toward function. This technology platform will confer structural resolution toward the investigation of proteoglycan structure-function relationships in cell biology.

Keywords: Proteoglycan, glycosaminoglycan, syndecan, glycan

Introduction

Proteoglycans (PGs) are ubiquitous macromolecular glycoconjugates that serve dynamic and important roles in cell biology (1). There is incredible detail embedded within the PG architecture, as it is composed of a core protein covalently decorated with poly-sulfated glycosaminoglycans (GAGs). Moreover, the conjugate product can be differentially displayed as soluble or cell surface-anchored molecules (**Fig. 1A**). Despite the overwhelming presence of PGs at the glycocalyx (2), the systematic exploration of the roles of PG architecture as replete glycoconjugates is lacking. Because of the polymorphic nature of PGs, current approaches focus on achieving structural resolution of the individual components (either the protein or GAG chains), but in doing so, they sacrifice the total architecture of PGs.

PG biogenesis begins with the ribosomal synthesis of the core protein (**Fig. 1B**), which is composed of an extracellular-facing ectodomain as well as transmembrane and cytoplasmic domains. The core protein is then post-translationally modified with a tetrasaccharide linker (GlcA-Gal-Gal-Xyl) at serine/threonine residues and extensively elongated by the action of glycosyltransferases, which sequentially install monosaccharides to the growing chain. The elongated chain is composed of repeating units of an amino sugar (GlcNAc or GalNAc) and an uronic acid (glucuronic or iduronic acid), where the disaccharide compositions and glycosidic linkages dictate whether the GAG is classified as a heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), or keratan sulfate (KS). The GAG chains are further modified with sulfate groups at various positions (e.g., 2-, 3-, and 6-O, or *N*-sulfation) by sulfotransferases. In total, an elaborate symphony of >40 enzymes are responsible for generating polydisperse GAGs (3).

PG glycoconjugates transact with their environment via their extracellular-facing ectodomains, which carry out a majority of their functions. Although the covalently attached GAGs can dominate PG sizes, the core protein itself deserves attention. Core proteins can complex with their own set of unique interactors (e.g. receptor tyrosine kinases to activate signaling cascades) that are distinct from its family members, and some ligands even prefer to bind deglycosylated PGs (4). Recent work has also demonstrated that the core protein can even dictate the extent of GAG sulfation (5). From a structural perspective, the core proteins also organize the GAG chains into multivalent copies, and this arrangement can be important for function. The decorating GAGs also serve important roles, where different sulfation patterns can encode binding motifs for GAG-binding proteins, including Fibroblast Growth Factor-2 (FGF2) to regulate cell growth and differentiation. (6, 7) Through variation of these elements, PGs can shape a range of functions and cell behavior. For example, a PG called syndecan-1 (SDC1) is present at similar levels on

both simple and stratified epithelial cells, yet because they exhibit various GAG glycosylation states, they can act as rheostats to meet the vastly different demands of the epithelium to achieve diversified cell adhesion. (8)

Efficient methods to access defined and complete PGs that permit a holistic way to modulate their architectures, are currently lacking. Although advances in the chemical and chemoenzymatic synthesis of carbohydrates have provided well-defined GAG structures (9), these techniques often require substantial technical expertise in carbohydrate chemistry, and to date, only short GAG oligomers (~10 residues) (10, 11) or glycopeptides (12) have been made. Such short oligomers fail to span the chemical space occupied by large GAG chains (100-500 monosaccharide residues). While heparin is commonly used as a surrogate molecule for HS, their structures differ quite dramatically. Native HS is composed of domains of sulfated and non-sulfated regions and is much longer and less sulfated compared to heparin (~0.6 vs. 2.6 sulfates/disaccharide). Moreover, while heparin is often used in its soluble format, native HS is present as both a soluble or a cell surface-anchored molecule, and this differing presentation can have profound impact on function. Neither the oligomers, glycopeptides, nor heparin recapitulate the compartmentalized domain arrangement of native GAGs, where regions of sulfation flank the non-sulfated regions. Failure to control the sulfation patterns of GAGs can lead to contradictory experimental observations. (13, 14) Macromolecular synthetic polymers that can present glycans in a multivalent manner to span larger distances have been employed to recreate GAG functions. (15-17) However, these approaches overlook the functional roles of the core protein. (15, 18) Additionally, simplified genetic techniques to manipulate PG architecture may fail to provide a complete picture of PG biology. Targeting the expression of PG core proteins may result in dysregulated GAG chains, whereas targeting the GAG biosynthetic machinery can alter PG expression patterns or even lethality (19). Moreover, manipulation of sheddases, the enzymes that cleave cell surface-anchored ectodomains to soluble fragments, often results in the off-target removal of many other proteins (20).

Here, we present a straightforward platform to control all of the components of a PG ectodomain, the core protein, the appended GAGs, and its membrane localization (**Fig. 1A**). This platform, rooted in chemical biology, offers unique opportunities to study the biological functions of PGs in high molecular detail, while maintaining a focus on the structures of the individual components. To generate functionalizable PG ectodomains, we chose to use unnatural amino acid (UAA) incorporation to install site-specific bioorthogonal reactive handles. In order to produce materials that recapitulate the native sizes and domain arrangement of native GAGs, we hijacked

the cellular biosynthetic machinery using synthetic azidoxylosides. Finally, to tune the display of the edited PGs onto cell surfaces, we employed a metal-based cell surface engineering strategy. Our engineered PGs have permitted a renewed understanding of how the presence of GAGs or the core proteins influence the binding of cytokines and extracellular matrix components, as well as showcase the importance of properly considering the membrane localization of PGs. We used the engineered SDC1 ectodomains to systematically study interactions with the $\alpha_v\beta_3$ integrin or FGF2 and catalogued their corresponding effects towards mouse embryonic stem cell (mESCs) differentiation and cancer cell spreading.

Results

Production of functionalizable ectodomains through protein engineering. We have developed a modular platform that permits the tailored installation of GAGs and membrane anchors onto PG core proteins. Our platform (**Fig. 1C**) entails the recombinant expression of poly-histidine-tagged core protein ectodomains in *E. coli* hosts, where serine/threonine GAG glycosylation sites have been modified to alkyne-containing unnatural amino acids, such as *p*-propargyl tyrosine (pPY). This modification is accomplished by re-assigning the amber codon, TAG, to encode for pPY through engineered tRNA and corresponding aminoacyl tRNA synthetases (aaRS) (21). We co-transformed competent Rosetta DE3 *E. coli* with individual plasmids encoding mouse syndecans 1-4 (SDC1-4) ectodomains and the suppressor plasmid, pULTRA-CNF (Addgene # 48215) (21). This plasmid encodes for an orthogonal tRNA/aaRS pair derived from *Methanococcus jannaschii*. As a proof of concept, we expressed pPY-containing SDC core protein ectodomains with C-terminal poly-histidine (poly-His) tags for nickel-nitriloacetic acid (NTA)-based purification (22) or complexation (23). We prepared pPY-modified versions of all four members of the syndecan family (SDC1-4), with each protein incorporating one to three pPY amino acids at residues known to incorporate either a HS or CS GAG (**Fig. 1C, Fig. 1D**; pPY positions depicted by subscripts). The SDCs are a family of type I transmembrane core proteins (5), where each SDC bears unique ectodomain sequences, of which SDC1 is the most studied within the family (24). SDC1 is comprised of three conserved sites for GAG attachment; position 37 is known to be modified by CS (sometimes HS), whereas residues 45 and 47 are thought to be mostly modified by HS, and its multivalent arrangement can be required for function. (25) We prepared a monovalent version of SDC1 with one pPY site at residue 37, SDC1₃₇ (**Fig. S1**), and we also explored the possibility of creating fusion proteins of SDC1 for various applications. Hence, we also prepared a APEX2 peroxidase-conjugated version of SDC1 bearing two pPY sites

at residues 45 and 47, separated by a TEV protease-cleavable tag (A-SDC1_{45,47}; **Fig. S2A, B**). The divalent SDC1_{45,47} was prepared by treatment of the APEX2 conjugate with TEV protease (**Fig. S2C**). We focused our efforts towards the three conserved GAG-modified residues within the SDC family, as most functions are ascribed to them (25). As expected, all SDC ectodomains were expressed in their de-glycosylated states (*E. coli* does not include the pertinent GAG biosynthesis machinery; **Fig. S3**). (20) These proteins were efficiently purified from lysates using Ni/NTA-based chromatography, and except for A-SDC1_{45,47}, they migrated on a SDS-PAGE gel as dimers, consistent with previous reports (**Fig. 1E**). (20) We confirmed the presence of the pPY amino acids by copper(I)-catalyzed [3+2] cycloaddition (CuAAC) (26) with a TAMRA azide fluorophore (**Fig. 1F**). The pPY-modified SDCs were expressed in moderate to high yields (0.3-3.0 mg/L), even with three proximally-located pPY sites. We also confirmed that SDC1₃₇ retained its capacity to be selectively processed by known sheddases, as it was processed by MMP-9 and not by ADAMTS-1 (**Fig. S4**). Using the same pULTRA-CNF plasmid, we also expressed modest yields of another UAA-modified proteoglycan, glypican-1 (GPC1), with three of its known HS-modified sites (residues 485, 487, 489) mutated to another propargyl-lysine (**Fig. S5**).

Figure 1. General proteoglycan architecture and biosynthesis. (A) Proteoglycan architecture is composed of multiple elements, a core protein is covalently decorated with glycosaminoglycan (GAG) chains, and the entire glycoconjugate can be differentially displayed as soluble or cell surface-anchored molecules. (B) The ribosomally-synthesized core protein is enzymatically modified with a tetrasaccharide (GlcA-Gal-Gal-Xyl) linker and further elongated (by Exostosin 1/2, Ext1/2). Changes in disaccharide composition and linkages result in the various classes of GAGs (e.g. heparan, chondroitin, dermatan, keratan sulfate). Sulfotransferases (e.g. Ndsts, Hs6sts) transfer sulfate groups to GAGs, shown here are sulfotransferases specific for heparan sulfate PGs. GAG chains are heterogeneous in size and composition, and they can encode specificity for growth factors, such as Fibroblast Growth Factor-2 (FGF2). (C) Amber codon reassignment is used to produce p-propargyltyrosine (pPY)-modified SDC1-4 core proteins. SDC1

cDNA containing modifications at GAG attachment sites to the TAG amber codon were co-transformed in *E. coli* along with a plasmid that encodes for a re-engineered tRNA/aaRS pair, resulting in cells expressing SDC ectodomains with pPY at 1-3 sites (positions denoted by subscripts). **(D)** Table summarizing the PG ectodomain sequences, expected molecular weights, and expression yields. **(E)** SDS-PAGE analysis of the expressed PGs. **(F)** Copper-catalyzed azide-alkyne cycloadditions (CuAAC) with rhodamine-azide shows the presence of fluorescent protein bands, indicative of successful pPY incorporation.

Metabolic engineering with azido xylosides to hijack GAG production. To generate GAGs that recapitulate the length and domain organization of native biomolecules, we prepared recombinant HS (rHS) and CS (rCS) GAGs using a metabolic oligosaccharide engineering strategy, where we hijacked GAG biosynthesis (**Fig. 1B**) using synthetic small molecules (27). Previous work has shown that xyloside molecules derivatized with hydrophobic aglycones, such as phenyl groups, can efficiently serve as decoys and compete for native GAG elongation processes (28-31). We posited that this strategy could be co-opted towards the production of GAGs with functionalizable handles at the reducing ends, by simply appending the hydrophobic moieties with azide groups. Starting from a bromo-xyloside intermediate and *p*-azidophenol, we generated *p*-azidophenol-xyloside **1** (**Fig. 2A; Fig S6**). We modeled how compound **1** docks into the β 4GalT7 galactosyltransferase active sites and observed that it minimizes to -6.3 kcal/mol (**Fig. 2B**), similar to the phenyl derivative lacking the azide group (**Fig. S7**). Furthermore, **1** is positioned such that it can be accessed for subsequent elongation in the GAG biosynthesis pathway. These experiments gave us confidence that the azide group minimally perturbs the aglycone, and that compound **1** could potentially be used to produce recombinant GAGs from cells.

Thus, we incubated **1** (400 μ M) with wild-type Chinese hamster ovary (CHO)-K1 cells (**Fig. 2C**), a cell line commonly used for GAG priming studies, and observed minimal effects in cell viability. We observed significant enhancements in azide-primed GAGs with **1** when incubation times were increased from 48 to 72 hours (**Fig. S8A**). After incubation, both the conditioned medium (CM) and cell extracts (E) were collected. To verify the presence of the azide group, the isolated materials were reacted with AlexaFluor488-alkyne using CuAAC and the resulting products analyzed by SDS-PAGE (**Fig. 2C**). Upon fluorescence visualization, we observed the presence of broad and diffuse fluorescent smears migrating from 8-30 kDa, indicating extensive molecular heterogeneity. As expected, the fluorescent GAGs were mostly present in the CM, as expected, with little to no fluorescent materials was observed in cell extracts (**Fig. S8B**). This observation is consistent with the reactions where copper was excluded, confirming that the fluorescence signal arose due to CuAAC.

When we selectively digested the total recombinant GAGs (rGAGs) with either a cocktail of heparinases (HSase) or chondroitinases (CSase; **Fig. 2D**), we observed that CSase digestion yielded visibly smaller molecular weight fluorescent fragments, while HSase digestion did not cause an immediately obvious change in the migration of the fluorescent band (**Fig. 2C**). This SDS-PAGE experiment suggested that majority of the material from CHO cells was composed of CS, consistent with the *Ext1*-hemizygous nature of CHO cells. Moreover, it suggests that the recombinant GAGs generated from azidoxylsode **1** faithfully captures host cell GAG production. Indeed, when we quantified the composition of the material from the CM, we observed that >80% is composed of CS (**Fig. 2E**) and that ~30% of total HS included the azide motif (**Fig. 2F**). The high concentrations of compounds required for priming are consistent with the proposed metabolic engineering strategy. Although only ~30% of the total HS GAGs are equipped with the azide motif, the unprimed non-azide tagged GAGs are inert towards subsequent CuAAC, and the isolated CM can be used without further purification.

We compared the sizes of the rGAGs isolated from CHO-K1 cells to heparin and observed that the recombinant materials differed in size (**Fig. 2G**). While rHS and rCS seemed to migrate similarly by size exclusion chromatography, both have shorter retention times compared to heparin, suggesting their comparatively large sizes (**Fig. 2G**). Although strategies to sequence HS GAGs remain limited, their disaccharide compositions can be used as a metric to fingerprint their identities (**Fig. S9**). Thus, we also analyzed the composition of the rHS from CHO-K1 by enzymatic depolymerization into the repeating disaccharide units (**Fig. 2H**) (32). As anticipated, the non-sulfated D0A0 disaccharide motif accounted for a large part of these rHS GAGs (33). This resulting composition is expected of native HS and is consistent with its domain-like arrangement, where regions of non-sulfated D0A0 disaccharides flank the sulfated regions. This disaccharide sulfation pattern is in stark contrast to that of heparin, which is mostly composed of the tri-sulfated D2S6 disaccharide motif. We also observed similar HS and CS disaccharide compositions in the rHS and rCS compared to native HS and CS produced by the CHO-K1 cells (**Fig. S10**), as well as in rCS from 48 and 72 hours of priming (**Fig. S11**). Overall, these observations indicate that the priming process produces GAG of similar composition to native host cell GAGs.

Figure 2. Azidoxyloside 1 hijacks GAG biosynthesis to produce recombinant GAGs. (A) Synthetic strategy to produce azidoxyloside **1**. (B) Docking of the enzyme acceptor $\beta 4\text{GalT7}$ with compound **1**. (C) By incubating mammalian cells with azidoxyloside **1** ($400 \mu\text{M}$), we generated recombinant GAGs that bear functional azide handles at the reducing ends. Fluorescence SDS-PAGE gel confirms that GAGs isolated from conditioned media (CM) retain the azide functional handle (confirmed via CuAAC with AF488-alkyne). Most of the GAGs were recovered from the conditioned media (CM) versus cell extracts (E). (D) Selective enzymatic digestion of the total GAG mixture with chondroitinase (CSase) or heparinases (HSase) was used to obtain either recombinant HS (rHS) or recombinant CS (rCS), respectively. (E) As expected from CHO-K1 cells, the GAG mixture is mostly composed of CS. Bar graphs depict means and standard deviations. (F) About 30% of total HS isolated from CHO-K1 cells bear the azide functional group. (G) Recombinant azide-primed HS differs in charge and size compared to heparin. (H) Recombinant HS bears GAG composition akin to native HS and significantly differs from heparin. Here, are shown the top eight most common motifs found in HS.

CuAAC generates proteoglycan glycoconjugates. To construct GAG-decorated PG ectodomains, we subjected azide-terminating GAGs, azido-heparin or azido-CS (**Fig. S13**) or our recombinant GAGs (**Fig. 2**) to CuAAC with the pPY-modified SDC1 ectodomain SDC1₃₇. After one hour, we analyzed the crude reactions using anion exchange chromatography and observed significant conversion to the higher molecular weight and more anionic GAG-conjugated product (**Fig. 3A**). We consistently observed high conversion yields (~90%) regardless of the GAG, core protein, or number of conjugation sites (**Fig. S14-S16**). The reaction conversions and migration patterns of the conjugates were similar to those of a model pPY-modified green fluorescent protein (GFP; **Fig. S17**). We analyzed the fractions from the reaction product of SDC1₃₇ conjugated to total rGAGs from CHO-K1 (generated via 48 or 72 hr incubation periods) and observed SDC1-positive bands using an anti-SDC1 antibody (clone 281-2), which displays reactivity towards both the GAG conjugate and to a lesser extent, the core protein (**Fig. 3B**). In contrast to the localized band for the SDC1₃₇ core protein, the glycoconjugates appeared as large, diffuse signals spanning molecular weights up to 260 kDa when probed using 281-2 or anti-HS antibody clone 10E4 (**Fig. 3C**). Through selective digestion of the rGAG mixture (**Fig. 2D**) and subsequent CuAAC conjugation to SDC1₃₇ or SDC1_{45,47}, we also prepared SDC1_{37rHS} and SDC1_{45,47rHS}. Performing the conjugation step in vitro and purification of the resulting conjugates by anion exchange and size exclusion chromatography further ensures that the purified materials are free of copper to mitigate potential toxicity concerns in live cells. Overall, we demonstrate that despite the poly-anionic nature of the GAGs and the SDC1 ectodomain (pI = 4.7), CuAAC is a robust method to generate PG glycoconjugates.

PG editing enables systematic dissection of structure-binding relationships. We first analyzed whether our PGs had discernable differences in binding to known SDC1-binding proteins, such as the $\alpha_v\beta_3$ integrin (34) or FGF2 (35). Using an ELISA with the engineered SDC1 ectodomains immobilized onto the plate, we observed that $\alpha_v\beta_3$ integrin (**Fig. 3D; Table S2**) and FGF2 (**Fig. 3E; Table S3**) display overall greater affinity for GAG-conjugated SDC1₃₇ compared to the non-conjugated ectodomain alone. Effective binding constants (EC₅₀; **Fig. 3F**) measured for the conjugates against $\alpha_v\beta_3$ followed the ranking: SDC1_{45,47HEP} < SDC1_{37HEP} < SDC1_{45,47HS} < SDC1_{37HS} < SDC1_{37CS}. The $\alpha_v\beta_3$ integrin has been predicted to bind the SDC1 core protein via residues 88-121 (36) and more recently, to heparin and HS via a series of basic lysine residues from both the α_v and the β_3 chains (37). The most potent binder in the series, the divalent SDC1_{45,47HEP} conjugate displayed an approximately 2-fold lower apparent affinity constant relative to the monovalent SDC1_{37HEP} (EC₅₀: 7.3 vs. 13 nM). This observation suggests that the extent of binding is highly influenced by the number of GAG chains, as similarly observed in other studies

(25). The higher affinities observed for the heparin conjugates compared to rHS suggests that the extent of sulfation, or the presence of 3-O-sulfation in heparin, also strongly enhances affinity. The GAG conjugates displayed significantly greater affinity for FGF2 as well, similar to $\alpha_v\beta_3$. EC_{50} values derived for binding FGF2 ranked in the order $SDC1_{37HEP} < SDC1_{37rHS} < SDC1_{45,47HS} \sim SDC1_{37CS} < SDC1_{45,47HEP}$. Notably, we observed an increase in affinity for the divalent $SDC1_{45,47HEP}$ relative to its monovalent counterpart $SDC1_{37HEP}$ (EC_{50} : 9.8 vs. 17 nM).

Figure 3. CuAAC generates replete proteoglycan glycoconjugates. (A) pPY-modified ectodomains can be efficiently reacted in high conversion yields with GAGs (e.g. azide-heparin, rGAGs from CHO-K1 cells) using CuAAC. Similar traces and yields were obtained with rHS and rCS obtained from cells. (B) Fractions collected from purification of the glycoconjugates resulting from recombinant GAGs primed for 48 vs 72 hours were further analyzed by Western blotting. Probing for SDC1 using an anti-SDC1 antibody (clone 281-2) demonstrates greatly enhanced priming with longer incubation. Although we expect a product of ~85 kDa (25 kDa core protein + 60 kDa (recombinant GAG), the large smears (50-260 kDa) are likely due to the unusual migration of these conjugates in SDS-PAGE. Clone 281-2 displays weak affinity for the core

protein, which is enhanced upon conjugation with any GAG. **(C)** Probing the same blot for the presence of HS-modified conjugates using an anti-HS antibody (clone 10E4) demonstrates that despite its low abundance in the mixture relative to CS, it can still be conjugated to SDC1₃₇. **(D)** ELISA with immobilized SDC1 ectodomains against recombinant $\alpha_v\beta_3$ or **(E)** FGF2. **(F)** Tabulated apparent affinity constants (EC_{50}) shown for each PG and target.

Cell surface engineering and stem cell differentiation. We next explored whether altered cell surface presentations (*i.e.* soluble vs. membrane-anchored) altered SDC1-mediated effects in a relevant biological system. To create cell surface-anchored SDC1 ectodomains, we employed a cell surface engineering approach (**Fig. 4A**), where live cells are first incubated with a lipid functionalized with NTA (e.g. cholesterol-PEG2000-NTA; cholPEGNTA) before incubation with SDC1 PGs in the presence of Ni(OAc)₂ (38). The affinity of the C-terminal poly-His tag for the tetradentate Ni-NTA chelating ligand results in complexation and the presentation of the SDC1 ectodomain at the cell surface. We used cholesterol as a lipid anchor, because of its high solubility, its ability to passively incorporate into any phospholipid membrane and to recycle back to the cell surface upon endocytosis (39). We chose to use mouse embryonic stem cells (mESCs) as the backdrop for these experiments, given the known requirement of HS PGs for successful exit from pluripotency and differentiation into neural precursor cells (40). Upon treatment of HS-deficient Ext1^{-/-} mESCs with cholPEGNTA, followed by a poly-His-tagged GFP (GFPHis₆), we observed dose-dependent increases in cellular fluorescence in flow cytometry analysis (**Fig. S18**). Fluorescence was dependent upon the concentrations of cholPEGNTA and GFPHis₆, and maximal fluorescence was achieved at higher doses, indicating that the presentation of molecules (~650,000 molecules of equivalent soluble fluorochrome; **Fig. S18**) using this strategy could be saturated.

Using these optimized conditions, we subjected Ext1^{-/-} mESCs to a six-day adherent differentiation protocol (**Fig. 4B**) (41), where wild-type mESCs gradually lose expression of pluripotency markers and differentiate into SOX1- and β III-tubulin (TubB3)-positive neural precursor cells. On the other hand, Ext1^{-/-} mESCs retain their pluripotent states unless they are supplemented with HS-presenting PGs or exogenous heparin (**Fig. 4C**). We incubated Ext1^{-/-} mESCs with various soluble (-cholPEGNTA) or membrane-anchored (+cholPEGNTA) SDC1 ectodomains for 1 hour only at the onset of the differentiation protocol (at D0) and analyzed the cells for expression of markers of ectodermal lineage and pluripotency after six days (D6, **Fig. 4D**, **Fig. S19**). As expected, we observe increased SOX1 expression in cells treated with our GAG-conjugated SDC1 constructs compared to untreated and SDC1₃₇-treated Ext1^{-/-} mESCs. This is concurrent with a marked decrease in the expression of the pluripotency marker Nanog, indicating successful exit from pluripotency in cells treated with replete SDC1 constructs (**Fig. S19**). Similar

results were obtained from a three-dimensional embryoid body differentiation format (**Fig. S20**), which mimics hallmarks of early embryonic development. Remarkably, the extent of differentiation with our conjugates was similar to mESCs treated with soluble heparin despite its prolonged treatment, (heparin is present in culture from D0-D2). In contrast, mESCs were treated with our soluble engineered ectodomains for only 1 hour on D0, with excess, unbound material washed away. The dependence on GAG-conjugation for differentiation reinforces their requirement for binding FGF2 to trigger ERK1/2 phosphorylation (**Fig. S21**). FGF2 is known to initiate differentiation in mESCs by forming ternary complexes with FGFR and HSPGs, such as SDC1. Ext1^{-/-} mESCs stain positively for SDC1, which is presumably conjugated to CS (**Fig. S22**), yet these cells remain pluripotent when void of external triggers. These observations overall highlight the requirement for HS GAGs for mESC differentiation, with less of a requirement for membrane presentation.

Cell surface engineering and cancer cell spreading. Syndecans are implicated the attachment and spreading of cells on the extracellular matrix (42). Thus, we also evaluated whether our SDC1 constructs could also rescue the spreading of human mammary cancer cells (MDA-MB231) in an established cell spreading assay (**Fig. 4E**) (42). Using SDC1-knockdown and GAG-digested cells (**Fig. S23**), we observed that all soluble forms of the SDC1₃₇ ectodomain failed to rescue cell spreading on vitronectin matrices, and that only membrane-anchored GAG-conjugated variants were able to cause cell spreading (**Fig. 4F, Fig. 4G**). Similar results were obtained for SDC1-knockout cells obtained via CRISPR-Cas9 (**Fig. S24**). The most significant rescue was observed with the divalent heparin SDC1_{45,47HEP}, followed by the monovalent SDC1_{37HEP}. Additionally, the preference for heparin conjugates over HS is observed when comparing SDC1_{37HEP} vs SDC1_{37HS}. The requirements of membrane-anchoring and GAGs are consistent with the significant affinity of $\alpha_v\beta_3$ integrin for heparin (**Fig. 3F**), (34) and they are concurrent with previous observations by Beauvais et al., highlighting the dependence of membrane-bound SDC1 in vitronectin-mediated cell spreading (42). Because cell surface $\alpha_v\beta_3$ binds both cell surface HS-conjugated SDC1 and the immobilized vitronectin (with an apparent $K_D < 10$ nM), it is likely that cell spreading is mediated by $\alpha_v\beta_3$ -SDC1-vitronectin ternary complexes and fine-tuned by *cis* and *trans*-occurring interactions, respectively. While deglycosylated SDC1 may also facilitate cell spreading via other protein-protein interactions or ligands, (42) due to the large enhancement in binding afforded by GAG-conjugated SDC1s, it is likely that the relevant biological mechanism is GAG-mediated.

Figure 4. Cell surface engineering with edited SDC1 ectodomains and their effects on mESC differentiation and cancer cell spreading. (A) A two-step complexation is used to present engineered SDC1 ectodomain variants onto cell surfaces. Live cells are first incubated with cholPEGNTA (1hr, 37°C), followed by Ni(OAc)₂ and the SDC1 ectodomain equipped with a poly-His tag (red tail; 1 hr, 37°C). Excess material is washed off at each step. We did not observe significant toxicity in mESCs with the introduction of any of these materials. (B) A six-day differentiation protocol for mouse embryonic stem cells, illustrating the loss of pluripotency marker Nanog and increased SOX1 expression, indicative of neuroectoderm differentiation. (C) In contrast to wild-type (WT) cells, mouse embryonic stem cells deficient in heparan

sulfate (Ext1^{-/-}) fail to differentiate due to their reduced ability to bind FGF2 (orange) effectively. **(D)** GAG-conjugated SDC1 ectodomains (2 μ M) can rescue differentiation of Ext1^{-/-} mESCs, as evidenced by expression of neural precursor markers Sox1 (green) and Tubb3 (red). Scale bar: 50 μ m **(E)** Workflow for MDA-MB231 cancer cell remodeling spreading assay on vitronectin-coated surfaces. **(F)** Representative fluorescence microscopy images of SDC1-knockdown MDA-MB231 cells following incubation with or without cholPEGNTA (10 μ M, 1 hr, 37°C) and with the engineered SDC1 ectodomains, and stained for phalloidin (red). Cells deficient in SDC1 fail to spread on vitronectin surfaces unless supplemented with membrane-anchored SDC1 conjugated to GAGs. Scale bar: 50 μ m **(G)** Quantification of cell spreading on vitronectin surfaces. Bar graphs represent means and error bars represent standard deviations. For each condition, n >5 images. One-way ANOVA **** $p < 0.0001$, ** $p: 0.0017$.

Proteoglycan editing platform is compatible with live cell proximity tagging. In addition to using our engineered SDC1 ectodomains in phenotypic assays, we also envisioned using the APEX2-tagged PGs to map the interactomes for PGs in live cells. PGs, such as SDC1, possess many binding partners as a consequence of their multi-component nature (43), which may vary as a result of differential glycosylation or cell surface localization. Understanding the structural requirements for PG interactors can aid in assigning mechanisms of action for their biological functions. As an initial proof of concept study, we used A-SDC1_{45,47} and its heparin conjugate A-SDC1_{45,47}HEP in a proximity tagging experiment to capture their interactors in live mESCs (**Fig. 5A**). Proximity tagging has gained recognition as a powerful approach to capture the interacting proteins for a given protein bait. In this approach, the APEX2 peroxidase enzyme catalyzes the formation of short-lived (< 1 ms) and proximal (<20 nm radius) biotinyl radicals that covalently react with proteins interactors for the bait protein. As such, the interacting proteins are biotinylated while the cells are alive, without disrupting their endogenous localization. The biotinylated proteins can then be probed using fluorophore-conjugated streptavidin and imaged using fluorescence microscopy or enriched for mass spectrometry-based identification. The APEX2 approach is especially apt for glycan-protein interactions that are heavily influenced by three-dimensional presentations of ligands, and which may be dynamic or weak (44).

We qualitatively compared the differential interactomes of A-SDC1_{45,47} when it is presented as a soluble or a cell-surface anchored molecule. Following incubation with cholPEGNTA, mESCs were similarly incubated with Ni(OAc)₂ and the A-SDC1_{45,47} ectodomains. We observed significant dose-dependent (**Fig. S25**) biotinylation of mESCs following proximity tagging (**Fig. 5B**). The majority of the signals arose from the cell surface, in both WT and Ext1^{-/-} mESCs (**Fig. S26**) consistent with the known localization of SDC1. Interestingly, we also observed varying intensities of biotinylation depending on the presence of the GAG chain and the presentation of the fusion protein, with highest signals were observed from membrane-anchored, deglycosylated A-SDC1_{45,47}. The detection of fluorescence in mESCs treated with soluble versions of A-SDC1_{45,47}

was expected, given that SDC1 is known to interact with cell surface ligands, such as integrins (**Fig. 3D**). While we have yet to assign the differential interactomes resulting from changes in PG architecture, we demonstrate that fusion proteins of our engineered PGs have tremendous utility and promise to fine-tune our understanding of native PG interactions in various contexts with minimal perturbation.

Figure 5. Proximity tagging with A-SDC1. (A) Live mESCs were incubated with A-SDC1_{45,47} or A-SDC1_{45,47HEP} (2 μ M, 1 hr), with or without pre-treatment with cholPEGNTA (10 μ M, 1 hr), to generate cell surface-anchored or soluble SDC1 ectodomains, respectively. Following treatment with biotin-phenol (yellow circle, 0.5 mM, 30 min) and a brief incubation with H₂O₂ (100 mM, 1 min), the APEX2 peroxidase catalyzes the formation of reactive biotin radicals that are short-lived (< 1 ms) and react with proteins located proximally (< 20 nm) from the complex. Biotinylated interactors are detected using Cy5-streptavidin. (B) Wild-type mESCs subjected to the proximity tagging protocol demonstrate robust fluorescence signals from biotinylated interactors. The signals are mostly located at the cell surface, compared to the non-treated control. Membrane anchoring and glycosylation appear to regulate the extent of fluorescence, and thus the interactomes, of SDC1 fusion proteins.

Discussion

PGs are important information-dense molecules that reside at the cell surface. Because glycosylation is not directly controlled by the genetic code, PGs naturally exist as molecular polymorphs of various GAG motifs attached to a single core protein. It is widely speculated that the molecular heterogeneity found within PGs is important as it largely contributes to their

functional diversity, although this remains poorly understood due to the challenges of studying PG architecture. To begin systematically dissecting the effects of PG molecular heterogeneity toward function, we have assembled a holistic platform to chemically edit PGs using an integrated approach that merges unnatural amino acid incorporation, metabolic oligosaccharide engineering, and cell surface engineering. Starting with the core proteins of SDCs 1-4, we incorporated unnatural amino acids at sites of native GAG glycosylation for bioorthogonal conjugation (**Fig. 1**). Notably, we recapitulated the endogenous multi-glycosylated nature of SDC1, which is known to be modified by HS at two sites.

GAGs are among the most complex and ubiquitous forms of glycan chains. While much heroic work has focused on the total synthesis of short GAG oligomers and proteoglycan-mimetic glycoconjugates, thus far, these products remain limited in their ability to capture the sophistication of PG architectures. We demonstrate the use of a metabolic oligosaccharide engineering approach to append bioorthogonal labels in GAGs, which are traceable using CuAAC (**Fig. 2**). Owing to the common tetrasaccharide linker for both HS and CS GAGs, the azidoxylside **1** generates azide-modified versions of both. Using selective enzymatic digestion, we are also able to generate homogenous GAGs (i.e. HS vs CS) despite the fact they are produced from cells as a mixture. These homogeneous GAG samples will be especially valuable toward understanding the functional relevance of PGs that are present as various hybrids of HS or CS. While we have yet to fully map the intermediate metabolic conversions the compound undergoes once inside the cell, it likely undergoes similar mechanisms as other xylosides (28, 31, 45).

We demonstrate here that CuAAC, a widely used bioorthogonal reaction, can be used to conjugate a highly polyanionic GAG molecule to one or two sites of an equally large PG core protein (**Fig. 3**). Owing to its bioorthogonal nature, this conjugation method is relatively inert, and it is compatible with heme-containing proteins, such as the APEX2 peroxidase. Although the GAGs produced by compound **1** remain heterogeneous in sizes, they retain the macromolecular feature and domain-like arrangement of sulfation motifs of natural GAGs. In studying the binding interactions between the engineered SDC1 ectodomains, we uncovered that the divalent SDC1_{45,47HEP} binds more tightly toward $\alpha_v\beta_3$ compared to the monovalent SDC1_{37HEP}, which was functionally recapitulated in cell spreading. These observations could be attributed to potential synergistic or cooperative effects, although we cannot rule out the potential for positional importance, where specific positions among the three GAG sites of SDC1 can contribute more towards binding than others (25). While we have made significant strides to achieve defined yet native architectures for PGs, there remain opportunities to expand the repertoire of edited PGs using multiple UAAs (46) and other synthetic xylosides primers.

In contrast to standard genetic approaches, our platform allows the facile interrogation of the roles of specific GAG structures within the context of the native PG core protein, and it permits us to tune the presentation of PG ectodomains. Altering the presence and composition of GAG chains via genetic means can disproportionately cause higher expression levels of certain SDC1 mutants or affect the glycosylation of other GAG sites (47), complicating the analysis of the effects of GAG deletions (25). Our platform not only provides a means to generate defined PG ectodomains, but coupled to cell surface engineering strategies, it greatly facilitates the study of native PG glycoconjugate interactions functions at surfaces of live cells, especially in those that are not readily amenable to genetic manipulation (e.g. primary cells, stem cells). The ability to perform these structure-activity studies is crucial, as many organisms and proteins that interact with HS GAGs *in vitro* have been observed to not interact to native HSPGs (48). These differences could be due to additional binding contributions of the core protein itself, the differing fine structure of HS among tissues, or the absence of the multivalent architectural arrangement naturally present in PGs. To our knowledge, this work represents the most advanced effort toward the generation of full PG glycoconjugates with defined GAG chains.

We have demonstrated the utility of our platform in two functional phenotypic assays, mESC differentiation (**Fig. 4D**) and cancer cell spreading (**Fig. 4G**). Both systems illustrate the requirement for the correct GAG compositions in order to exhibit positive effects, yet it is only with vitronectin-mediated cancer cell spreading that we observe an obligate requirement for cell surface anchoring. These observations could be due to the fact that the FGF2 is also present as a soluble molecule in mESC differentiation, whereas in cancer cell spreading, both vitronectin and the $\alpha_v\beta_3$ integrin are anchored either as a substrate or on the cell surface, such that ternary complex formation is much more difficult to achieve with soluble SDC1. However, we cannot rule out the possibility that there may be different extents of ternary complex formation in these two systems, and an auto-inhibitory effect may be achieved with soluble SDC1 at higher concentrations (49).

The ability to tag interactors in live mESCs by combining cell surface engineering with proximity tagging open up new opportunities to study PG interactions. Notably, SDC1 alone exhibits over 100 interactors in different contexts. (43) Our cell surface engineering approach is applicable to any cell with a phospholipid membrane, and it is especially useful for mESCs, which are generally recalcitrant towards transient genetic manipulations. As such, stem cells are not amenable to structure-activity relationship studies of PGs that require knock-in/knockdown of individual components. Studying PG biology at the cell surface, where they are free to interact

with their ligands and move along the plane of the membrane, may also allow focal concentration of the PGs to enact their corresponding functions. SDC1 is an important PG that mediates myriad complex biological functions. When anchored at the cell surface, it can serve as a functional receptor for microbes (50) and cytokines, and promote the assembly of extracellular matrix components (51). Interestingly, when it is processed by matrix metalloproteinases or sheddases, the soluble ectodomains can carry out competing functions to membrane-anchored SDC1 (52). Thus, we expect that further inquiry into these biological contexts may benefit from our platform. While there are 16 classically known PGs, of which we have made progress on four, recent work has identified that other proteins not conventionally designated as PGs, can bear GAG chains and carry out important biological functions (53, 54). These discoveries therefore create even further opportunities for the use of glyco-engineering platforms, like ours, to aid in our understanding of the biology of protein glycoconjugates.

Conclusion

In summary, we have generated a modular platform that permits the tailored assembly of proteoglycan ectodomains, replete with sequence-specific core proteins, defined GAGs, and the ability to present them onto cell surfaces. The capacity to site-specifically incorporate a growing number of functionalized unnatural amino acids into PG backbones provides new opportunities to study their structural conformations, as well as chemoselectively conjugate additional moieties (55, 56). The efficient metabolic oligosaccharide engineering approach we used to generate working amounts of azide-modified GAGs is compatible with multiple cell types, and sulfation-mutant cell lines will greatly facilitate studies on native GAGs beyond heparin (57-59). The GAG-conjugated SDC1 proteoglycans used here represent the most defined semi-synthetic materials (60) to date, and they capture many functional facets of PG ectodomains. We have demonstrated that the systematic variation of GAG composition and multivalency on SDC1 core proteins can yield refined insights into structure-function relationships, as it relates to mESC differentiation and cancer cell spreading. These materials may also serve beneficial for controlled localized delivery of GAG-binding proteins (60). The process for producing these conjugates is relatively simple - over the course of these studies, we have produced sub-milligram quantities of the most complex material in the series, SDC1_{rGAG}. Lipid-based cell surface engineering and radical-based proximity tagging approaches are readily assimilated into the engineered proteoglycans, and we anticipate that these materials and this platform will be further utilized to dismantle the complexity of proteoglycans in many other biological contexts.

Methods

Materials and reagents. Unless otherwise noted, chemical reagents were used as purchased and obtained from commercial vendors. Recombinant human $\alpha_v\beta_3$ (R&D Biosystems # 3050-AV-050) is composed of the α_v (Phe31-Val992, Accession # NP_002201) and the β_3 (Gly27-Asp718, Accession # AAA52589) subunits. Recombinant human FGF2 (NovusBio # NBP2-76301) is composed of 154 amino acids (Accession Number: P09038). Wild-type mESCs and MDA-MB231 cells were purchased from ATCC, and Ext1^{-/-} mESCs were a gift from Kamil Godula, Catherine Merry, and Jeffrey Esko.

Expression of pPY-modified PG ectodomains. The syndecan ectodomain cDNA sequences were derived from EMbL, mutated as desired, codon-optimized for bacterial expression, synthesized (Genscript, NJ, USA), and cloned into the NheI/NotI sites of the pET-28a(+) vector with kanamycin resistance. Rosetta 2 (DE3) competent (Novagen) bacterial cells were co-transformed with the SDC-containing plasmid and pULTRA-CNF (Addgene # 48215, a gift from Peter Schultz) [14]. Following spectinomycin and kanamycin-mediated selection, single colonies were picked and expanded in LB media containing the same antibiotics and cultured until OD₆₀₀ ~ 0.6-0.8. IPTG and the pPY unnatural amino acid (400 mg/L) were then added, and the culture was further incubated overnight. The cells were then harvested, lysed, and purified using FPLC systems equipped with nickel-NTA columns to harvest the SDC1 ectodomains. Following elution with imidazole and subsequent purification by anion exchange or size exclusion chromatography, proteins were stored in PBS with 25% glycerol, flash frozen in methanol/dry ice and stored at -80 °C.

Chemical synthesis of azidoxyloside 1 (Fig. 2B). See supporting information for full chemical synthesis and purification.

Molecular docking of azidoxyloside 1 with β 4GalT7. Compound 1 was docked into the active site of β 4GalT7 in complex with UDP-galactose (PDB ID 4M4K) (61) using Autodock Vina (62).

Production of recombinant GAGs by azidoxyloside priming. CHO-K1 cells were seeded on 15 cm² dishes overnight and incubated with 1 (400 μ M, 48-72 hr) in basal DMEM media. After the desired incubation period, the conditioned media (CM) was collected, lyophilized, and purified using manual DEAE columns. Dialysis with water was performed to remove excess salt.

Disaccharide analysis and quantification of azide-primed GAGs. Harvested GAGs were digested overnight with 2.5 U/mL heparinases I, II, III (Sigma) or 0.5 U/mL chondroitinase ABC (Sigma). Disaccharides were isolated with a 3000 MWCO filter (Thermo) and labeled with anthranilamide (63). Labeled disaccharides were analyzed with a Propac PA1 column (ThermoFisher Scientific) on an Ultimate 3000 UHPLC with fluorescence detection (57). To gauge the extent of azide incorporation, GAGs were reacted with sulfo-DBCO-biotin (Sigma) (6.7 mM, 20 h, 37 °C, PBS). Excess DBCO-biotin was removed with size-filtration, and GAGs were incubated with streptavidin agarose (Sigma). Unbound GAGs were separated from GAGs captured on the streptavidin beads, and both fractions were treated with either heparinase or

chondroitinase. Disaccharides were labeled and analyzed and proportions of azide-labeled GAGs were measured by comparing fluorescence signals for bead-captured vs. unbound fractions.

CuAAC conjugations. CuAAC conjugations (60) were performed in PBS (pH 6-8) with the proteoglycan (25 μ M), azide-GAG (900 μ M), aminoguanidine hydrochloride (5.0 mM), CuSO₄ (320 μ M), Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; 1.6 mM), and sodium ascorbate (21 mM), 37 °C for 2 hr. Conjugates were purified using weak anion exchange DEAE chromatography (WAX-10 4x250 mm column, 0 to 400 mM NaCl, 15 min, then 1.75 M NaCl to elute). DEAE eluate was loaded onto a Ni-NTA column, and washed to remove unconjugated azide-GAGs (1 M NaCl tris buffer). The reaction product was eluted with imidazole, desalted, and concentrated prior to analysis or subsequent use.

ELISA with immobilized syndecans. Each SDC1 construct was immobilized in duplicate using pH 9.6 carbonate buffer onto 96-well high-binding plates (4°C, overnight, rocking, 50 mL, 10 mg/mL) (Thermo Maxisorp plates, 439454). Following blocking (2% BSA/PBST, 100 mL) and washing 3X with PBST (100 mL), wells were incubated with titrations of biotinylated FGF2 or $\alpha_v\beta_3$ (PBST, 4°C, overnight, rocking, 50 mL). Following 3X washes with PBST, wells were probed with HRP-streptavidin (Biolegend 405210, 1000:1) or biotinylated anti- $\alpha_v\beta_3$ antibody (Biolegend 304412, 1:1000) (RT, 1 hr, rocking, 100 mL) followed by washing and HRP-streptavidin (RT, 1 hr, rocking). Wells were then washed 3X with PBST and incubated with a TMB substrate (BioFX, 100 mL) until a solid blue color develops, and the reaction was quenched with sulfuric acid (2N, 100 mL). Absorbance values at 450 nm was read and subtracted from blank (no FGF2 or $\alpha_v\beta_3$) wells, normalized to 100%, and fitted into a non-linear logarithmic curve (Prism).

Cell surface engineering. Cells were washed with DPBS and incubated with cholPEGNTA (Nanocs # PG2-CSNT-2k; 10 μ M, 1 hr, 37°C) After washing twice with DPBS, cells were incubated with SDC1 constructs (pre-centrifuged at 14,000 xg, 10 min. to exclude potential aggregates in KO-DMEM media (1 hr, 37°C). Cells were then washed 2X with DPBS and re-suspended in N2B27 media for 6 days to differentiate. For neuronal differentiation, cells were cultured in differentiation (N2B27) media for 6 days (**Fig. 4B**). For the cell spreading assay, cells were plated onto vitronectin-coated plates (**Fig. 4E**).

Neuronal differentiation of mouse embryonic stem cells (mESCs). For adherent mESC differentiation (64), cells were seeded overnight (37°C, 5% CO₂, ~18 hr) in maintenance media (MM) at a density of 1x10⁴ cells/cm² in 24-well gelatin-coated 24-well plates. The following day (D0), mESCs were washed 1X with PBS and the media was replaced with N2B27 neural differentiation media. Subsequent washes and media replacements were performed at D2, D4, and the cells were fixed at D6, stained for various markers, and imaged using fluorescence microscopy.

Cell Spreading Assay. SDC1-knockdown GAG-digested MDA-MB-231 cells harvested using a non-enzymatic cell dissociation buffer were seeded (5x10⁴ cells per well) in 24-well plates pre-coated with poly-D-lysine and vitronectin (2 hr at 37°C, 5% CO₂). After washing 1X with PBS and fixation, cell spreading was visualized using rhodamine-conjugated phalloidin and imaged using fluorescence microscopy.

Live cell proximity labeling. Proximity tagging was adapted from previously published procedures (44, 65). mESCs seeded on gelatin were treated with cholPEGNTA (10 μ M, 1 hr, 37°C) and further incubated with APEX2 fused proteoglycans (10 μ M, 1 hr, 37°C) after washing 2X DPBS. After further washes, cells were incubated with biotin phenol (500 μ M in KO-DMEM, 30 mins at 37°C), followed by H₂O₂ (1 mM, 1 min, RT). The reaction was stopped with 3X washes of quenching solution (5 mM Trolox, 10 mM sodium ascorbate, and 10 mM sodium azide in DPBS). mESCs were then fixed, blocked, probed for biotinylated interactors by Cy5-streptavidin, and imaged by fluorescence microscopy.

Author contributions

T.O., M.C., T.N.S., and M.L.H. conceived and designed the research. T.O. performed molecular docking simulations. T.O. and X. Y. expressed engineered proteoglycans and performed bioconjugation reactions. T.N.S., N.B., and R.H. performed xyloside priming experiments and T.O., T.N.S., and N.B. performed GAG analysis. M.C. performed mESC differentiation, cell spreading assays, and proximity tagging experiments. X. Y. contributed to biological assays. T.O., M.C., T.N.S., and M.L.H contributed to manuscript preparation.

Acknowledgements

We thank Drs. Sourav Chatterjee and Abdullah Hassan for their assistance in preparing the azidoxyloside compound. This work, T.N.S. (HD090292-S1), M.C., and M.L.H are supported by the NIH K99/R00 Pathway to Independence Award (R00-HD090292). M.L.H is grateful for the support and scientific counsel of Profs. Kamil Godula and Jeffrey Esko for this work.

References

1. U. Hacker, K. Nybakken, N. Perrimon, Heparan sulphate proteoglycans: the sweet side of development. *Nat Rev Mol Cell Biol* **6**, 530-541 (2005).
2. M. Critcher, T. O'Leary, M. L. Huang, Glycoengineering: scratching the surface. *Biochem J* **478**, 703-719 (2021).
3. M. Salmivirta, K. Lidholt, U. Lindahl, Heparan sulfate: a piece of information. *FASEB J* **10**, 1270-1279 (1996).
4. Y. H. Zhang *et al.*, Targeting of Heparanase-modified Syndecan-1 by Prosecretory Mitogen Lacritin Requires Conserved Core GAGAL plus Heparan and Chondroitin Sulfate as a Novel Hybrid Binding Site That Enhances Selectivity. *Journal of Biological Chemistry* **288**, 12090-12101 (2013).
5. M. Bernfield *et al.*, Biology of the Syndecans - a Family of Transmembrane Heparan-Sulfate Proteoglycans. *Annu Rev Cell Biol* **8**, 365-393 (1992).
6. P. W. Park, O. Reizes, M. Bernfield, Cell surface heparan sulfate proteoglycans: selective regulators of ligand-receptor encounters. *J Biol Chem* **275**, 29923-29926 (2000).
7. C. I. Gama *et al.*, Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2**, 467-473 (2006).
8. R. D. Sanderson, M. Bernfield, Molecular polymorphism of a cell surface proteoglycan: distinct structures on simple and stratified epithelia. *Proc Natl Acad Sci U S A* **85**, 9562-9566 (1988).

9. S. E. Tully *et al.*, A chondroitin sulfate small molecule that stimulates neuronal growth. *Journal of the American Chemical Society* **126**, 7736-7737 (2004).
10. C. Noti, J. L. de Paz, L. Polito, P. H. Seeberger, Preparation and use of microarrays containing synthetic heparin oligosaccharides for the rapid analysis of heparin-protein interactions. *Chem-Eur J* **12**, 8664-8686 (2006).
11. Z. Wang *et al.*, Preactivation-Based, One-Pot Combinatorial Synthesis of Heparin-like Hexasaccharides for the Analysis of Heparin-Protein Interactions. *Chem-Eur J* **16**, 8365-8375 (2010).
12. W. Z. Yang *et al.*, Chemical synthesis of human syndecan-4 glycopeptide bearing O-, N-sulfation and multiple aspartic acids for probing impacts of the glycan chain and the core peptide on biological functions. *Chem Sci* **11**, 6393-6404 (2020).
13. S. Nadanaka, A. Clement, K. Masayama, A. Faissner, K. Sugahara, Characteristic hexasaccharide sequences in octasaccharides derived from shark cartilage chondroitin sulfate D with a neurite outgrowth promoting activity. *Journal of Biological Chemistry* **273**, 3296-3307 (1998).
14. P. A. Brittis, D. R. Canning, J. Silver, Chondroitin Sulfate as a Regulator of Neuronal Patterning in the Retina. *Science* **255**, 733-736 (1992).
15. M. L. Huang, R. A. A. Smith, G. W. Triege, K. Godula, Glycocalyx Remodeling with Proteoglycan Mimetics Promotes Neural Specification in Embryonic Stem Cells. *Journal of the American Chemical Society* **136**, 10565-10568 (2014).
16. Y. Miura, T. Fukuda, H. Seto, Y. Hoshino, Development of glycosaminoglycan mimetics using glycopolymers. *Polym J* **48**, 229-237 (2016).
17. M. Rawat, C. I. Gama, J. B. Matson, L. C. Hsieh-Wilson, Neuroactive chondroitin sulfate glycomimetics. *Journal of the American Chemical Society* **130**, 2959-2961 (2008).
18. A. Pulsipher, M. E. Griffin, S. E. Stone, J. M. Brown, L. C. Hsieh-Wilson, Directing Neuronal Signaling through Cell-Surface Glycan Engineering. *Journal of the American Chemical Society* **136**, 6794-6797 (2014).
19. E. Forsberg, L. Kjellen, Heparan sulfate: lessons from knockout mice. *J Clin Invest* **108**, 175-180 (2001).
20. J. M. Kim, K. Lee, M. Y. Kim, H. I. Shin, D. Jeong, Suppressive effect of syndecan ectodomains and N-desulfated heparins on osteoclastogenesis via direct binding to macrophage-colony stimulating factor. *Cell Death Dis* **9** (2018).
21. K. C. Schultz *et al.*, A genetically encoded infrared probe. *J Am Chem Soc* **128**, 13984-13985 (2006).
22. J. Porath, J. Carlsson, I. Olsson, G. Belfrage, Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599 (1975).
23. C. M. Csizmar, J. R. Petersburg, C. R. Wagner, Programming Cell-Cell Interactions through Non-genetic Membrane Engineering. *Cell Chem Biol* **25**, 931-940 (2018).
24. S. Saunders, M. Jalkanen, S. Ofarrell, M. Bernfield, Molecular-Cloning of Syndecan, an Integral Membrane Proteoglycan. *Journal of Cell Biology* **108**, 1547-1556 (1989).
25. J. K. Langford, M. J. Stanley, D. Cao, R. D. Sanderson, Multiple heparan sulfate chains are required for optimal syndecan-1 function. *J Biol Chem* **273**, 29965-29971 (1998).
26. B. C. Bundy, J. R. Swartz, Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein-Protein Click Conjugation. *Bioconjugate Chem* **21**, 255-263 (2010).
27. D. H. Dube, C. R. Bertozzi, Metabolic oligosaccharide engineering as a tool for glycobiology. *Curr Opin Chem Biol* **7**, 616-625 (2003).
28. T. A. Fritz, J. D. Esko, Xyloside priming of glycosaminoglycan biosynthesis and inhibition of proteoglycan assembly. *Methods Mol Biol* **171**, 317-323 (2001).
29. V. M. Tran, B. Kuberan, Synthesis of Fluorophore-Tagged Xylosides That Prime Glycosaminoglycan Chains. *Bioconjugate Chem* **25**, 262-268 (2014).

30. S. Dahbi *et al.*, Synthesis of a library of variously modified 4-methylumbelliferyl xylosides and a structure-activity study of human beta 4GalT7. *Org Biomol Chem* **15**, 9653-9669 (2017).
31. J. S. Chua, B. Kuberan, Synthetic Xylosides: Probing the Glycosaminoglycan Biosynthetic Machinery for Biomedical Applications. *Accounts Chem Res* **50**, 2693-2705 (2017).
32. S. M. Carnachan, S. F. R. Hinkley, Heparan Sulfate Identification and Characterisation: Method II. Enzymatic Depolymerisation and SAX-HPLC Analysis to Determine Disaccharide Composition. *Bio-Protocol* **7** (2017).
33. R. Lawrence, H. Lu, R. D. Rosenberg, J. D. Esko, L. Zhang, Disaccharide structure code for the easy representation of constituent oligosaccharides from glycosaminoglycans. *Nat Methods* **5**, 291-292 (2008).
34. C. Faye *et al.*, Molecular Interplay between Endostatin, Integrins, and Heparan Sulfate. *Journal of Biological Chemistry* **284**, 22029-22040 (2009).
35. M. S. Filla, P. Dam, A. C. Rapraeger, The cell surface proteoglycan syndecan-1 mediates fibroblast growth factor-2 binding and activity. *J Cell Physiol* **174**, 310-321 (1998).
36. A. C. Rapraeger, Synstatin: a selective inhibitor of the syndecan-1-coupled IGF1R- α v β 3 integrin complex in tumorigenesis and angiogenesis. *FEBS J* **280**, 2207-2215 (2013).
37. L. Ballut, N. Sapay, E. Chautard, A. Imberty, S. Ricard-Blum, Mapping of heparin/heparan sulfate binding sites on α v β 3 integrin by molecular docking. *J Mol Recognit* **26**, 76-85 (2013).
38. S. Boonyarattanakalin, S. Athavankar, Q. Sun, B. R. Peterson, Synthesis of an artificial cell surface receptor that enables oligohistidine affinity tags to function as metal-dependent cell-penetrating peptides. *J Am Chem Soc* **128**, 386-387 (2006).
39. E. C. Woods, N. A. Yee, J. Shen, C. R. Bertozzi, Glycocalyx Engineering with a Recycling Glycopolymer that Increases Cell Survival In Vivo. *Angew Chem Int Ed Engl* **54**, 15782-15788 (2015).
40. C. E. Johnson *et al.*, Essential alterations of heparan sulfate during the differentiation of embryonic stem cells to Sox1-enhanced green fluorescent protein-expressing neural progenitor cells. *Stem Cells* **25**, 1913-1923 (2007).
41. S. Chatterjee, T. N. Stephenson, A. L. Michalak, K. Godula, M. L. Huang, Silencing glycosaminoglycan functions in mouse embryonic stem cells with small molecule antagonists. *Method Enzymol* **626**, 249-270 (2019).
42. D. M. Beauvais, B. J. Burbach, A. C. Rapraeger, The syndecan-1 ectodomain regulates α v β 3 integrin activity in human mammary carcinoma cells. *J Cell Biol* **167**, 171-181 (2004).
43. M. A. Stepp, S. Pal-Ghosh, G. Tadvalkar, A. Pajoohesh-Ganji, Syndecan-1 and Its Expanding List of Contacts. *Adv Wound Care* **4**, 235-249 (2015).
44. E. Joeh *et al.*, Mapping glycan-mediated galectin-3 interactions by live cell proximity labeling. *Proc Natl Acad Sci U S A* **117**, 27329-27338 (2020).
45. B. J. Beahm *et al.*, A visualizable chain-terminating inhibitor of glycosaminoglycan biosynthesis in developing zebrafish. *Angew Chem Int Ed Engl* **53**, 3347-3352 (2014).
46. A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. *Biochemistry* **52**, 1828-1837 (2013).
47. A. S. Eriksson, D. Spillmann, The Mutual Impact of Syndecan-1 and Its Glycosaminoglycan Chains-A Multivariable Puzzle. *J Histochem Cytochem* **60**, 936-942 (2012).

48. P. W. Park, G. B. Pier, M. T. Hinkes, M. Bernfield, Exploitation of syndecan-1 shedding by *Pseudomonas aeruginosa* enhances virulence. *Nature* **411**, 98-102 (2001).
49. M. Fannon, K. E. Forsten, M. A. Nugent, Potentiation and inhibition of bFGF binding by heparin: a model for regulation of cellular response. *Biochemistry* **39**, 1434-1445 (2000).
50. R. S. Aquino, Y. H. Teng, P. W. Park, Glycobiology of syndecan-1 in bacterial infections. *Biochem Soc Trans* **46**, 371-377 (2018).
51. A. Jinno, A. Hayashida, H. F. Jenkinson, P. W. Park, Syndecan-1 Promotes *Streptococcus pneumoniae* Corneal Infection by Facilitating the Assembly of Adhesive Fibronectin Fibrils. *mBio* **11** (2020).
52. T. Manon-Jensen, H. A. B. Multhaupt, J. R. Couchman, Mapping of matrix metalloproteinase cleavage sites on syndecan-1 and syndecan-4 ectodomains. *Febs Journal* **280**, 2320-2331 (2013).
53. P. Zhang *et al.*, Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships. *Cell* **174**, 1450-+ (2018).
54. F. Noborn *et al.*, Identification of Chondroitin Sulfate Linkage Region Glycopeptides Reveals Prohormones as a Novel Class of Proteoglycans. *Mol Cell Proteomics* **14**, 41-49 (2015).
55. K. Lang, J. W. Chin, Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem Rev* **114**, 4764-4806 (2014).
56. T. A. Nguyen, M. Cigler, K. Lang, Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew Chem Int Edit* **57**, 14350-14361 (2018).
57. H. Qiu *et al.*, A mutant-cell library for systematic analysis of heparan sulfate structure-function relationships. *Nature Methods* **15**, 889-+ (2018).
58. M. Ishihara, Animal cell mutants defective in glycosaminoglycan biosynthesis. *Trends Glycosci Glyc* **9**, 235-236 (1997).
59. Y. H. Chen *et al.*, The GAGome: a cell-based library of displayed glycosaminoglycans. *Nat Methods* **15**, 881-888 (2018).
60. S. I. Presolski, V. P. Hong, M. G. Finn, Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Curr Protoc Chem Biol* **3**, 153-162 (2011).
61. A. Siegbahn *et al.*, Exploration of the active site of β 4GalT7: modifications of the aglycon of aromatic xylosides. *Organic & Biomolecular Chemistry* **13**, 3351-3362 (2015).
62. O. Trott, A. J. Olson, Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **31**, 455-461 (2010).
63. J. C. Bigge *et al.*, Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid. *Anal Biochem* **230**, 229-238 (1995).
64. Q. L. Ying, M. Stavridis, D. Griffiths, M. Li, A. Smith, Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-186 (2003).
65. S. S. Lam *et al.*, Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51-54 (2015).