Xylosyltransferase Bump-and-hole Engineering to Chemically Manipulate Proteoglycans in Mammalian Cells

Zhen Li[a,b], Lucia Di Vagno[b,c], Aisling Ni Cheallaigh[d], Douglas Sammon[e], David C. Briggs[f], Nara Chung[g], Vincent Chang[g], Keira E. Mahoney[g], Anna Cioce[a,b], Lloyd D. Murphy[h], Yen-Hsi Chen[i,j], Yoshiki Narimatsu[i], Rebecca L. Miller[i], Lianne I. Willems[h], Stacy A. Malaker[g], Gavin J. Miller[d], Erhard Hohenester[e], Benjamin Schumann[a,b]

[a] Department of Chemistry, Imperial College London, W12 0BZ, London, United Kingdom.
[b] Chemical Glycobiology Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom
[c] Proteomics Science Technology Platform, The Francis Crick Institute, London NW1 1AT, United Kingdom.
[d] Lennard-Jones Laboratory, School of Chemical & Physical Sciences, and Centre for Glycoscience, Keele University, Keele, Staffordshire, ST5 5BG United Kingdom
[e] Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.
[f] Signalling and Structural Biology Laboratory, The Francis Crick Institute, London NW1 1AT, United Kingdom.
[g] Department of Chemistry, Yale University, CT 06511, New Haven, United States.
[h] York Structural Biology Laboratory and York Biomedical Research Institute, Department of Chemistry, University of York, York YO10 5DD, United Kingdom
[i] Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen; Copenhagen-N, Denmark.
[j] Current address: GlycoDisplay ApS, Copenhagen, Denmark.

Abstract

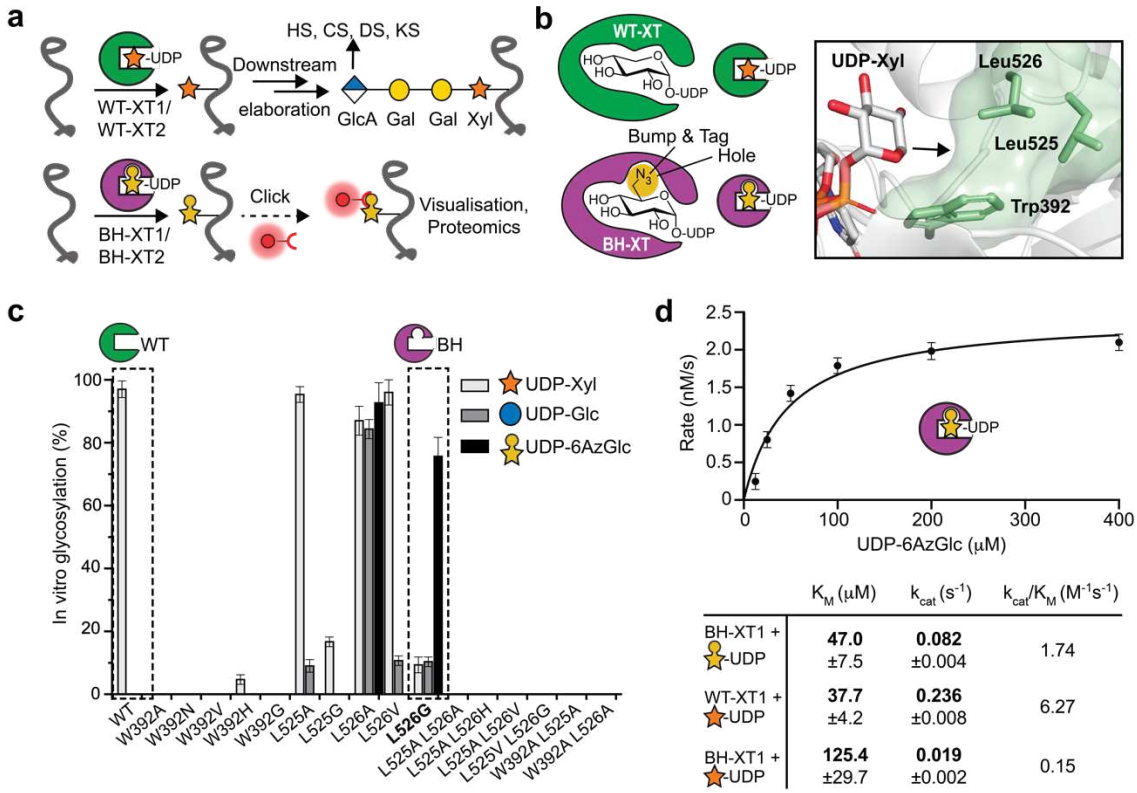
Mammalian cells orchestrate signalling through interaction events on their surfaces. Proteoglycans are an intricate part of these interactions, carrying large glycosaminoglycan polysaccharides that recruit signalling molecules. Despite their importance in development, cancer and neurobiology, a relatively small number of proteoglycans have been identified. In addition to the complexity of glycan extension, biosynthetic redundancy in the first protein glycosylation step by two xylosyltransferase isoenzymes XT1 and XT2 complicates annotation of proteoglycans. Here, we develop a chemical strategy that allows profiling of cellular proteoglycans. By employing a tactic termed bump-and-hole engineering, we engineer the two isoenzymes XT1 and XT2 to specifically transfer a chemically modified xylose analogue to target proteins. The chemical modification contains a bioorthogonal tag, allowing the ability to visualise and profile target proteins modified by both transferases in

mammalian cells as well as pinpointing glycosylation sites by mass spectrometry. Engineered XT enzymes permit a view into proteoglycan biology that is orthogonal to conventional techniques in biochemistry.

Main

Proteoglycans are large biomolecules that consist of a core protein and one or more glycosaminoglycan (GAG) modifications. Ubiquitous on cell surfaces and within the extracellular matrix in higher eukaryotes, proteoglycans form a myriad of interactions with cellular receptors as well as soluble signalling molecules, and provide structural support in connective tissues such as cartilage.[1,2] Growth factors, neurotrophic factors and chemokines can be recruited to target cells through GAG binding sites, rendering proteoglycans important determinants for development.[3,4] Consequently, dysfunctions in GAG biosynthesis cause severe phenotypes from embryonic lethality to skeletal and muscular deficiencies.[5] Binding events between proteoglycans and their receptors are impacted by the core protein as well as the identity of GAG polysaccharides that are classified into either heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS) or keratan sulfate (KS).[6,7] Biochemistry and genetic engineering have linked proteoglycan physiology to the GAG structures on particular cell types or even on distinct subcellular locations.[8–13] Despite their relevance in physiology, only a relatively small number of less than 50 proteoglycans is known in humans.[14,15]

An impediment for the identification of proteoglycans is the large size of GAG modifications that renders analysis by mass spectrometry (MS) challenging. While GAG-carrying glycopeptides contain common amino acid signatures such as acidic patches and a central O-glycosylated Ser with often flanking Gly or Ala residues, there is no consensus sequence to predict GAG glycosylation in the Golgi.[6,9,16,17] Common strategies to identify proteoglycans feature enzymatic digestion of GAG chains either before or after isolation of glycopeptides.[9,18–23] While powerful, such procedures make use of complex digestion and purification protocols and focus solely on the GAG-carrying glycopeptide, without the advantages of shotgun (glyco-)proteomics methods that employ the full MS peptide coverage of individual proteins for detection.



	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
BH-XT1 + UDP	**47.0** ±7.5	**0.082** ±0.004	1.74
WT-XT1 + UDP	**37.7** ±4.2	**0.236** ±0.008	6.27
BH-XT1 + UDP	**125.4** ±29.7	**0.019** ±0.002	0.15

Fig. 1: Design of a xylosyltransferase bump-and-hole system. **a**, principle of the BH approach. WT-XTs transfer Xyl to substrate proteins that can be extended to GAG chains. BH-XTs transfer a bioorthogonal Xyl analogue for visualisation and MS profiling. **b**, structural considerations of XT1 engineering to accept UDP-6AzGlc instead of UDP-Xyl. Insert: Gatekeeper residues in the XT1 crystal structure (PDB 6EJ7) and structural trajectory of 6-azidomethyl modification in UDP-6AzGlc. **c**, *in vitro* glycosylation of a fluorescently labelled bikunin substrate peptide by WT- or mutant XT1 and different UDP-sugars. Data are means ± SD from three technical replicates from one out of two independent experiments. **d**, Michaelis-Menten kinetics of *in vitro* peptide glycosylation by different XT1/UDP-sugar combinations. Data are means ± SD from three technical replicates.

The biosynthesis of HS and CS commences via a common O-linked glycan "linker" modification consisting of a glucuronic acid (GlcA), two galactoses (Gal) and a xylose (Xyl) in the GlcA(β-3)Gal(β-3)Gal(β-4)Xyl(β-)Ser sequence (Fig. 1a), with optional modifications such as phosphorylation on the core Xyl.[7,24] The first glycosylation step attaching Xyl to Ser is subject to biosynthetic redundancy by the xylosyltransferase isoenzymes XT1 and XT2 that utilize uridine diphosphate (UDP)-Xyl as a substrate. The isoenzymes share 60% amino acid identity but display tissue-specific expression patterns and dysfunctions are associated with different genetic disorders - Desbuquois Dysplasia type 2 and Spondylo-Ocular Syndrome for patient *XYLT1* and *XYLT2* mutations, respectively.[25–28] Differential roles in physiology have been attributed to XT1 and XT2.[29,30] Although XT2 appears to be the dominant isoenzyme in cell lines and serum,[31,32] *Xylt1* and *Xylt2* knockout (KO) mice display

differential defects in development.[29,30] Despite their importance in physiology and the unanswered questions about substrate repertoires, it is not possible to directly profile the substrate proteins or even individual glycosylation sites of XT isoenzymes.

Here, we employ a chemical biology tactic termed bump-and-hole (BH) engineering to probe the substrates of human xylosyltransferases in living cells. Based on structural considerations, we mutate a bulky amino acid in the active site of XT1 to a smaller residue to accept a chemically modified UDP-sugar that is not accepted by the wildtype (WT) enzyme. The chemical modification contains an azide group for bioorthogonal incorporation of fluorophores or biotin (Fig. 1a). After in-depth biochemical characterisation, we install the XT1 BH system in mammalian cells to directly visualise and probe proteoglycans. Using MS glycoproteomics, we find confirm that BH-engineered XT1 modifies the native glycosylation site of a model proteoglycan in living mammalian cells. We further show that BH engineering can be applied to the isoenzyme XT2, allowing differential glycoproteomics in future applications. By developing an XT BH system, we simultaneously gain the ability to find new proteoglycans and define the protein substrate specificities of XT isoenzymes.

Results

Design of a xylosyltransferase bump-and-hole system

Our XT BH design was prompted by biosynthetic and structural considerations (Fig. 1b). In the absence of an acetamide group that has featured in numerous tool development approaches of other sugars,[33,34] the most common approach to develop bioorthogonal reporters of monosaccharides is the replacement of hydroxyl with azido groups.[35–37] Beahm et al. developed a 4-azido-substituted Xyl analogue that is incorporated into proteoglycans by XT1,[38] but the corresponding UDP-sugar could not be biosynthesized in mammalian cells since there is no salvage pathway for Xyl.[39] The bump-and-hole tactic uses substrate analogues that would normally not be accepted by GTs.[40–42] We sought to employ this feature to our benefit and reprogram XTs to accept a UDP-sugar that is not accepted by WT-XT1 but can be biosynthesized in mammalian cells. WT-XT1 has been reported to use UDP-6AzGlc with approx. 20-fold lower enzymatic efficiency than UDP-Xyl.[43] We opted to develop a mutant with reversed selectivity to accept UDP-6AzGlc over UDP-Xyl. We recently reported the crystal structure of XT1, revealing a two-lobe architecture.[16] Since XT1 contains an unusually constricted UDP-Xyl binding site that prevents the use of larger UDP-sugars such as UDP-6AzGlc, we deemed it possible to generate additional space (a “hole”) in the active site by mutation. We identified several bulky “gatekeeper” amino acids in close

proximity to C-5 of UDP-Xyl, namely Trp392, Leu525 and Leu526 (Fig. 1b). We designed, expressed and purified from Expi293 cells a total of 16 XT1 single and double mutants in

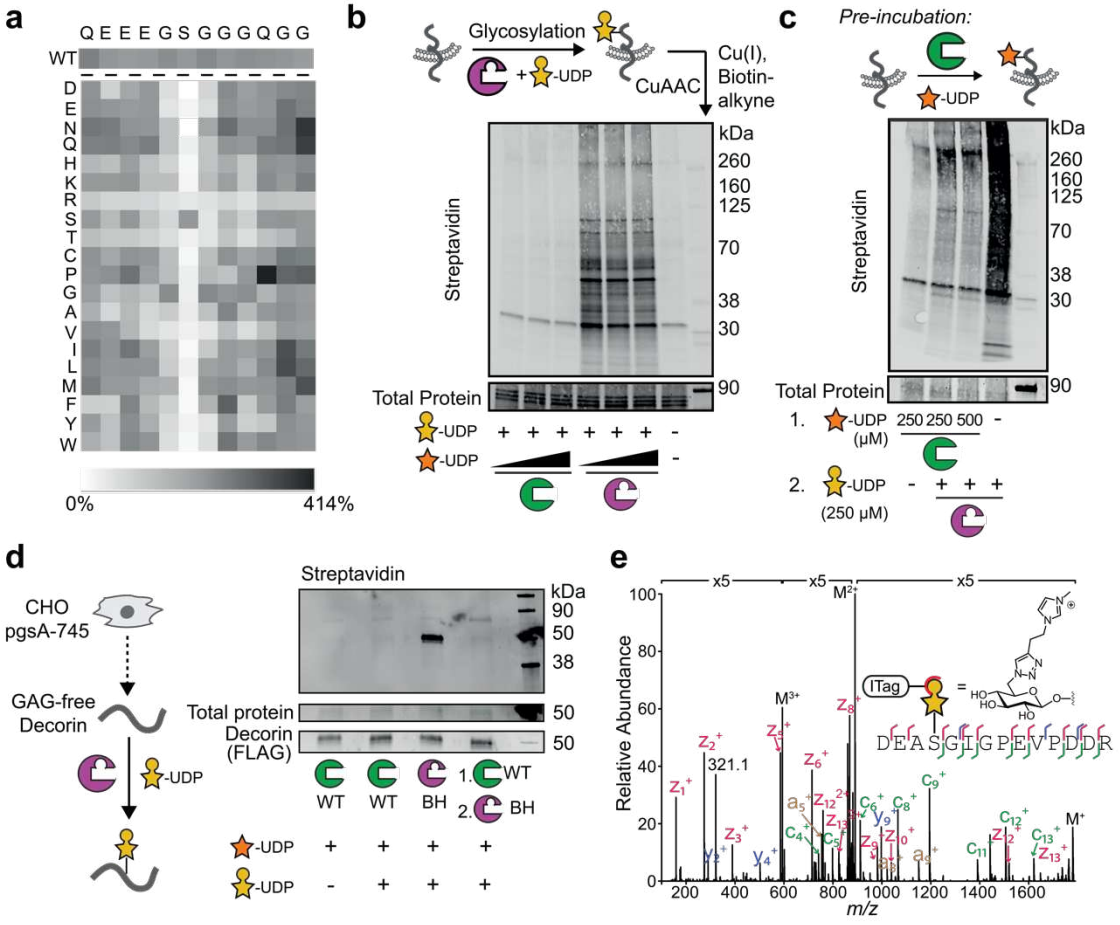


Fig. 2: BH engineering preserves protein substrate specificity of XT1. **a**, *in vitro* glycosylation of a peptide substrate panel based on the bikunin peptide indicated at the top of the panel as assessed by a luminescence assay, with every position substituted for each of the 20 amino acids. Intensity of grey scale indicates % turnover. The WT bikunin peptide was present 12 times in the panel (copied in the top row) and all other peptide reactions were normalised on the average of these 12 data points. Data are from one out of three independent experiments. **b**, *in vitro* glycosylation of a membrane protein fraction of XT2-KO CHO$^{KO\ Xyl2t}$ cells as assessed by streptavidin blot. Reactions contained 250 μM UDP-6AzGlc and 100, 200 or 300 μM UDP-Xyl, respectively, and were reacted with biotin-alkyne before blotting. Data are from one out of two independent experiments. **c**, glycosylation by BH-XT1/UDP-6AzGlc can be prevented by pre-incubation with WT-XT1/UDP-Xyl. Reactions were processed as in **b**. Data are from one experiment. **d**, *in vitro* glycosylation of a GAG-free preparation of human decorin purified from pgsA-745 CHO cells as assessed by streptavidin blot. Reactions were run with 250 μM UDP-sugars and processed as in **b** and **c**. Data are from one out of two independent experiments. **e**, analysis of the glycosylation site on decorin introduced by BH-XT1 by mass spectrometry and ETD fragmentation. Decorin was *in vitro* glycosylated as in **c**, subjected to CuAAC with ITag-azide,[44] digested and subjected to MS-glycoproteomics. Fragments are annotated on the tryptic peptide from mature decorin. 321.1 denotes a signature ion from chemically modified sugar. Data are from one experiment.

which these residues were replaced with smaller amino acids (Fig. 1c, Supporting Fig. 1). We assessed *in vitro* glycosylation of a well-known bikunin substrate peptide in an HPLC-based assay to assess glycosylation from the sugar donors UDP-Xyl, UDP-6AzGlc and, as a substrate of intermediate size of the "bump", UDP-glucose.[16]

WT-XT1 displayed exclusive activity for UDP-Xyl in our hands, in contrast to recently reported residual use of UDP-6AzGlc.[43] Most mutants displayed either no activity at all or were still selective for UDP-Xyl, with some displaying low activity toward UDP-Glc. Strikingly, the mutant Leu526Gly preferred UDP-6AzGlc as a substrate, with 7-8-fold higher turnover than using UDP-Xyl or UDP-Glc in an endpoint assay. Compared to the Leu526Gly mutant (henceforth termed "BH-XT1"), the construct Leu526Ala displayed no such selectivity, with equal activity on all three UDP-sugars (Fig. 1c). These data highlight the need for a detailed structure-function investigation in the design of bump-and-hole mutants. We next determined the kinetic constants for the native and BH enzyme-substrate pairs (Fig. 1d, Supporting Figs 2, 3). We found that the K_M of the BH pair was conserved compared to the WT pair, while the k_{cat} was reduced threefold. In contrast, BH-XT1 uses UDP-Xyl with an approx. 10-fold lower catalytic efficiency than UDP-6AzGlc, indicating that the native substrate UDP-Xyl might not be able to outcompete UDP-6AzGlc in cells. Taken together, we established a sensitive structure-activity relationship in the development of a suitable XT1 bump-and-hole mutant. BH-XT1 fulfilled the crucial pre-requisite of preferring a chemically modified substrate that is not used by the WT enzyme.

Bump-and-hole engineering retains the peptide specificity of WT-XT1

To assess whether BH engineering retains the peptide substrate preference of WT-XT1, we tested the bump-and-hole enzyme-substrate pair (BH-XT1, UDP-6AzGlc) with a panel of 240 substrate peptides in an *in vitro* glycosylation assay. The panel contained derivatives of the well-characterised bikunin XT1 substrate peptide in which each amino acid was substituted with each of the 20 proteinogenic amino acids. We had previously used the same peptide panel to extract peptide substrate preferences of the native enzyme-substrate pair (WT-XT1, UDP-Xyl) in a luminescence-based assay.[16] We found that the peptide substrate preferences were remarkably conserved between WT- and BH-XT1 (Fig. 2a, for WT peptide preference see Briggs and Hohenester[16]). For instance, introducing basic Lys or Arg residues anywhere in the substrate peptide lowered enzyme activity, while acidic Glu and Asp tended to increase activity. A notable data point was the swap of Glu at -4 position to Asp which led to a decrease in turnover for both WT- and BH-XT1.[16] As in WT-XT1, substitutions of glycine residues at positions -1 and +1 of the central Ser were not well-tolerated by BH-XT1. An

exception was substitution of Gly at +1 position to hydrophobic amino acids Leu, Met or Phe, which led to residual activity in BH- but not WT-XT1. Since this Gly is in contact with Leu526 in WT-XT1, we reasoned that the Leu526Gly "hole" left space for substitutions to larger hydrophobic amino acids in substrate peptides. Since these were the only reproducible differences between WT and BH-XT1 (other replicates in Supporting Fig. 4) and relatively low in number in a 240-member peptide library, we concluded that BH engineering exhibits conservation of peptide substrate preference *in vitro*.

BH-XT1 glycosylates proteoglycans at GAG attachment sites *in vitro*

We next assessed whether BH-XT1 retains the activity of WT-XT1 to prime GAG attachment sites on proteoglycan backbones. We prepared membrane fractions from Chinese Hamster Ovary (CHO) cells with or without KO for endogenous xylosyltransferase genes *Xylt1* and *Xylt2*.[10] Membrane fractions were incubated with recombinant WT- or BH-XT1 as well as synthetic UDP-Xyl and UDP-6AzGlc, followed by reaction with alkyne-biotin under copper-catalysed azide-alkyne cycloaddtion (CuAAC) "click" conditions. Analysis by streptavidin blot suggested labelling of lysate proteins with 6AzGlc only when BH-XT1, but not WT-XT1 was present (Supporting Fig. 5). Since *Xylt2* is the major xylosyltransferase gene expressed in CHO cells,[10] we used CHO$^{KO\ Xylt2}$ cells for further *in vitro* glycosylation experiments. We established that labelling by BH-XT1 with UDP-6AzGlc could not be outcompeted with increasing concentrations of UDP-Xyl, suggesting that BH-XT1 specifically and potently recognises UDP-6AzGlc as a substrate (Fig. 2b). Pre-incubation of the membrane protein fraction with WT-XT1 and UDP-Xyl abrogated incorporation of 6AzGlc by BH-XT1, suggesting that the same glycosylation sites are introduced by both enzymes (Fig. 2c).

We next confirmed *in vitro* that BH-XT1 emulates the activity of WT-XT1 to glycosylate proteoglycans. Human decorin has a single site of GAG attachment. Recombinant expression in the CHO cell mutant pgsA-745 that lacks endogenous XT activity results in a GAG-free decorin preparation.[7,8] We incubated this GAG-free decorin with either WT- or BH-XT1 in the presence of UDP-Xyl and/or UDP-6AzGlc, followed by CuAAC with alkyne-biotin and streptavidin blot. (Fig. 2d). While WT-XT1 activity did not lead to discernible streptavidin signal on decorin, BH-XT1 in the presence of UDP-6-AzGlc led to intense streptavidin signal that could be abrogated by pre-incubation of decorin with WT-XT1 and UDP-Xyl. These data indicate that the single GAG attachment site was blocked with a xylose residue by WT-XT1, preventing BH-XT1 activity. We observed the same behaviour in GAG-free preparation of human glypican 1 (Supporting Fig. 6), suggesting that BH-XT1 recapitulates the activity of WT-XT1 across a range of proteoglycans.

We confirmed the glycosylation site modified by BH-XT1 on decorin by tandem mass spectrometry (MS). Two fragmentation methods are routinely employed for O-glycopeptides.

High energy collision-induced dissociation (HCD) primarily fragments the glycosidic bond to detect glycan oxonium ions while electron transfer dissociation (ETD) fragments the peptide backbone to allow glycan site annotation. The clickable azide tag was essential to improve sugar identification in mass spectra, allowing incorporation of functional groups that are beneficial to analysis. Specifically, we employed a clickable imidazolium tag (ITag) that carries a permanent positive charge and increased the charge state of glycopeptides, allowing direct glycosylation site annotation.[44] We used HCD first to fragment decorin-derived (glyco-)peptides (Supporting Fig. 7a) and upon detection of an ITag-containing, 6AzGlc-derived signature ion, triggered ETD on the same glycopeptide.[34,40,42,44]

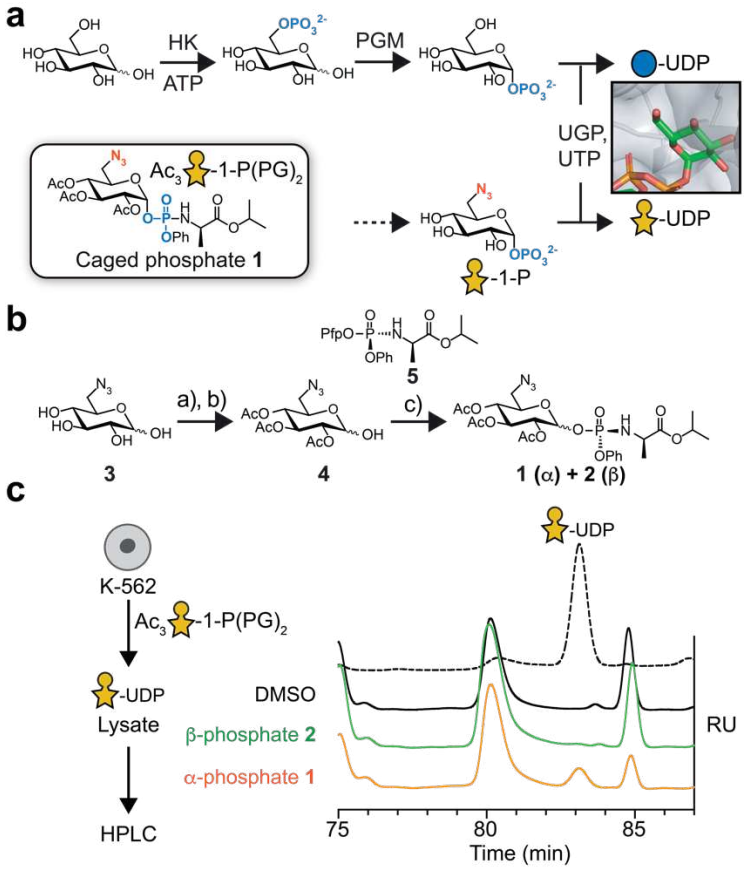


Fig. 3: Biosynthesis of UDP-6AzGlc. **a**, biosynthetic pathway of UDP-Glc via hexokinase (HK), phosphoglucomutase (PGM) and UDP-Glc pyrophosphorylase (UGP). Biosynthesis of UDP-6AzGlc from caged phosphate **1** bypasses the HK and PGM steps. Insert: crystal structure of UGP with UDP-Glc indicating that the 6-OH group protrudes into an open cavity (PDB 4R7P). **b**, synthesis of caged sugar-1-phosphates **1** and **2** from 6AzGlc **3**. **c**, biosynthesis of UDP-6AzGlc in K562 cells as assessed by ion exchange HPLC of lysates fed with compounds **1** or **2**. Data are from one out of two independent experiments. Reagents and conditions: a) Ac_2O, pyridine, DMAP, r.t., 90%; b) AcOH, ethylene diamine, r.t. 45-73%; c) 2 M LDA in THF, **5**, -78°C to -70°C, (**1**) 60% (**2**) 9.8%. DMAP = 4-Dimethylaminopyridine; LDA = lithium diisopropylamide. RU = relative units.

Decorin is proteolytically processed during secretion to remove a propeptide and shorten the N-terminus.[45] ETD allowed for direct identification of Ser34 as the attachment site of 6AzGlc by BH-XT1 on this mature form, consistent with Ser34 being the site of cellular GAG attachment (Fig. 2e).[46] Taken together, these results suggest that the XT1 bump-and-hole enzyme-substrate pair glycosylates native GAG attachment sites in proteoglycans *in vitro*.

Development of a cellular XT1 bump-and-hole system

Application of a GT BH system in living cells requires biosynthesis of the nucleotide-sugar. In general, caged, membrane-permeable monosaccharide precursors are employed with ester modifications that are deprotected in the cytosol. Free monosaccharides can then be converted to UDP-sugars before transport to the Golgi.[41,42,47] Although human cells are devoid of a salvage pathway for UDP-Xyl, our strategic use of UDP-6AzGlc provided an opportunity for cellular application by hijacking the biosynthetic pathway for UDP-Glc instead. Glc is activated in mammalian cells first by phosphorylation to Glc-6-phosphate and, subsequently, isomerization by phosphoglucomutase PGM to Glc-1-phosphate (Fig. 3a). Conversion to UDP-Glc then features the enzymes UDP-Glc pyrophosphorylase 1 or 2, UGP1/2. Since phosphorylation at the 6-position was prevented by the presence of an azido group, we sought to bypass the kinase and PGM steps and provide a sugar-1-phosphate as a direct substrate for UGP1/2. We were encouraged by analysis of the UGP1/UDP-Glc co-crystal structure where the UDP-Glc 6-hydroxyl group is solvent-exposed, suggesting that an azido group at that position should be tolerated by the enzyme (Fig. 3a).[48] While we and others have made sugar-1-phosphates caged as labile bis-*S*-acetylthioethyl (SATE) phosphotriesters, synthesis of SATE-caged 6AzGlc-1-phosphate failed in our hands.[42,49,50] Instead, we took inspiration from the increasingly popular protide technology that has gained attention to cage phosphates in antiviral nucleotides.[51,52] To our knowledge, such chemistry has not been applied to sugar-1-phosphates yet, but for instance to sugar-6-phosphates.[53,54] We synthesized phosphoamidate diester **1** to be deprotected by esterases in the cytosol (Fig. 3a).[51] The synthesis proceeded from lactol mixture **3** via the intermediate triacetate **4.** Treatment of **4** with **5** under basic conditions yielded both α-phosphate **1** (60% yield) and β-phosphate **2** (9.8% yield) (Fig. 3b).[55] Since UGP1/2 is naturally restricted to a-configured Glc-1-phosphate, β-phosphate **2** served as a negative control in feeding experiments. Feeding K-562 cells the α-phosphate **1** led to notable and reproducible biosynthesis of UDP-6AzGlc (Fig. 3c). In turn, β-configured **2** led to negligible UDP-6AzGlc levels. These data suggest that α-phosphate **1** is a suitable precursor to deliver UDP-6AzGlc to mammalian cells by entering the UDP-Glc biosynthetic pathway.

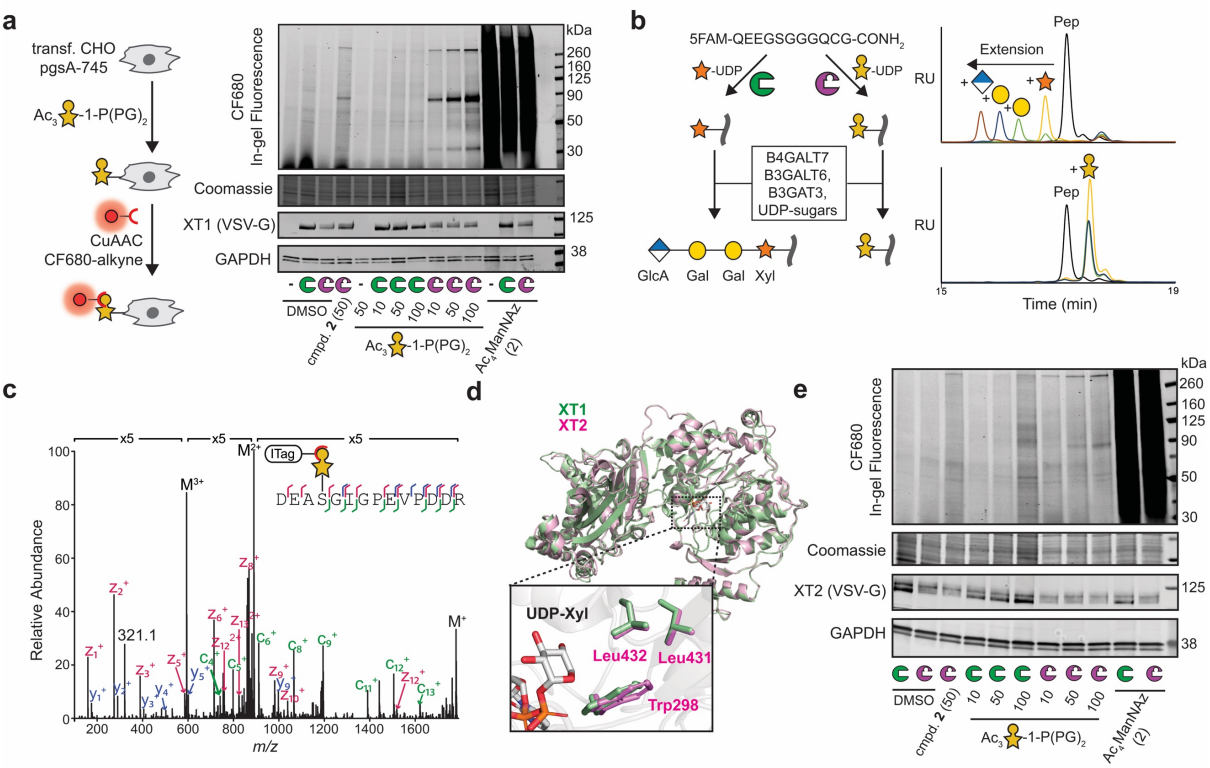


Fig. 4: BH-engineered xylosyltransferases label proteoglycans in mammalian cells. **a**, chemical tagging of proteoglycans on pgsA-745 CHO cells as assessed by in-gel fluorescence. Cells stably expressing WT- or BH-XT1 or non-transfected were fed with compounds in the indicated concentrations in μM before on-cell CuAAC. Data are from one out of two independent replicates. **b**, 6AzGlc is not elongated by GAG linker enzymes. Fluorescently labelled bikunin-derived peptide was incubated with WT-XT1/UDP-Xyl or BH-XT1/UDP-6AzGlc and subsequently with the indicated soluble glycosyltransferases[7] and corresponding UDP-sugars. Elongation was assessed on each step by HPLC. Data are from one experiment each with different combinations of transferases. **c**, analysis of the glycosylation site on decorin introduced by BH-XT1 in living cells by mass spectrometry and ETD fragmentation. Decorin was co-expressed with BH-XT1 in cells fed with **1**, then subjected to CuAAC with ITag-azide,[44] digested and subjected to MS-glycoproteomics. Fragments are annotated on the tryptic peptide from mature decorin. Not all fragments are shown. 321.1 denotes a signature ion from chemically modified sugar. Data are from one experiment. **d**, structural alignment between the crystal structure of human XT1 (PDB 6EJ7, green) and the AlphaFold structure of human XT2 (accession no. Q9H1B5, purple) with aligned gatekeeper residues in the insert. **e**, BH-XT2 chemically tags proteoglycans on mammalian cells. Cells stably transfected with WT- or BH-XT2 were fed and treated as in **a**. Data are from one out of two independent experiments.

In accordance with biosynthetic experiments, the β-configured 6AzGlc-1-phosphate **2** yielded a weak and diffuse labelling signal, indicating that UDP-6AzGlc biosynthesis is a direct prerequisite for cellular chemical tagging of glycoproteins by BH-XT1.

We next employed UDP-6AzGlc delivery to establish a cellular XT1 bump-and-hole system. Stable transfection of pgsA-745 CHO cells with WT- or BH-XT1 was followed by feeding the 6AzGlc-1-phosphate precursor **1**. After overnight incubation, a CuAAC reaction was performed to attach clickable alkyne-CF680 on the cell surface while keeping cells alive.[34,42,47] Surplus click reagents were washed away, cells lysed and fluorophore incorporation assessed by SDS-PAGE and in-gel fluorescence (Fig. 4a). Minimal background fluorescence was observed in cells fed with DMSO or only expressing WT-XT1, even when fed with increasing concentrations of 6AzGlc-1-phosphate precursor **1**. In the presence of BH-XT1, clear fluorescent proteins were observed at 38 kDa, 90 kA and 260 kDa. With increasing feeding concentration of **1**, a dose-dependent increase of fluorescence was observed, along with labelled protein bands of lower intensity, especially at 50 kDa.

XT bump-and-hole engineering identifies proteoglycans in mammalian cells

Xylosyltransferase BH engineering is poised to allow the identification of new proteoglycans, a feat that normally requires elaborate methods of glycopeptide enrichment and characterisation.[9,20,22] To establish an MS-glycoproteomics workflow, it was important to assess whether 6AzGlc, like Xyl, was extended to a functional GAG linker tetrasaccharide. We recently reported an enzymatic method for extension of xylosylated glycopeptides by recombinant preparations of the glycosyltransferases B4GALT7, B3GALT6 and B3GAT3 (termed linker enzymes) in the presence of UDP-galactose (UDP-Gal) and UDP-glucuronic acid (UDP-GlcA). Employing a fluorescently labelled bikunin-derived peptide, we first confirmed by HPLC that the linker enzymes extend a Xyl moiety introduced by WT-XT1 to the full tetrasaccharide. In contrast, a 6AzGlc-modified peptide did not shift in retention time upon incubation with the GTs and UDP-sugars (Fig. 4b). We concluded that 6AzGlc is a chain-terminating modification that is not extended to functional GAG chains. We interpreted this feature is an advantage for the discovery of proteoglycans by MS due to the substantially decreased complexity of the modification that is convenient to analyse. To this end, we next performed an MS-glycoproteomics experiment to identify the glycosylation site of a model proteoglycan in living cells. FLAG-tagged human decorin was overexpressed in pgsA-745 CHO cells that expressed BH-XT1 and were fed with 6AzGlc-1-phosphate precursor **1**. Following immunoprecipitation, decorin was subjected to a CuAAC reaction with an ITag-azide, digested and subjected to mass spectrometry by HCD-triggered ETD fragmentation. We confirmed unambiguously that Ser34 was glycosylated by BH-XT1 inside mammalian cells, confirming the bump-and-hole approach as suitable to identify proteoglycans including native Xyl attachment sites (Supporting Fig. 7b, Fig. 4c).

It is currently not known why the mammalian genome encodes two xylosyltransferase isoenzymes. To distinguish their protein substrate profiles, we extended the BH method from XT1 to the second isoenzyme xylosyltransferase 2. A structural overlay between the XT1 crystal structure and the XT2 AlphaFold2 structure highlighted conservation of amino acids interacting with UDP-Xyl (Fig. 4d). In parallel to XT1, we cloned the BH-XT2 mutant L432G and stably transfected pgsA-745 CHO cells with either WT- or BH-XT2. Feeding with 6AzGlc-1-phosphate precursor **1** and cell-surface CuAAC reaction indicated that BH-XT2 labels a similar protein repertoire as BH-XT1 but at lower intensity (Fig. 4e).

Discussion

The importance of proteoglycans in physiology is undisputed, as the vast majority of signalling events between cells or with the extracellular matrix are modulated by the associated GAG chains. While great efforts are being made to understand the details of GAG polysaccharide sequence on biology,[10,14,21,56,57] we still lack important information on the first step of glycosylation to the protein backbone. The two human xylosyltransferases display limited tissue selectivity and differences in attached GAG sequences but we still lack fundamental detail on their individual biological functions.[27,30–32,58] Furthermore, only a relatively small number of annotated mammalian proteoglycans has been mapped, with new annotations requiring considerable effort.[9] A chemical tool to dissect XT1/2 biology must accurately report on XT1/2 activity while being orthogonal to other glycosylation events in the secretory pathway and deliverable to living mammalian cells. The use of a UDP-Glc analogue matched these pre-requisites. Both catalytic efficiency and peptide substrate preference of the XT1 bump-and-hole enzyme-substrate pair were remarkably conserved, which we attribute to the careful structure-based design of the pair. Labelling signal was dependent on the presence of BH-XT1, indicating that 6AzGlc does not enter other major glycosylation pathways. The finding that 6AzGlc is not extended to the common GAG linker tetrasaccharide was an advantage for mass spectrometry since the sugar is structurally well-defined and did not require any glycosidase treatment steps prior to analysis. Annotation was further simplified by the availability of the ITag technology to facilitate mass spectrometry. Furthermore, a chain-terminating, clickable inhibitor of chain extension has the potential to be employed to study GAG biology *in vitro* or *in vivo*,[38,59,60] substantially expanding our toolbox.

Establishing a cellular XT bump-and-hole system required a biosynthetic entry point for UDP-6AzGlc. Previously used per-acetylated 6AzGlc was not a suitable precursor for glycosylation in our hands,[35] but we note that cell lines from different organisms can vary in their biosynthetic potential.[61] Nevertheless, a protide-based caged sugar-1-phosphate was a reliable precursor for UDP-6AzGlc to fashion a cellular bump-and-hole system. We did not

assess whether UDP-6AzGlc was used by other glycosyltransferases, but cell surface labelling experiments suggested that BH-XT activity was necessary to introduce the chemical modification into glycoproteins.

While XT2 appears to be the dominant isoenzyme expressed in humans, dysfunctions in both enzymes lead to severe yet differential disorders in mouse models and in patients.[29–32] After fully characterising an XT1 BH system, we designed a functional BH-XT2 mutant simply based on structural homology, showcasing the reliability of the tactic as well as the importance of structural data. Our work will establish the fine differences between XT1 and XT2 and identify proteoglycans in a range of different model systems.

Acknowledgements

We thank Ganka Bineva-Todd and Andrea Marchesi for help with chemical synthesis and analysis, and Khadija Babiker for help with cloning similar constructs. We thank Chloë Roustan, Svend Kjaer and the Francis Crick Institute Structural Biology Science Technology Platform for help with protein expression and purification. We further thank Tania Auchynnikava and Mark Skehel for help with mass spectrometry, and the Crick Proteomics, Chemical Biology and Cell Services Science Technology Platforms for valuable support. This work was supported by the Francis Crick Institute (to B. S.) which receives its core funding from Cancer Research UK (CC2127, CC2068), the UK Medical Research Council (CC2127, CC2068) and Wellcome Trust (CC2127, CC2068), the BBSRC (BB/T01279X/1 to E. H. and B. S., BB/V008439/1 to B. S.), the EPSRC (EP/T007397/1 to G.J.M) and the MRC (MR/T019522/1 to G.J.M). We also thank the EPSRC UK National Mass Spectrometry Facility (NMSF) at Swansea University. This work was supported by the Novo Nordisk Foundation grant No. NNF22OC0073736 (to R. L. M.). L. I. W. gratefully acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 851448). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Xu, D. & Esko, J. D. Demystifying Heparan Sulfate–Protein Interactions. *Annu Rev Biochem* **83**, 129–157 (2014).
2. Bishop, J. R., Schuksz, M. & Esko, J. D. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* **446**, 1030–1037 (2007).
3. Pickford, C. E. *et al.* Specific Glycosaminoglycans Modulate Neural Specification of Mouse Embryonic Stem Cells. *Stem Cells* **29**, 629–640 (2011).
4. Kreuger, J. *et al.* Fibroblast growth factors share binding sites in heparan sulphate. *Biochem J* **389**, 145–150 (2005).

5. Mizumoto, S. & Yamada, S. Congenital Disorders of Deficiency in Glycosaminoglycan Biosynthesis. *Front Genet* **12**, 717535 (2021).
6. Zhang, L. & Esko, J. D. Amino acid determinants that drive heparan sulfate assembly in a proteoglycan. *J of Biol Chem* **269**, 19295–19299 (1994).
7. Sammon, D. *et al.* Molecular mechanism of decision-making in glycosaminoglycan biosynthesis. *Nat Commun 2023* **14**, 1–13 (2023).
8. Esko, J. D., Stewart, T. E. & Taylor, W. H. Animal cell mutants defective in glycosaminoglycan biosynthesis. *Proc Natl Acad Sci U S A* **82**, 3197–3201 (1985).
9. Zhang, P. *et al.* Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships. *Cell* **174**, 1450-1464.e23 (2018).
10. Chen, Y. H. *et al.* The GAGOme: a cell-based library of displayed glycosaminoglycans. *Nat Methods* **15**, 881–888 (2018).
11. Qiu, H. *et al.* A mutant-cell library for systematic analysis of heparan sulfate structure–function relationships. *Nat Methods 2018* **15**, 889–899 (2018).
12. Ji, S. K. *et al.* PTPσ functions as a presynaptic receptor for the glypican-4/LRRTM4 complex and is essential for excitatory synaptic transmission. *Proc Natl Acad Sci U S A* **112**, 1874–1879 (2015).
13. Kamimura, K. *et al.* Perlecan regulates bidirectional Wnt signaling at the Drosophila neuromuscular junction. *J Cell Biol* **200**, 219–233 (2013).
14. Merry, C. L. R., Lindahl, U., Couchman, J. & Esko, J. D. Proteoglycans and Sulfated Glycosaminoglycans. *Essentials of Glycobiology* (2022)
15. Noborn, F. & Sterky, F. H. Role of neurexin heparan sulfate in the molecular assembly of synapses – expanding the neurexin code? *FEBS J* **290**, 252–265 (2023).
16. Briggs, D. C. & Hohenester, E. Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* **26**, 801-809.e3 (2018).
17. Zhang, L., David, G. & Esko, J. D. Repetitive Ser-Gly Sequences Enhance Heparan Sulfate Assembly in Proteoglycans. *J Biol Chem* **270**, 27127–27135 (1995).
18. Ramarajan, M. G. *et al.* Mass spectrometric analysis of chondroitin sulfate-linked peptides. *J Protein Proteomics 2022* **13**, 187–203 (2022).
19. Ly, M., Laremore, T. N. & Linhardt, R. J. Proteoglycomics: Recent progress and future challenges. *OMICS* **14**, 389–399 (2010).
20. Noborn, F., Nilsson, J. & Larson, G. Site-specific glycosylation of proteoglycans: A revisited frontier in proteoglycan research. *Matrix Biology* vol. 111 289–306 (2022).
21. Persson, A., Nikpour, M., Vorontsov, E., Nilsson, J. & Larson, G. Domain Mapping of Chondroitin/Dermatan Sulfate Glycosaminoglycans Enables Structural Characterization of Proteoglycans. *Mol Cell Proteom* **20**, 100074 (2021).
22. Noborn, F. *et al.* A Glycoproteomic Approach to Identify Novel Proteoglycans. *Meth Mol Biol* **2303**, 71–85 (2022).
23. Noborn, F. *et al.* Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. *Sci Rep* **6**, 1–11 (2016).
24. Wen, J. *et al.* Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc Natl Acad Sci U S A* **111**, 15723–15728 (2014).
25. Bui, C. *et al.* XYLT1 mutations in desbuquois dysplasia type 2. *Am J Hum Genet* **94**, 405–414 (2014).
26. Munns, C. F. *et al.* Homozygosity for frameshift mutations in XYLT2 result in a spondylo-ocular syndrome with bone fragility, cataracts, and hearing defects. *Am J Hum Genet* **96**, 971–978 (2015).
27. Wilson, I. B. H. The never-ending story of peptide O-xylosyltransferase. *Cell Mol Life Sci* **61**, 794–809 (2004).
28. Götting, C., Kuhn, J., Zahn, R., Brinkmann, T. & Kleesiek, K. Molecular cloning and expression of human UDP-D-xylose: Proteoglycan core protein β-D-xylosyltransferase and its first isoform XT-II. *J Mol Biol* **304**, 517–528 (2000).
29. Ferencz, B. *et al.* Xylosyltransferase 2 deficiency and organ homeostasis. *Glycoconj J* **37**, 755 (2020).

30. Taieb, M., Ghannoum, D., Barré, L. & Ouzzine, M. Xylosyltransferase I mediates the synthesis of proteoglycans with long glycosaminoglycan chains and controls chondrocyte hypertrophy and collagen fibers organization of in the growth plate. *Cell Death Dis 2023 14:6* **14**, 1–13 (2023).
31. Roch, C., Kuhn, J., Kleesiek, K. & Götting, C. Differences in gene expression of human xylosyltransferases and determination of acceptor specificities for various proteoglycans. *Biochem Biophys Res Commun* **391**, 685–691 (2010).
32. Kuhn, J. *et al.* Xylosyltransferase II is the predominant isoenzyme which is responsible for the steady-state level of xylosyltransferase activity in human serum. *Biochem Biophys Res Commun* **459**, 469–474 (2015).
33. Choi, J. *et al.* Engineering Orthogonal Polypeptide GalNAc-Transferase and UDP-Sugar Pairs. *J Am Chem Soc* **141**, 13442–13453 (2019).
34. Debets, M. F. *et al.* Metabolic precision labeling enables selective probing of O-linked N-acetylgalactosamine glycosylation. *Proc Natl Acad Sci U S A* **117**, 25293–25301 (2020).
35. Darabedian, N., Gao, J., Chuh, K. N., Woo, C. M. & Pratt, M. R. The Metabolic Chemical Reporter 6-Azido-6-deoxy-glucose Further Reveals the Substrate Promiscuity of O-GlcNAc Transferase and Catalyzes the Discovery of Intracellular Protein Modification by O-Glucose. *J Am Chem Soc* **140**, 7092–7100 (2018).
36. Darabedian, N. *et al.* O-Acetylated Chemical Reporters of Glycosylation Can Display Metabolism-Dependent Background Labeling of Proteins but Are Generally Reliable Tools for the Identification of Glycoproteins. *Front Chem* **8**, 529502 (2020).
37. Daughtry, J. L., Cao, W., Ye, J. & Baskin, J. M. Clickable Galactose Analogues for Imaging Glycans in Developing Zebrafish. *ACS Chem Biol* **15**, 318–324 (2020).
38. Beahm, B. J. *et al.* A visualizable chain-terminating inhibitor of glycosaminoglycan biosynthesis in developing zebrafish. *Angew Chem Int Ed Engl* **53**, 3347–3352 (2014).
39. Bakker, H. *et al.* Functional UDP-xylose transport across the endoplasmic reticulum/golgi membrane in a Chinese hamster ovary cell mutant defective in UDP-xylose synthase. *J Biol Chem* **284**, 2576–2583 (2009).
40. Cioce, A. *et al.* Cell-specific bioorthogonal tagging of glycoproteins. *Nat Commun* **13**, 1–18 (2022).
41. Cioce, A., Malaker, S. A. & Schumann, B. Generating orthogonal glycosyltransferase and nucleotide sugar pairs as next-generation glycobiology tools. *Curr Opin Chem Biol* **60**, 66–78 (2021).
42. Schumann, B. *et al.* Bump-and-Hole Engineering Identifies Specific Substrates of Glycosyltransferases in Living Cells. *Mol Cell* **78**, 824-834.e15 (2020).
43. Gao, J. *et al.* Exploration of Human Xylosyltransferase for Chemoenzymatic Synthesis of Proteoglycan Linkage Region. *Org Biomol Chem* **19**, 3374–3378 (2021)
44. Calle, B. *et al.* Benefits of Chemical Sugar Modifications Introduced by Click Chemistry for Glycoproteomic Analyses. *J Am Soc Mass Spectrom* **32**, 2366–2375 (2021).
45. von Marschall, Z. & Fisher, L. W. Decorin is processed by three isoforms of bone morphogenetic protein-1 (BMP1). *Biochem Biophys Res Commun* **391**, 1374–1378 (2010).
46. Seo, N. S., Hocking, A. M., Höök, M. & McQuillan, D. J. Decorin Core Protein Secretion Is Regulated by N-Linked Oligosaccharide and Glycosaminoglycan Additions. *J Biol Chem* **280**, 42774–42784 (2005).
47. Cioce, A. *et al.* Optimization of Metabolic Oligosaccharide Engineering with Ac4GalNAlk and Ac4GlcNAlk by an Engineered Pyrophosphorylase. *ACS Chem Biol* **16**, 1961–1967 (2021).
48. Führing, J. I. *et al.* A Quaternary Mechanism Enables the Complex Biological Functions of Octameric Human UDP-glucose Pyrophosphorylase, a Key Enzyme in Cell Metabolism. *Sci Rep* **5**, 1–11 (2015).

49. Yu, S. H. *et al.* Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proc Natl Acad Sci U S A* **109**, 4834–4839 (2012).
50. Murphy, L. D. *et al.* Synthesis of biolabile thioalkyl-protected phosphates from an easily accessible phosphotriester precursor. *Chem Sci* **14**, 5062–5068 (2023).
51. Mehellou, Y., Rattan, H. S. & Balzarini, J. The ProTide Prodrug Technology: From the Concept to the Clinic. *J Med Chem* **61**, 2211–2226 (2018).
52. Guinan, M., Huang, N., Smith, M. & Miller, G. J. Design, chemical synthesis and antiviral evaluation of 2′-deoxy-2′-fluoro-2′-C-methyl-4′-thionucleosides. *Bioorg Med Chem Lett* **61**, 128605 (2022).
53. Morozzi, C. *et al.* Targeting GNE Myopathy: A Dual Prodrug Approach for the Delivery of N-Acetylmannosamine 6-Phosphate. *J Med Chem* **62**, 8178–8193 (2019).
54. McGuigan, C. *et al.* Phosphate prodrugs derived from N-acetylglucosamine have enhanced chondroprotective activity in explant cultures and represent a new lead in antiosteoarthritis drug discovery. *J Med Chem* **51**, 5807–5812 (2008).
55. Ross, B. S., Ganapati Reddy, P., Zhang, H. R., Rachakonda, S. & Sofia, M. J. Synthesis of diastereomerically pure nucleotide phosphoramidates. *J Org Chem* **76**, 8311–8319 (2011).
56. Barnett, M. W., Fisher, C. E., Perona-Wright, G. & Davies, J. A. Signalling by glial cell line-derived neurotrophic factor (GDNF) requires heparan sulphate glycosaminoglycan. *J Cell Sci* **115** 4495–4503 (2002).
57. Sterner, E., Meli, L., Kwon, S. J., Dordick, J. S. & Linhardt, R. J. FGF-FGFR signaling mediated through glycosaminoglycans in microtiter plate and cell-based microarray platforms. *Biochemistry* **52**, 9009–9019 (2013).
58. Götting, C., Kuhn, J. & Kleesiek, K. Human xylosyltransferases in health and disease. *Cell Mol Life Sci* **64**, 1498–1517 (2007).
59. Maciej-hulme, M. L. *et al.* Selective inhibition of heparan sulphate and not chondroitin sulphate biosynthesis by a small, soluble competitive inhibitor. *Int J Mol Sci* **22**, 6988 (2021).
60. O'Leary, T. R. *et al.* Chemical editing of proteoglycan architecture. *Nat Chem Biol* **18**, 634–642 (2022).
61. Batt, A. R., Zaro, B. W., Navarro, M. X. & Pratt, M. R. Metabolic Chemical Reporters of Glycans Exhibit Cell-Type-Selective Metabolism and Glycoprotein Labeling. *ChemBioChem* **18**, 1177–1182 (2017).