



HAL
open science

Chemo-enzymatic synthesis of tetrasaccharide linker peptides to study the divergent step in glycosaminoglycan biosynthesis

Farah Fouladkar, Margot Weber, Elisabetta Boeri-Erba, Rebekka Wild

► **To cite this version:**

Farah Fouladkar, Margot Weber, Elisabetta Boeri-Erba, Rebekka Wild. Chemo-enzymatic synthesis of tetrasaccharide linker peptides to study the divergent step in glycosaminoglycan biosynthesis. *Glycobiology*, 2024, 34 (5), 10.1093/glycob/cwae016 . hal-04624891

HAL Id: hal-04624891

<https://hal.science/hal-04624891>

Submitted on 25 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Chemo-enzymatic synthesis of tetrasaccharide linker peptides to study the divergent step in glycosaminoglycan biosynthesis

Marie Bourgeais , Farah Fouladkar , Margot Weber , Elisabetta Boeri-Erba ,
Rebekka Wild* 

Institut de Biologie Structurale, UMR 5075, University Grenoble Alpes, CNRS, CEA, 71, Avenue des Martyrs, 38044 Grenoble, France.

*Corresponding author: Rebekka Wild. Email: rebekka.wild@ibs.fr

Glycosaminoglycans are extended linear polysaccharides present on cell surfaces and within the extracellular matrix that play crucial roles in various biological processes. Two prominent glycosaminoglycans, heparan sulfate and chondroitin sulfate, are covalently linked to proteoglycan core proteins through a common tetrasaccharide linker comprising glucuronic acid, galactose, galactose, and xylose moieties. This tetrasaccharide linker is meticulously assembled step by step by four Golgi-localized glycosyltransferases. The addition of the fifth sugar moiety, either N-acetylglucosamine or N-acetylgalactosamine, initiates further chain elongation, resulting in the formation of heparan sulfate or chondroitin sulfate, respectively. Despite the fundamental significance of this step in glycosaminoglycan biosynthesis, its regulatory mechanisms have remained elusive. In this study, we detail the expression and purification of the four linker-synthesizing glycosyltransferases and their utilization in the production of fluorescent peptides carrying the native tetrasaccharide linker. We generated five tetrasaccharide peptides, mimicking the core proteins of either heparan sulfate or chondroitin sulfate proteoglycans. These peptides were readily accepted as substrates by the EXTL3 enzyme, which adds an N-acetylglucosamine moiety, thereby initiating heparan sulfate biosynthesis. Importantly, EXTL3 showed a preference towards peptides mimicking the core proteins of heparan sulfate proteoglycans over the ones from chondroitin sulfate proteoglycans. This suggests that EXTL3 could play a role in the decision-making step during glycosaminoglycan biosynthesis. The innovative strategy for chemo-enzymatic synthesis of fluorescent-labeled linker-peptides promises to be instrumental in advancing future investigations into the initial steps and the divergent step of glycosaminoglycan biosynthesis.

Key words: chemo-enzymatic synthesis; glycosaminoglycans; glycosyltransferase enzymes.

Introduction

Glycosaminoglycans (GAGs) are intricate and long polysaccharides covalently bound to diverse core proteins, collectively forming proteoglycans (PGs). These molecular entities are ubiquitous in nearly all metazoan organisms (Yamada et al. 2011). Depending on their disaccharide unit and modifications, the attached polysaccharide chains categorize into distinct types, including heparan sulfate (HS), heparin (Hep), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS) (Merry et al. 2022). Despite the seemingly limited varieties of Heparan Sulfate Proteoglycans (HSPGs) (20) and Chondroitin Sulfate Proteoglycans (CSPGs) (80), they have been found to interact with an astonishingly extensive array of protein partners, numbering over 3,000 to date (Vallet et al. 2022). These interaction partners include growth factors, cytokines, morphogens, signaling receptors, chemokines, and pathogens, implicating them in a myriad of biological processes (Bishop et al. 2007; Connell et al. 2016). This expansive interactome is mirrored by the intricate structural complexity of GAG chains, which poses a considerable challenge in identifying specific GAG-protein interactions (Annaval et al. 2020). The imperative for a more profound understanding is underscored by the involvement of proteoglycans in diverse pathological processes, spanning from cancer and inflammation to Alzheimer's disease and pathogen infections

(Iozzo and Sanderson 2011; Connell and Lortat-Jacob 2013; Gschwandtner et al. 2017; Gopal 2020; Ozsan McMillan et al. 2023). Yet, despite these critical roles, the availability of drugs targeting proteoglycans remains limited, likely due to the lack of a detailed understanding of the structure-function relationship (Morla 2019; Theocharis et al. 2019). Thus, there is a strong need for innovative methodologies to probe GAG biosynthesis, comprehend its regulation, and facilitate the development of GAG mimetics with precisely defined sequences.

The biosynthesis of Hep, HS, CS and DS commences with the step-wise addition of a common [GlcA β 1-3Gal β 1-3Gal β 1-4Xyl] tetrasaccharide linker onto the serine residue of the core protein. This process takes place in the Golgi lumen, facilitated by the peptide-O-xylosyltransferases, XylT1/XylT2, which transfer a xylose (Xyl) from UDP-Xyl onto the side chain of the serine residue (Götting et al. 2000; Pönighaus et al. 2007). The crystal structure of XylT1, in complex with acceptor peptide substrates, has revealed the specific recognition mechanism of the Ser-Gly motif (Briggs and Hohenester 2018). Subsequently, a galactose (Gal) molecule is transferred from a UDP-Gal substrate by β 4GalT7 (Almeida et al. 1999; Tsutsui et al. 2013), followed by the addition of a second Gal molecule through the action of β 3GalT6 (Bai et al. 2001; van Damme et al. 2018). Linker synthesis is completed by the glucuronyltransferase GlcAT-1

Received: September 22, 2023. Revised: January 29, 2024. Accepted: February 22, 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

catalyzing the transfer of glucuronic acid (GlcA) from UDP-GlcA (Pedersen et al. 2000, 2002; Kitagawa et al. 2001).

The addition of the fifth sugar molecule dictates the fate of the resulting glycosaminoglycan chain. Transfer of a N-acetylglucosamine (GlcNAc) molecule by EXTL3 or EXTL2 leads to the formation of Hep/HS chains (Kitagawa et al. 1999; Kim et al. 2001; Nadanaka et al. 2013; Wilson et al. 2022). In contrast, N-acetylgalactosamine (GalNAc) transfer by CSGALNACT-1 or CSGALNACT-2 directs subsequent biosynthesis towards the generation of CS/DS chains (Sato et al. 2003; Gulberti et al. 2012). The regulatory mechanism governing the addition of either GlcNAc or GalNAc, thus determining the formation of HSPG or CSPG respectively, remains enigmatic.

One conceivable mechanism posits that HSPGs and CSPGs exhibit distinct, yet conserved features in their amino acid sequence and/or three-dimensional structure. These features might be specifically recognized by the EXTL3 or CSGALNACT-1/CSGALNACT-2 enzymes. Previous bioinformatic studies have indeed revealed conserved attributes in the amino acid sequence surrounding the glycosaminoglycan attachment site: (i) an enrichment of serine and glycine residues, alongside large, hydrophobic residues in close proximity to the attachment site in HSPGs, and (ii) an accumulation of negatively charged residues in CSPGs (Zhang et al. 1995; Briggs and Hohenester 2018; Toledo et al. 2020). Despite these characteristics, accurately predicting whether a given protein contains a GAG chain remains challenging. Additionally, predicting the specific type of GAG chain that will be added remains impossible.

Thus, there is a strong need for innovative tools that facilitate a more detailed investigation of the divergent step in GAG biosynthesis. In this study, we present the chemo-enzymatic synthesis of fluorescent-labeled peptides bearing the native tetrasaccharide linker, achieved through the use of recombinantly expressed and purified linker-synthesizing glycosyltransferases (GTs). A similar approach was recently described by Sammon et al. (Sammon et al. 2023). This methodology will significantly advance future studies aimed at characterizing both the initial and the divergent steps of GAG biosynthesis.

Materials and methods

Construction of expression vectors

XylT1

The coding sequence for human XylT-I (Uniprot #Q86Y38), encompassing residues 232 to 959, was synthesized with an N-terminal BM40 signal sequence, an 8xHistidine tag, and a TEV cleavage site. This construct was then inserted into the pTWIST-CMV-WPRE-Neo plasmid (Twist Bioscience).

β 4GalT7

The coding sequence for human β 4GalT7 (Uniprot #Q9UBV7), spanning residues 81–327, was cloned utilizing *NcoI* and *NotI* restriction sites into the pETM30 vector backbone. This vector includes an N-terminal 6xHistidine tag, a GST, and a TEV site cleavage (Chatron-Colliet et al. 2017).

β 3GalT6

The coding sequence for human β 3GalT6 (Uniprot #Q96L58), ranging from residues 53–329, was synthesized (Twist Bioscience) and integrated into the pHR-CMV-TetO2 vector using *BshTI* and *KpnI* restriction sites. This resulted in an

N-terminal chicken RPTP σ signal sequence, followed by a SUMO fusion protein, a 6xHistidine tag, and a 3C cleavage site (Elegheert et al. 2018; Behiels and Elegheert 2021).

GlcAT-1

The coding sequence for human GlcAT-1 (Uniprot #O94766), spanning residues 76 to 335, was custom-synthesized for bacterial expression (Twist Bioscience). It was then cloned into the pETM30 vector using *NcoI* and *NotI* restriction sites. This construct featured an N-terminal 6xHistidine tag, a GST tag, and a TEV cleavage site.

EXTL3

The coding sequence for human EXTL3 (Uniprot #O43909), ranging from residues 53 to 919, was custom-synthesized (Twist Bioscience) and integrated into the pHR-CMV-TetO2 vector using *BshTI* and *KpnI* restriction sites. The resultant vectors encoded an N-terminal chicken RPTP σ signal sequence, followed by a SUMO fusion protein, a 6xHistidine tag, and a 3C cleavage site.

Protein expression and purification

XylT1

Two batches of 225 mL of 293-F cells at a concentration of 1.10^6 cells/mL were prepared in FreeStyle™ 293 medium (Thermo Fisher). Cultures were incubated with agitation at 37 °C in a 5% CO₂ environment for all described steps. Transfections with the pTWIST-Xylt1_232-959 were performed 24 h later, once cultures reached approximately 2.10^6 cell/s/mL, using a mass ratio of polyethylenimine (PEI) to DNA in a 2:1 ratio (500 μ g PEI: 250 μ g DNA) and resulting in a total culture volume of 250 mL per Erlenmeyer flask. After 96 h of incubation, cell cultures were centrifuged for 5 min at $500 \times g$. The resulting supernatants were supplemented with protease inhibitor cocktail tablets (Roche) and further centrifuged for 15 min at $6,000 \times g$ at 4 °C. The supernatants were then filtered through a Durapore 0.45 μ m PVDF membrane (Merck Millipore) and 50 mM Tris (pH 7.4), 250 mM NaCl, 10 mM MgCl₂, and 10 mM imidazole were added. The protein suspension was loaded onto a HisTrap HP™ 5 mL column (Cytiva) that had been pre-equilibrated with Tris buffer A (50 mM Tris, pH 7.4, 250 mM NaCl). Elution was carried out using a gradient of imidazole concentrations (20 mM, 50 mM, 100 mM, 200 mM, 500 mM, and 750 mM) achieved by mixing Tris buffer A and Tris buffer B (50 mM Tris, pH 7.4, 250 mM NaCl, 1 M imidazole, and 5% glycerol (v/v)). All elution fractions were collected and subsequently analyzed by SDS-PAGE. Fractions containing the protein of interest were pooled and transferred into a dialysis membrane with a molecular weight cut-off of 3.5 kDa (SpectraPor®3 Dialysis Membrane Standard Rc Tubing, Spectrum Labs™). The protein was dialyzed against 2 L of Tris buffer A containing 5% glycerol for 16 h at 4 °C. The resulting supernatant was concentrated to a final concentration of 0.5 mg/mL (total yield of 1.7 mg) using a 30 kDa molecular weight cut-off centrifugal filter (Amicon Ultra, Millipore, Sigma-Aldrich). Aliquots were prepared, snap-frozen in liquid nitrogen, and stored at -80 °C.

β 4GalT7

The pETM30- β 4GalT7_81-327 plasmid was transformed into *E. coli* Rosetta(DE3)pLysS bacterial cells. Four liters of Luria-Bertani broth medium supplemented with kanamycin (15 μ g/mL) and chloramphenicol (34 μ g/mL) were inoculated

with a preculture and cultivated at 37 °C with shaking. Cultures were shifted to 16 °C upon reaching an OD value of 0.2 at 600 nm and induced with 0.25 mM IPTG when the OD value reached 0.6 at 600 nm. Incubation was extended overnight at 16 °C. Bacteria were pelleted at 6,000 × *g* for 15 min and washed with 100 mL of PBS. Following an identical centrifugation step, pellets were resuspended in 120 mL of lysis buffer, composed of 25 mM HEPES (pH 7.4), 500 mM NaCl, 1 M urea, 5% glycerol (v/v), 10 mM imidazole, 200 U of DNase I (Roche), and 2 protease inhibitor tablets (EDTA-free, Roche cOmplete). Cells were lysed by sonication on ice (using a 4 min program of 2 s on, 8 s off). The resulting lysate was diluted two times in cold lysis buffer, spun down at 35,000 × *g* for 30 min at 4 °C, and the supernatant was sterile-filtered. All subsequent purification steps were conducted at 4 °C and analyzed by SDS-PAGE. The supernatant was applied to a HisTrap HP™ 5 mL column (Cytiva) pre-equilibrated with HEPES buffer (25 mM HEPES, pH 7.4, containing 500 mM NaCl and 5% glycerol (v/v)). After washing with HEPES buffer containing 200 mM imidazole, β4GalT7 was eluted with 500 mM imidazole in HEPES buffer. Two milligrams of His-tagged Tobacco Etch Virus (TEV) protease were added, and reactions were dialyzed overnight against 4 L of HEPES buffer. The supernatant was reloaded onto the HisTrap HP™ column and cleaved β4GalT7 eluted at 50 mM imidazole. Pooled fractions were concentrated using a 30 kDa molecular weight cut-off centrifugal filter (Amicon Ultra, Millipore, Sigma-Aldrich) and injected onto a Superdex 200 Increase 10/300 GL (Cytiva) pre-equilibrated in S200 Buffer (25 mM MES, pH 6.5, 200 mM NaCl, 5% glycerol (v/v)). Elution fractions containing pure β4GalT7 were pooled, concentrated to 4.9 mg/mL (total yield of 3.6 mg), and snap-frozen for storage at –80 °C.

SUMO-β3GalT6

Transfection of 600 mL of 293-F cells with the pHR-β3GalT6_53-329 plasmid was carried out, followed by a one-step IMAC purification procedure, similar to the method employed for XylT1 as described above. The purified protein was dialyzed against Tris buffer A supplemented with 5% glycerol (v/v). Subsequently, the protein was concentrated to a final concentration of 2.8 mg/mL (total yield of 2.1 mg). The concentrated protein was then snap-frozen in liquid nitrogen and stored at –80 °C for future use.

GlcAT-1

The pETM30-GlcAT-1_76-335 plasmid was transformed into competent *E. coli* BL21(DE3)pLysS bacterial cells. Protein purification followed a protocol similar to that employed for β4GalT7, as described above, with a minor deviation. After removing the 6xHis-GST tag, the protein was retrieved in the flow-through of the second IMAC step. Subsequently, the protein was concentrated to 1.8 mg/mL and snap-frozen for storage (total yield of 1.4 mg).

EXTL3

Transfection of 600 mL of 293-F cells with the pHR-EXTL3_53-919 plasmid, followed by the initial IMAC purification step, was conducted in a manner analogous to the procedure employed for XylT1 as described above. Subsequently, 2 mg of His-tagged 3C protease were introduced, and the reactions were dialyzed overnight against 2 L of Tris buffer A supplemented with 5% glycerol (v/v). The resulting supernatant was reloaded onto the HisTrap HP™ column,

and cleaved EXTL3 was eluted with Tris buffer A containing 25 mM imidazole. Pooled fractions were concentrated and simultaneously buffer exchanged into Tris buffer A with 5% glycerol (v/v) using a 30 kDa molecular weight cut-off centrifugal filter (Amicon Ultra, Millipore, Sigma-Aldrich). This yielded a final concentration of 5.4 mg/mL, totaling 2.5 mg. The concentrated protein was promptly snap-frozen in liquid nitrogen and stored at –80 °C for future use.

Nano differential scanning fluorimetry measurements

Thermal stability of purified XylT1, β4GalT7, SUMO-β3GalT6, and GlcAT-1 were assessed using nano differential scanning fluorimetry technology (nanoDSF). Protein samples were diluted to a concentration of 0.3 mg/mL in their corresponding storage buffers: XylT1 and SUMO-β3GalT6 in Tris A buffer with 5% glycerol (v/v), β4GalT7 in S200 buffer, and GlcAT-1 in HEPES A buffer. Standard-grade glass capillaries (Nanotemper) were filled with 15 μL of protein suspension and loaded onto the tray of Prometheus NT.48 instrument (NanoTemper Technologies). The experiment was run using a 1.5 °C per minute thermal ramp over a temperature range from 25 °C to 95 °C at 80%–100% excitation power. Acquisition was followed by data analysis using the PR.ThermControl v.2.3.1 software in order to establish melting temperatures for each sample.

Mass photometry analysis

Mass photometry analyses were conducted using the Refeyn OneMP device (Oxford, UK). Purified proteins were diluted in the following buffers: XylT1 and SUMO-β3GalT6 in Tris buffer A with 5% glycerol (v/v), β4GalT7 in S200 buffer, and GlcAT-1 in HEPES buffer. Cover slides (24×50 mm, 170 μm) were thoroughly rinsed with demineralized water and isopropanol, and subsequently dried under a clean stream of compressed air. A reusable culture well gasket (Grace Bio-Labs) was placed onto a cover slide. To establish focus, 19 μL of buffer was added to the well and autofocus mode was engaged using the Refeyn AcquireMP v.2.4.0 software. Movies, consisting of 60 s (6,000 frames), were recorded with a standard field-of-view, and the data was processed using DiscoverMP v.2.5.0 software. Calibration was performed by adding 1 μL of 1:20 diluted NativeMark Protein Standard (Thermo Fisher Scientific) in buffer. For each acquisition, either 18 μL or 19 μL of buffer were employed to set the focus, followed by the addition of 1 μL to 2 μL of the protein sample to achieve a final concentration between 50 nM and 125 nM. Duplicates were carried out for each sample.

Tetrasaccharide linker synthesis

Peptide sequences GGPSGDFE (SDC3_short), VSGGPS-GDFELPEEET (SDC3_long), GDSSGWD (TGFBR3), EEAS-GEAS (CSF1_short) and VP EEASGEASEIPVPQ (CSF1_long) were synthesized with an additional N-terminal TAMRA fluorophore (SB-PEPTIDE, SmartBioscience). Peptides SDC3_short, SDC3_long and TGFBR3 peptides were dissolved in 30% acetonitrile (v/v), while peptides CSF1_short and CSF1_long were suspended in demineralized water, resulting in a final concentration of 1 mM for each. All enzymatic assays were conducted at room temperature in the dark, using a reaction buffer consisting of 25 mM MES pH 6.5, 125 mM NaCl, 2.5 mM MgCl₂, and 2.5 mM MnCl₂. For

the first condition, a mixture of 100 μM of peptide, 1 mM UDP-Xyl, and 1 μM XylT1 was prepared. A second reaction was set up similarly, but with the addition of 1 mM UDP-Gal and 8.4 μM β4GalT7 . Both mixtures were incubated for 16 h and then stored at -20°C to halt the reactions. Reaction three was prepared as the second reaction, but 1 mM UDP-Gal and 3.15 μM SUMO- β3GalT6 were added after the initial 16 h incubation. Reaction four was set up as reaction 3, and 1 mM UDP-GlcA and 3.15 μM GlcAT-1 were simultaneously added with SUMO- β3GalT6 . Reactions three and four were incubated for an additional 24 h before being stored at -20°C . To analyze the glycan transfer on the peptides, gel electrophoresis was conducted. One microliter of each reaction was mixed with 39 μL of 20% glycerol (v/v) containing 1% phenol red (v/v), and 5 μL of the mixture were loaded onto a 25% polyacrylamide gel (not containing any sodium dodecyl sulfate (SDS)). Migration was carried out in a Tris-glycine running buffer, also lacking SDS. Visualization was achieved using a ChemiDoc MP instrument with Alexa 546 detection settings. The experiment was repeated three times for each peptide. Two identical replicate mixtures were pooled, adjusted to 200 μL with Tris A buffer, and injected onto two Superdex Peptide 10/300 GL (Cytiva) columns connected in line, pre-equilibrated with Tris buffer A.

Purification of tetrasaccharide linker peptides

For the larger-scale synthesis of the tetrasaccharide linker on peptides, a 300 μL reaction mixture was prepared, consisting of 100 μM of a specific peptide, 1 mM UDP-Xyl, 1 mM UDP-Gal, 1 μM XylT1, and 8.4 μM β4GalT7 in reaction buffer (25 mM MES pH 6.5, 125 mM NaCl, 2.5 mM MgCl_2 , and 2.5 mM MnCl_2). Following a 16 h incubation at room temperature in the dark, an additional 1 mM UDP-Gal, 1 mM UDP-GlcA, 3.15 μM SUMO- β3GalT6 , and 3.15 μM GlcAT-1 were introduced, bringing the final volume to 366 μL . Incubation was extended for an additional 24 h, and the reactions were halted by transferring the mixture to -20°C . The entire sample was then injected onto two Superdex Peptide 10/300 GL columns (Cytiva) connected in line, pre-equilibrated with Tris buffer A. Elution fractions of the SDC3_long peptide were analyzed by polyacrylamide gel electrophoresis. Peak fractions of the different glycosylated peptides were pooled and dialyzed using a dialysis membrane with a molecular weight cut-off of 1 kDa (SpectraPor). This dialysis was carried out against 4 L of water at 4°C in the dark for 3 h. A second dialysis step was performed overnight. The samples were then lyophilized using an Alpha 1–2 LD plus instrument (Martin Christ). Lyophilisates were dissolved in demineralized water and their concentrations were determined using titration curves prepared with non-glycosylated peptides. Absorbance values were measured at 554 nm using a Nanodrop 2,000 (Thermo Scientific). Peptides carrying the full linker were stored at -20°C .

Mass spectrometry analysis

A commercial Matrix Assisted Laser Desorption Ionisation (MALDI)-time-of flight (TOF)/TOF mass spectrometer (Autoflex MAX, Bruker Daltonics, Bremen, Germany) was used for the measurements. Ionization was achieved using a Bruker-proprietary (Nd:YAG) 355 nm smartbeamTM II solid state laser (maximal frequency: 2 kHz). The laser energy was just above the threshold required for the formation of ions. The measurements of peptides were performed

in reflector positive ion mode, detecting 700–5,000 m/z range. Each mass spectrum was the average of 10,000 laser shots randomly acquired at different sample area. We used Bruker AnchorChipTM plate as target. We diluted unmodified peptide (15 μM) 1:3 using 5% formic acid (FA) solution to obtain a 5 μM solution. The glycosylated peptide was not diluted (final concentration: 5 μM). We used a procedure to deposit samples on the MALDI plate as described previously (Signor and Boeri Erba 2013). Briefly, we utilized a thin layer (TL) prepared as a saturated solution of α -cyano-4-hydroxycinnamic acid (CHCA) matrix in acetonitrile (ACN). On the TL, we deposited 0.8 μL of peptide of interest. On each peptide spot, we added 0.8 μL of a matrix (CHCA and 2,5-Dihydroxybenzoic acid (DHB)) as final step. The samples were left to dry under the extractor hood at room temperature. 20 mg/mL CHCA solution was prepared in ACN and 5% FA (70:30, vol/vol). Final concentration of solutions was: 70% ACN and 1.5% FA. 20 mg/mL DHB solution in ACN and 0.1% trifluoroacetic acid (TFA)(70:30, vol/vol). Final concentration of solution was 70% ACN and 0.03% TFA. Then, CHCA and DHB were mixed in 1:1 ratio (Signor and Boeri Erba 2013). Next to the samples, we deposited 0.6 μL of a calibrant standard (peptide standard II, Bruker Daltonics) and then added 0.6 μL of the matrix solution. Regarding data analyzes, the spectra of intact peptides were processed using FlexAnalysisTM (version 3.4). Sophisticated Numerical Annotation Procedure detection algorithm was used for the identification of peaks with a threshold of 2 signal-to-noise ratio and a number maximal of 500 peaks detected. After this, it was applied automatic labelling and we manually reviewed the peak labelling.

Purification of [Gal β 1–4Xyl]-SDC3_short peptide

A large-scale synthesis reaction was conducted to generate a [Gal β 1–4Xyl]-SDC3_short peptide for product inhibition experiments, similar to the process outlined above. In short, a 300 μL reaction mix was prepared containing 100 μM SDC3-short peptide, 1 mM UDP-Xyl, 1 mM UDP-Gal, 1 μM XylT1, and 8.4 μM β4GalT7 in the reaction buffer. This mixture was incubated in the dark for 16 h at room temperature. The glycosylated peptide was then purified by size exclusion chromatography and subsequently dialyzed against demineralized water, lyophilized and resuspended in demineralized water.

In vitro glycosyltransferase activity of β4GalT7

Various sets of activity assays were conducted to investigate the β4GalT7 enzyme activity towards the SDC3_short peptide in more detail. All enzymatic reactions were carried out in the dark at room temperature, utilizing a reaction buffer comprising 25 mM MES at pH 6.5, 125 mM NaCl, 2.5 mM MgCl_2 , and 2.5 mM MnCl_2 . Two control reactions were performed, each containing 100 μM SDC3_short peptide, 1 μM XylT1, and either no or 1 mM UDP-Xyl.

UDP-gal concentration

Reaction mixtures of 20 μL were prepared, containing buffer, 100 μM SDC3_short peptide, 1 mM UDP-Xyl, 1 μM XylT1, and 8.4 μM β4GalT7 , along with either 1 mM, 2 mM, or 5 mM UDP-Gal. After 16 h incubation, the reactions were stopped by transferring the mixture to -20°C and then analyzed by gel electrophoresis. The experiment was performed in triplicate.

Time-course

A reaction mixture of 20 μL , containing buffer, 100 μM SDC3_short peptide, 1 mM UDP-Xyl, 1 mM UDP-Gal, 1 μM XylT1, and 8.4 μM β4GalT7 , was prepared. After incubation for 1, 2, 4, 6, 16, 48, and 72 h, 1 μL of the mixture was combined with 39 μL of 20% glycerol (v/v) containing phenol red and heat-inactivated at 85 $^{\circ}\text{C}$ for 10 min and then analyzed by gel electrophoresis. This time-course experiment was conducted in triplicate.

Product inhibition

A first 10 μL reaction mixture, consisting of buffer, 100 μM SDC3_short, 1 mM UDP-Xyl, 1 mM UDP-Gal, 1 μM XylT1, and 8.4 μM β4GalT7 , was prepared. A second mix was set up similarly, but included 100 μM [Gal $\beta\text{1-4Xyl}$]-SDC3_short. After 16 h incubation, the reactions were stopped by storing the mix at -20°C . Both reaction products and an additional control, consisting of a [Gal $\beta\text{1-4Xyl}$]-SDC3_short peptide sample, were observed by gel electrophoresis. This experiment was performed in duplicate.

Specificity of β3GalT6

20 μL reaction mixture were prepared containing 100 μM SDC3_short peptide, 1 mM UDP-Xyl and 1 μM XylT1. After 16 h incubation, 1 mM UDP-Gal, and 3.15 μM SUMO- β3GalT6 were added to the reaction, followed by an additional 24 h incubation. Reaction was stopped by placing mixture at -20°C and then analyzed by gel electrophoresis. This experiment was realized in duplicate.

In vitro glycosyltransferase activity of EXTL3

Peptides SDC3_short, SDC3_long, TGFBR3, CSF1_short, and CSF1_long, all carrying a tetrasaccharide [Glc $\alpha\text{1-3Gal}\beta\text{1-3Gal}\beta\text{1-4Xyl}$]-linker, served as acceptor substrates to measure the GlcNAc-transferase activity of EXTL3. 10 μL of reaction mixture were prepared, containing 25 μM tetrasaccharide peptide, 3 mM UDP-GlcNAc, 1 μM EXTL3, and assay buffer (50 mM MES pH 6.5, 50 mM NaCl, 2.5 mM MgCl_2 , 2.5 mM MnCl_2). Reactions were incubated at 37 $^{\circ}\text{C}$ in the dark and stopped after 0, 1, 2, 4 and 6 h by transferring 1 μL of reaction mix into 9 μL 20% glycerol (v/v) containing phenol red and heating for 10 min at 85 $^{\circ}\text{C}$. The reactions were analyzed by gel electrophoresis using 25% polyacrylamide gels and band intensities quantified using Image Lab v.6.1. (Biorad). This experiment was realized in duplicate.

Results

Purification of the four linker-assembling glycosyltransferases

To produce GAG-mimicking peptides containing the native tetrasaccharide linker, we expressed and purified the GTs responsible for the initial four steps in GAG biosynthesis. All four GTs were expressed as soluble proteins, omitting their N-terminal transmembrane anchoring helix to facilitate purification and subsequent enzymatic reactions (Fig. 1A). In a first attempt, we tried to recombinantly express the enzymes β4GalT7 , β3GalT6 and GlcAT-1 in *E. coli* cells. However, due to low expression levels, we adapted a purification procedure for β3GalT6 from human embryonic kidney FreeStyle 293-F cells. XylT1 recombinant expression was performed in FreeStyle 293-F cells, following the methodology outlined in

a prior study (Briggs and Hohenester 2018). XylT1 (residues 232–959) was expressed carrying an N-terminal 8xHistidine tag, followed by a Tobacco Etch Virus (TEV) cleavage site. β4GalT7 (residues 81–327) was expressed in *E. coli* as a fusion protein with an N-terminal 6xHistidine tag, a glutathione S-transferase tag (GST), and a TEV cleavage site. For β3GalT6 (residues 53–329) expression in FreeStyle 293-F cells, a construct was designed containing an N-terminal Small Ubiquitin-like Modifier (SUMO) tag, followed by a 6xHistidine tag and a 3C protease cleavage site. GlcAT-1 (residues 76–335) was expressed in *E. coli* using the same construct design as for β4GalT7 . All proteins were purified using immobilized metal ion affinity chromatography (IMAC). The 6xHis-GST fusion tags of β4GalT7 and GlcAT-1 were removed by TEV protease treatment, followed by a second IMAC purification step. The SUMO- β3GalT6 fusion protein was not cleaved to improve protein stability. A final buffer exchange was performed using either dialysis or size exclusion chromatography. The obtained protein yields per liter of culture medium were 3.4 mg for XylT1, 0.9 mg for β4GalT7 , 3.5 mg for SUMO- β3GalT6 , and 0.35 mg for GlcAT-1. Purity of the samples was confirmed by SDS-PAGE analysis (Fig. 1B). For XylT1, in addition to the band at the expected molecular weight of 86 kDa, a high molecular weight smear was observed. One possible explanation could be that XylT1 undergoes various post-translational modifications.

Biophysical characterization of linker enzymes

Next, we assessed the stability of the purified proteins using nano differential scanning fluorimetry (nanoDSF). The melting temperature (T_m), which indicates the temperature at which half of the protein is unfolded, is a key parameter for describing thermostability. XylT1, β4GalT7 , and SUMO- β3GalT6 exhibited T_m values of 58.6 $^{\circ}\text{C}$, 56.4 $^{\circ}\text{C}$, and 58.8 $^{\circ}\text{C}$, respectively, signifying a high level of thermal stability (Fig. 1C). Only one distinct melting point was observed for the SUMO- β3GalT6 fusion protein. Given that the SUMO tag lacks tryptophans and possesses only one tyrosine, the observed changes in the absorbance ratio likely stem from the unfolding of β3GalT6 , which contains 9 tryptophans and 12 tyrosines. Notably, GlcAT-1 displayed a comparatively lower T_m value of 41.4 $^{\circ}\text{C}$. The nanoDSF experiments demonstrate that the proteins remained stable in their respective storage buffers. However, it seems best to maintain the optimal incubation temperature for enzymatic reactions below 35 $^{\circ}\text{C}$, particularly for GlcAT-1.

To gain further insight into sample homogeneity and the oligomeric structure of the four glycosyltransferases, we conducted mass photometry experiments. In all four cases, a prominent peak was observed, indicating both sample homogeneity and the presence of well-folded proteins in a defined oligomeric state. Specifically, XylT1 particles exhibited an average molecular weight of 87 kDa, aligning closely with the theoretical weight of 86 kDa for XylT1 in its monomeric state (Fig. 1D). On the other hand, the prevailing majority of particles for β4GalT7 , SUMO- β3GalT6 , and GlcAT-1 showed molecular weights of approximately 61 kDa, 94 kDa, and 58 kDa, respectively. These values correspond to the expected molecular weights for the proteins in their dimeric states. However, it is noteworthy that the anticipated molecular masses of monomeric β4GalT7 (29 kDa), SUMO- β3GalT6 (45 kDa), and GlcAT-1 (29 kDa) are in close proximity to/or

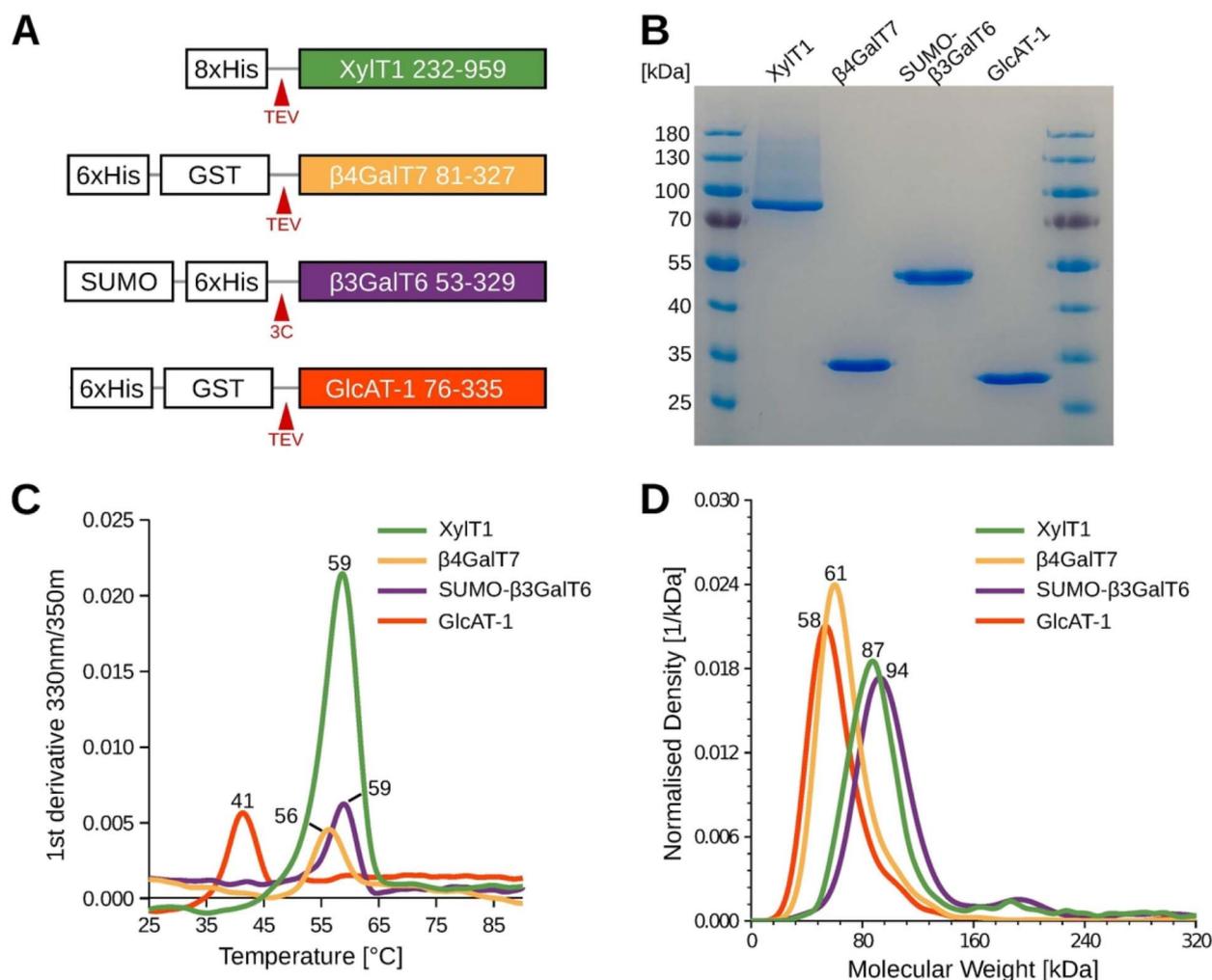


Fig. 1. Purification and biophysical characterization of linker-synthesizing glycosyltransferases. A) Schematic representation of expression constructs featuring histidine purification tags (His) in conjunction with either glutathione S-transferase (GST) or SUMO fusion proteins. Tags can be cleaved using either tobacco etch virus (TEV) or 3C proteases. B) Coomassie-stained SDS-PAGE analysis of purified glycosyltransferases. Expected molecular sizes are 86 kDa for XylT1, 29 kDa for β 4GalT7, 45 kDa for SUMO- β 3GalT6, and 29 kDa for GlcAT-1. C) NanoDSF analysis of purified proteins, with melting temperatures depicted in the graph. D) Mass photometry analysis of purified glycosyltransferases, with determined molecular masses indicated in the graph.

fall below the detection limit of mass photometry. Consequently, it cannot be ruled out that these three proteins may also exist in a monomeric form in solution. However, since we observed only one peak for β 4GalT7 and GlcAT-1 in size exclusion chromatography experiments, it is most likely that these proteins are in a dimeric state.

Design of peptides mimicking GAG attachment sites

The design of short peptide sequences was based on known sequences of proteoglycan core proteins, for which the precise GAG attachment sites have previously been experimentally determined (Noborn et al. 2022). This approach aimed to ensure that the chosen peptides would be readily recognized as substrates by the linker-synthesizing GTs. Specifically, two GAG attachment sites were selected based on heparan sulfate proteoglycans, namely syndecan-3 (SDC3) and TGF β type III receptor (TGFBR3), as well as one derived from a chondroitin sulfate proteoglycan, macrophage colony-stimulating factor 1 (CSF1) (Carey et al. 1992; Chernousov and Carey 1993; Kimura et al. 1994; López-Casillas et al. 1994). To gain further insight into the positioning of the GAG attachment

site, we examined the AlphaFold-predicted structures of the three proteoglycans (Jumper et al. 2021; Varadi et al. 2022). Upon comparison, an intriguing observation emerged: all three proteoglycans featured substantial unstructured regions (Fig. 2). Mapping of the serine residues serving as GAG attachment sites revealed their localization within these extensive unstructured regions, devoid of adjacent secondary structure elements.

In preparation for chemo-enzymatic synthesis, peptides of a length of eight amino acids were designed, including three preceding and four following the target serine residue. The first peptide, named SDC3_short and featuring the sequence ³¹¹GGPSGDFE was based on syndecan-3 (UniProt: O75056). It had been previously employed by Gao et al. for investigating XylT1 activity (Gao et al. 2021). Another peptide, TGFBR3, was derived from the TGF β type III receptor (UniProt: Q03167) and harbors the sequence ⁵³¹GDSSGWPD. A third peptide CSF1_short (³⁰⁶EEASGEAS) was based on the sequence of macrophage colony-stimulating factor 1 (UniProt: P09603). In order to study the effect of peptide lengths on enzymatic activity, we further designed longer peptides

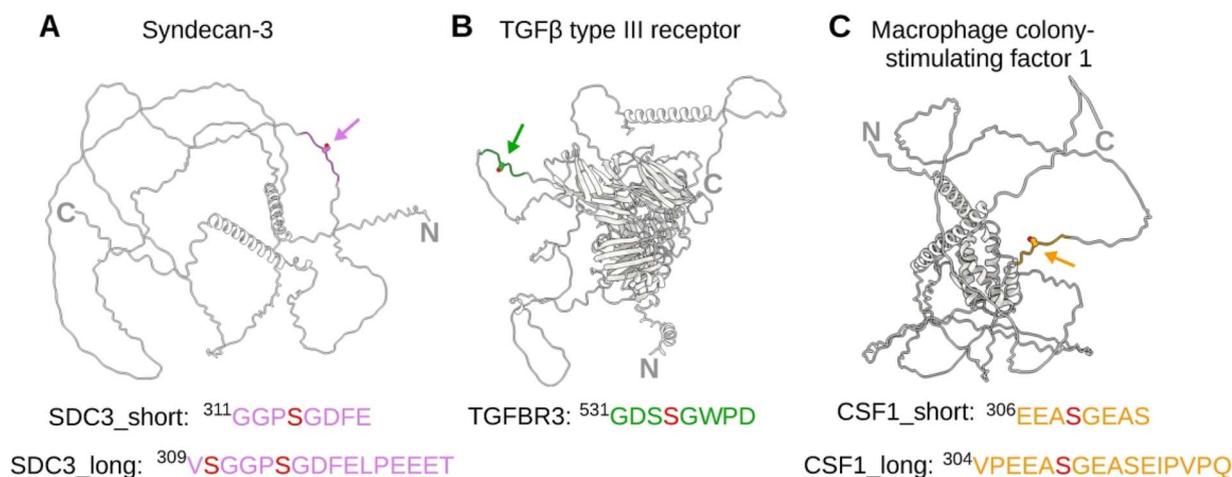


Fig. 2. Mapping of GAG attachment sites on AlphaFold predicted proteoglycan structures. A) A cartoon representation of syndecan-3 in grey, with the presumable heparan sulfate attachment site and surrounding residues colored in rose. The amino acid sequences of the peptides used in this study, SDC3_short and SDC3_long, are indicated alongside. B) TGFβ type III receptor model with heparan sulfate attachment site and surrounding residues colored in green. The sequence of the TGFBR3 peptide is indicated. C) Cartoon representation of the chondroitin sulfate proteoglycan macrophage colony-stimulating factor 1 with its glycosaminoglycan attachment site and surrounding residues colored in orange. The amino acid sequences of the peptides CSF1_short and CSF1_long are indicated alongside.

for syndecan-3 (SDC3_long, ³⁰⁹VSGGPSGDFELPEEET) and for the macrophage colony-stimulating factor 1 (CSF1_long, ³⁰⁴VPEEASGEASEIPVPQ).

Enzymatic synthesis of tetrasaccharide-carrying linker peptides

In order to produce peptides bearing the native tetrasaccharide [GlcAβ1–3Galβ1–3Galβ1–4Xyl], we conducted enzymatic reactions utilizing the four linker-synthesizing glycosyltransferases. For expedited and accurate analysis of the extension reactions, we designed fluorescently-labeled peptides with a N-terminal 5-Carboxytetramethylrhodamine (TAMRA) modification. The TAMRA fluorophore, with excitation and emission wavelengths of 541 nm and 568 nm respectively, enabled highly sensitive detection within the picomolar range. To monitor the addition of individual molecules, reactions were conducted in the presence of one, two, three, or all four linker enzymes (Fig. 3A).

The tetrasaccharide addition was performed in a two-step process. First, XylT1 and β4GalT7, along with their respective substrates UDP-Xyl and UDP-Gal, were incubated with the TAMRA-labeled peptides. Following a 16-hour incubation, SUMO-β3GalT6 and GlcAT-1 were added, along with UDP-GlcA and additional UDP-Gal. The reaction mixture was then incubated for another 24 h. The reaction products were separated using a 25% acrylamide gel and visualized through a fluorescence imaging system.

The successive addition of xylose, galactose, galactose, and glucuronic acid moieties onto the SDC3_short peptide (³¹¹GGPSGDFE) resulted in distinct band shifts (Fig. 3B, Supplementary Fig. 1A). Comparison of reaction 1 (lacking any UDP-sugar donor) with reaction 2 (containing XylT1 and UDP-Xyl) revealed a band shift, indicative of an increased molecular weight due to covalent glycan transfer. Notably, reaction 2 displayed only one band, suggesting complete reaction conversion. Following the addition of β4GalT7 and UDP-Gal (reaction 3), a further reduction in migration speed was observed, in line with the formation of a [Galβ1–4Xyl]-peptide product. However, the reaction was not entirely complete, as a band corresponding to the [Xyl]-peptide was also detected. An additional band shift was noted

between reaction 3 and reaction 4, indicating the formation of a [Galβ1–3Galβ1–4Xyl]-peptide. Interestingly, the band corresponding to the intermediate [Xyl]-peptide was notably fainter in reaction 4 compared to reaction 3. The product of reaction 5, which included GlcAT-1 and UDP-GlcA, exhibited faster migration despite an increased molecular size. This observation can be attributed to the enhanced negative charge of the resulting [GlcAβ1–3Galβ1–3Galβ1–4Xyl]-peptide following glucuronic acid addition.

Chemo-enzymatic reaction products were further analyzed by size-exclusion chromatography. Each addition of a single sugar molecule led to a distinct shift in elution volume, further affirming the accurate addition of glycan moieties onto the SDC3_short peptide substrate (Fig. 3C). Parallel chemo-enzymatic reactions for the peptide TGFBR3 (⁵³¹GDSSGWPD), CSF1_short (³⁰⁶EEASGEAS) and CSF1_long (³⁰⁴VPEEASGEASEIPVPQ) yielded highly similar results. Reactions with SDC3_long (³⁰⁹VSGGPSGDFELPEEET), in contrast, led to a more complex band pattern (Fig. 3D–G, Supplementary Fig. 1B–E) due to the presence of two glycosaminoglycan attachment sites (³¹⁰S and ³¹⁴S). The HS attachment to ³¹⁰S had not been described before, and its modification might only take place in the here described in vitro reaction conditions.

Larger scale synthesis reactions were conducted to generate peptides with a tetrasaccharide linker. These tetrasaccharide peptides were subsequently purified by size exclusion chromatography to eliminate biosynthetic enzymes, UDP-sugar substrates and reaction products (Supplementary Fig. 2). It also resulted in a relatively homogeneous sample for peptide SDC3_long, carrying two tetrasaccharide linkers, as confirmed by mass spectrometry analysis (Fig. 4). Mass spectrometry analysis further revealed the presence of small amounts of a non-canonical trisaccharide linker [GlcAβ1–3Galβ1–4Xyl].

Limiting factors in β4GalT7 activity

Intrigued by the fact that full conversion of the [Xyl]-peptides by β4GalT7 was reproducibly only observed, when the reaction was coupled to the subsequent step catalyzed by SUMO-β3GalT6, we set out to study the activity of β4GalT7 in more detail. When following galactose addition on the SDC3_short

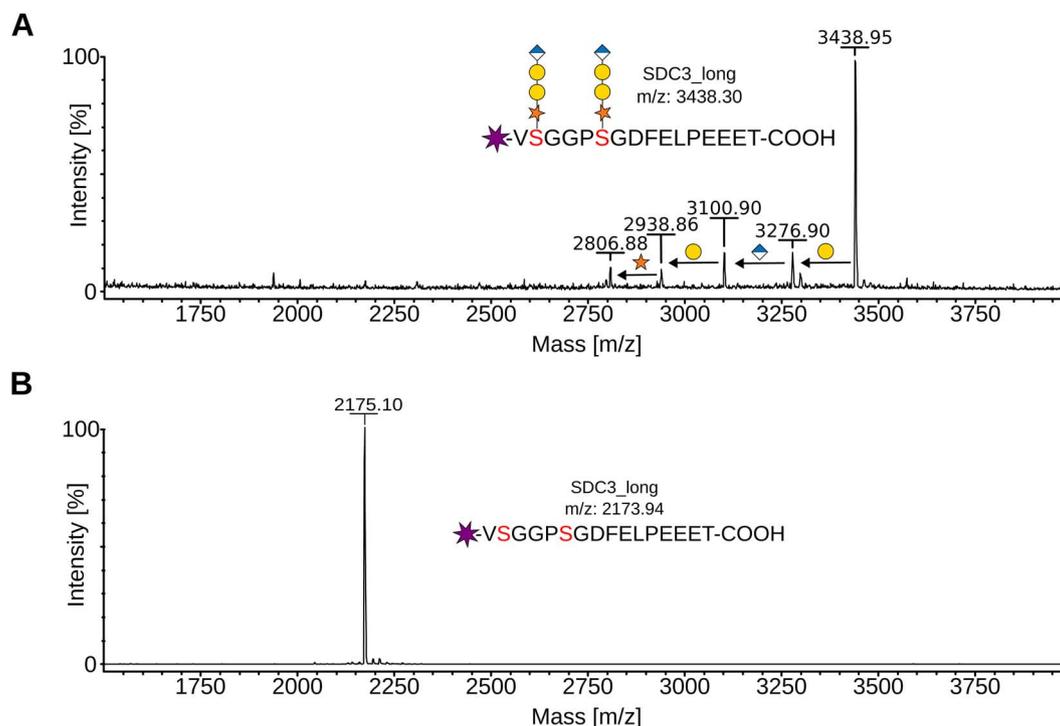


Fig. 4. Mass spectrometry analysis of glycosylated SDC3_long peptide. A) Purified tetrasaccharide-carrying SDC3_long peptide was analyzed by MALDI-TOF. The major peak reflects the SDC3_long peptide carrying two [GlcA β 1-3Gal β 1-3Gal β 1-4Xyl]-glycans. Four additional minor signals correspond to less glycosylated peptides. The accurate mass difference between glycopeptides allows the assignment of distinct forms. Monosaccharide symbols follow the SNFG (symbol nomenclature for Glycans) system (Varki et al. 2015). B) Control MALDI-TOF spectra of non-glycosylated SDC3_long peptide.

UDP-galactose concentrations could inhibit XylT1 activity. Another possible explanation for observing reaction completeness only in the presence of SUMO- β 3GalT6, is that the latter removes the [Gal β 1-4Xyl]-peptide product. To explore the potential inhibition of β 4GalT7 by its own product, we purified [Gal β 1-4Xyl]-SDC3_short peptide. The activity of β 4GalT7 in presence of its reaction product was found to be lower, as the non-converted [Xyl]-peptide increased by around 60% (Fig. 5C). Of note, SUMO- β 3GalT6 did not extend the [Xyl]-SDC3_short peptide in the absence of β 4GalT7 activity (Fig. 5C, Supplementary Fig. 3C).

Role of core protein sequence in the divergent step of GAG biosynthesis

The biosynthesis of various glycosaminoglycans commences with the assembly of the tetrasaccharide linker. The addition of the fifth sugar molecule marks a pivotal divergence point in the biosynthetic pathways: the incorporation of a GlcNAc residue initiates HS/Hep production, while the addition of GalNAc leads to the formation of a CS/DS chain. To gain a deeper insight into the regulatory mechanisms governing this critical step, we aimed on elucidating the substrate specificity of EXTL3, the enzyme responsible for transferring GlcNAc onto the tetrasaccharide linker. We successfully expressed and purified EXTL3 (Fig. 6A and B, Supplementary Fig. 4). SDS-PAGE analysis suggested that EXTL3 protein was highly pure and that it carries several post-translational modifications, e.g. N-glycans, as its apparent molecular mass was around 10 kDa higher than expected.

Next, we investigated the GlcNAc transfer activity of EXTL3 towards five tetrasaccharide peptides derived from the HSPGs syndecan-3 (SDC3_short and SDC3_long) and TGF β

type III receptor (TGFBR3), as well as the CSPG macrophage colony-stimulating factor 1 (CSF1_short and CSF1_long). EXTL3-catalyzed GlcNAc addition onto the tetrasaccharide peptides was followed for 1, 2, 4 and 6 h (Fig. 6C and D, Supplementary Fig. 5).

Highest reaction completeness after 2 h was observed for the SDC3_short peptide (84%). The remaining tetrasaccharide peptides were converted with decreasing efficiency: TGFBR3 (68%) > CSF1_short (42%) > CSF1_long (33%) > SDC3_long (16%) (Fig. 6C and D, Supplementary Fig. 5). This suggests that the short 8-amino acid peptides derived from HSPGs (SDC3_short, TGFBR3) are better substrates for EXTL3 compared to the one originating from CSGPs (CSF1_short). The longer peptides showed lower conversion rates than their shorter counterparts. For SDC3_long, which carries two tetrasaccharide linkers, we could see the addition of either one or two GlcNAc residues. However, the band corresponding to the addition of two GlcNAc residues was barely detectable.

Discussion

GAGs are essential linear polysaccharide chains found on cell surfaces and in the extracellular matrix. Among these, HS and CS are prominent, both attached to the core protein of proteoglycans through a shared tetrasaccharide linker consisting of glucuronic acid, galactose, galactose, and xylose. This linker is meticulously constructed by a group of glycosyltransferases localized in the Golgi apparatus. The addition of the fifth sugar, either GlcNAc or GalNAc, marks a pivotal point where HS and CS chains diverge, each serving distinct biological functions. The regulatory mechanisms governing

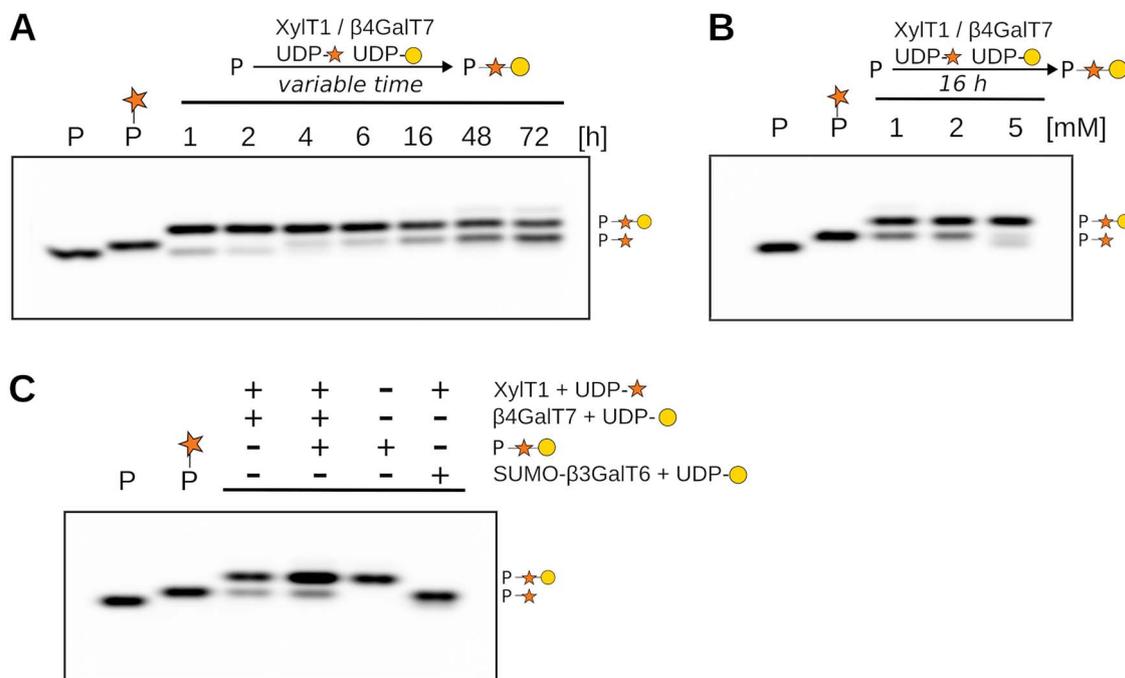


Fig. 5. β 4GalT7 activity towards SDC3_short peptide under various reaction conditions. A) Time-course experiment to follow galactose transfer by β 4GalT7 onto SDC3_short peptide substrate. Reactions were stopped after 1, 2, 4, 6, 16, 48, and 72 h by heat-inactivation. B) Effect of donor substrate concentration on β 4GalT7 activity was measured by varying UDP-Gal concentration from 1 mM to 2 mM and 5 mM. C) Potential influence of [Gal β 1-4Xyl]-SDC3_short reaction product on β 4GalT7 enzyme activity. Enzymatic reactions were carried out in the presence and absence of a purified [Gal β 1-4Xyl]-SDC3_short peptide, which was added at the equimolar amount as the SDC3_short peptide. Purified [Gal β 1-4Xyl]-SDC3_short alone served as negative control. Last lane shows reaction to study galactosyltransferase activity of β 3GalT6 onto [Xyl]-SDC3_short peptide. Monosaccharide symbols follow the SNFG (symbol nomenclature for Glycans) system (Varki et al. 2015).

this crucial step remain elusive. It is tempting to speculate that, considering glycosaminoglycan biosynthesis occurs post-translationally in the Golgi apparatus, the protein's amino acid sequence or its three-dimensional structure may encode the information dictating its fate as either an HSPG or CSPG. To explore this hypothesis, we established a methodology for generating fluorescent peptides carrying the shared GAG linker.

Peptides were designed based on key proteoglycans—syndecan-3 and TGF β type III receptor, emblematic of HSPGs, and macrophage colony-stimulating factor 1, a representative of CSPGs. Structural analysis using AlphaFold predictions unveiled that all peptides are situated within highly flexible regions (Fig. 2). It is noteworthy that while AlphaFold may not definitively confirm protein disorderliness, it does suggest regions amenable to structural flexibility (Ruff and Pappu 2021; Guo et al. 2022). These unstructured regions may adopt a more ordered conformation within the cell through interactions with other proteins or GAG chains (Gondelaud et al. 2021). The high flexibility surrounding the GAG attachment site potentially confers an advantage for post-translational glycan moiety addition, likely by facilitating access to catalytic sites (Bludau et al. 2022).

The designed peptides were readily accepted by the linker enzymes as substrates, with most reactions reaching completion (Fig. 3). However, the β 4GalT7 reaction, responsible for generating [Gal β 1-4Xyl]-peptide, displayed partial completeness, leaving a discernible [Xyl]-peptide band on the gel. Interestingly, this band diminished when β 3GalT6 was added after 16 h to the XylT1 and β 4GalT7 containing reaction. To eliminate the possibility that this

effect was merely a result of extended incubation, a time-course experiment was conducted. Surprisingly, we observed a decrease in the quantity of the generated [Gal β 1-4Xyl]-peptide after an initial accumulation, concurrently with the accumulation of the [Xyl]-peptide intermediate (Fig. 5A). One plausible explanation is that galactose undergoes hydrolysis from the generated [Gal β 1-4Xyl]-peptide, and with prolonged incubation times, the UDP-Gal substrate is depleted or β 4GalT7 loses activity, preventing the generation of new [Gal β 1-4Xyl]-peptide. The hydrolysis is likely enzyme-assisted, given that the [Gal β 1-4Xyl]-peptide remains stable upon heat inactivation of the sample. Considering both the hydrolysis of [Gal β 1-4Xyl]-peptide and the observed product inhibition affecting β 4GalT7 activity (Fig. 5C), we propose that β 3GalT6 enhances the efficiency of β 4GalT7 by eliminating its [Gal β 1-4Xyl]-peptide product. It should be mentioned, that β 4GalT7 and GlcAT-1 were expressed in *E. coli* cells, lacking their native post-translational modifications, including N-glycans. This absence could potentially impact their stability, activity, and specificity. Another possibility, which remains to be investigated, is a cooperative function between β 3GalT6 and β 4GalT7, possibly facilitated by functional or physical complexes (Kellokumpu et al. 2016; Li and Kusche-Gullberg 2016). It should be noted that when XylT1, β 4GalT7, and β 3GalT6 were introduced simultaneously in the reaction, none of the sugar transfers were completed. This resulted in a very low final proportion of peptide with [Gal β 1-3Gal β 1-4Xyl] (results not shown). This phenomenon might be attributed to the potential formation of a non-canonical Gal-Xyl-peptide (Persson et al. 2019; Delbaere et al. 2020) or the competition between β 4GalT7 and β 3GalT6 for the UDP-Gal substrate.

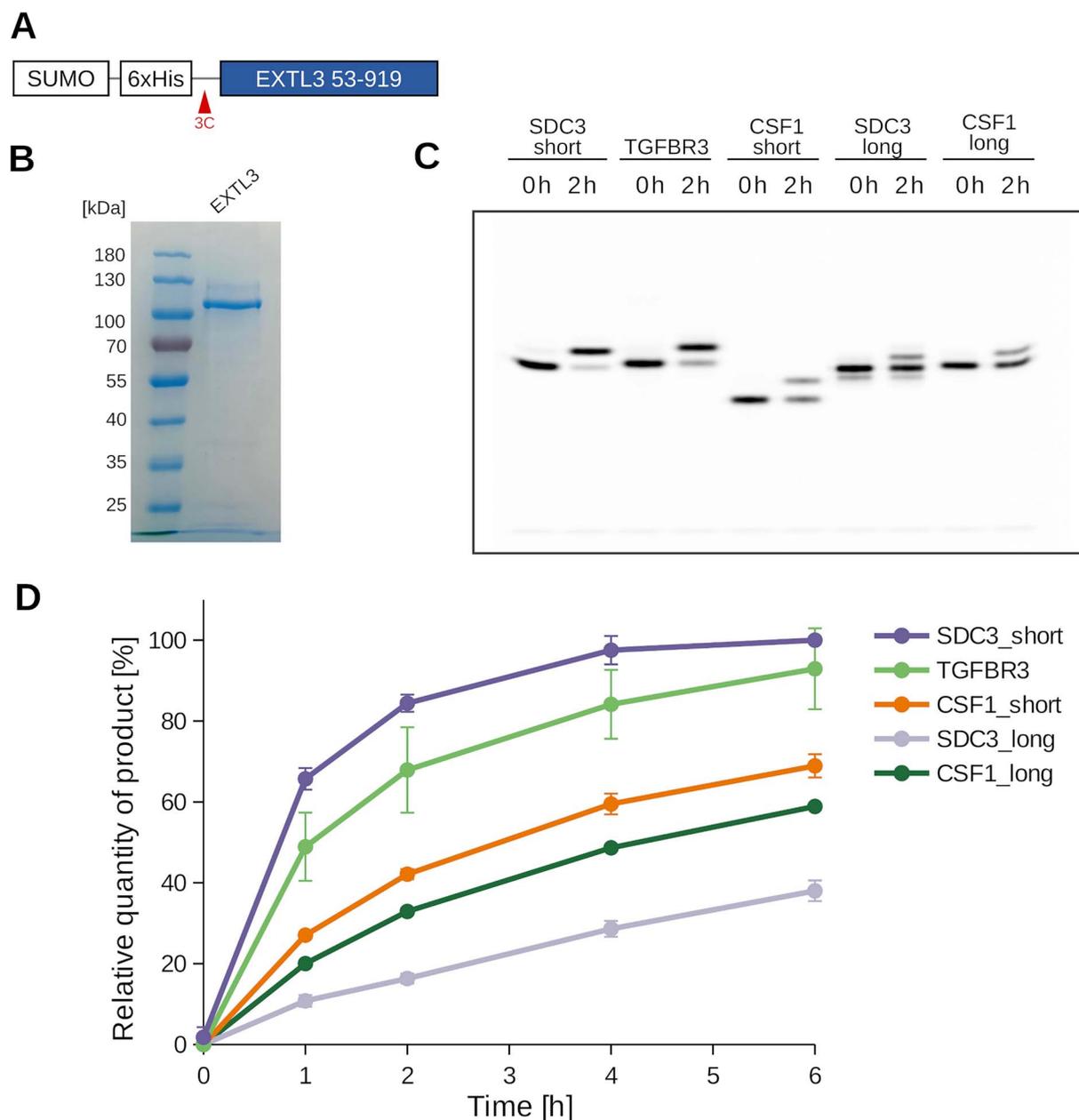


Fig. 6. Substrate specificity tests using recombinant EXTL3. A) Schematic representation of the recombinant EXTL3 construct. The N-terminal anchoring helix is omitted, and the construct features a cleavable N-terminal SUMO fusion protein followed by a 6xHistidine purification tag. B) Coomassie-stained SDS-PAGE analysis of purified EXTL3, with a theoretical molecular weight of 99 kDa. (C) Analysis of EXTL3 activity towards five distinct tetrasaccharide peptides: SDC3_short (³¹¹GGPSGDFE), TGFBR3 (⁵³¹GDSSGWP), CSF1_short (³⁰⁶EEASGEAS), SDC3_long (³⁰⁹VSGGSPGDFELPEEET) and CSF1_long (³⁰⁴VPEEASGEASEIPVPO). Reactions after 0 and 2 h incubation time were analyzed by gel electrophoresis (see Supplementary Fig. 4 for all reaction points). D) GlcNAc transfer by EXTL3 onto the five different peptides was quantified (see polyacrylamide gels in Supplementary Fig. 4). Product formation was followed over 6 h. For the SDC3_long peptide, which harbors two tetrasaccharides, the addition of either one or two GlcNAc moieties could be observed. For calculating the conversion rate, the addition of at least one GlcNAc residue was considered.

Linker generation was previously shown to be fine-tuned by xylose phosphorylation/dephosphorylation (Koike et al. 2014). Phosphorylation by the enzyme FAM20B most likely occurs after [Gal β 1-4Xyl]-peptide formation by β 4GalT7 (Gulberti et al. 2005) and was further shown to enhance β 3GalT6 activity (Wen et al. 2014; Sammon et al. 2023). Of note, we could observe full completeness for the β 3GalT6-catalyzed reaction in the absence of phosphorylation under our chosen reaction conditions, which included sufficient amounts of recombinant enzyme and relatively long incubation times (Fig. 3).

Next, we examined the substrate specificity of EXTL3 using purified tetrasaccharide peptides. The most rapid GlcNAc addition was observed for syndecan-3-derived peptide SDC3_short (HSPG), followed by the TGF β type III receptor-derived peptide TGFBR3 (HSPG) and the macrophage colony-stimulating factor 1 peptide CSF1_short (CSPG) (Fig. 6). Notably, this substrate preference of EXTL3 for peptides derived from HSPGs over CSPGs has recently been reported by Sammon et al. (Sammon et al. 2023). To test the effect of peptide length on EXTL3 activity, we designed peptides SDC3_long and CSF1_long that contained two additional

amino acid residues at the N-terminal end and six additional residues at the C-terminal end. Both peptides were converted less efficiently compared to their shorter versions, with a larger difference observed between peptides SDC3_short and SDC3_long. The SDC3_long peptide contains three additional Glu residues compared to SDC3_short and it has been recently proposed by Sammon et al. that acid residues bind to a basic surface patch in EXTL3. We would have thus expected an increased affinity for the peptide SDC3_long over SDC3_short. One possible explanation for a reduced activity of EXTL3 for the SDC3_long peptide could be that it carries two glycosylation sites, of which one has not been observed before, and that the additional tetrasaccharide blocks access to the active site in EXTL3. Our findings suggest that the sequence in close proximity to the glycosaminoglycan attachment site contains the information recognized by EXTL3. Therefore, EXTL3 emerges as a key player in the decision-making process during glycosaminoglycan chain biosynthesis.

Nevertheless, CSPG peptides were also processed by EXTL3 and the differences in conversion between the different peptides do not seem sufficient to explain the degree of specificity during glycosaminoglycan chain addition observed in the cell. We cannot exclude that this is due to our in vitro reaction conditions or the lack of xylose phosphorylation, which was shown to enhance EXTL3 activity (Sammon et al. 2023). Additionally, the absence of the N-terminal transmembrane anchoring helix in the EXTL3 construct could impact its activity. On the other hand, it might also hint that regulation of the divergent step is more intricate and might involve additional cellular factors or larger parts of the PG. It could also represent a potential rescue mechanism, in which EXTL3 adds HS chains onto CSPG core proteins in the absence of enzymes initiating CS biosynthesis (Takeuchi et al. 2013). These intricate regulatory mechanisms warrant further investigation, including substrate specificity assessments of CSGALNACT-1/CSGALNACT-2, competition assays between CSGALNACTs and EXTL3, and testing peptides/core proteins of varying lengths. Furthermore, the influence of modifications of the tetrasaccharide linker, such as phosphorylation of the Xyl residue, can be explored as additional regulatory mechanisms (Koike et al. 2014; Wen et al. 2014).

In conclusion, our established system for generating fluorescent peptides provides an efficient means to characterize different peptide substrates in the divergent step of GAG biosynthesis. This methodology paves the way for future research in unraveling the intricate regulatory network governing GAG biosynthesis, shedding light on its fundamental biological significance.

Acknowledgments

The authors thank S. Fournel-Gigleux for sharing the pETM30- β 4GalT7 expression plasmid, J. Rocha and C. Breton for the corresponding purification protocol, and J. Engelheert for providing the pHR-CMV-TetO2 vector. We are very grateful to F. Leisico for critically reading the manuscript and M. Friedel for assisting in enzymatic assays. We thank Caroline Mas for assistance and access to the Biophysical platform. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble

Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

Supplementary material

Supplementary material is available at *Glycobiology Journal* online.

Author contributions

R.W. designed and supervised the study. F.F. cloned β 3GalT6 and EXTL3 expression vectors. M.B. cloned XylT1 and GlcAT-1 expression vectors. M.B. and M.W. expressed and purified all proteins used in the study. F.F., M.B. and M.W. performed enzymatic reactions. M.B. and M.W. purified the glycosylated peptides. R.W., M.B., F.F. and M.W. analyzed the data. E.B.-E. performed and analyzed mass spectrometry experiments. M.B. and R.W. wrote the manuscript. All authors read and revised the manuscript.

CRedit author statement

Marie Bourgeais (Data curation [lead], Formal analysis [equal], Investigation [lead], Methodology [lead], Visualization [equal], Writing—original draft [equal]), Farah Fouladkar (Data curation [supporting], Investigation [supporting], Methodology [supporting], Writing—review & editing [supporting]), Margot Weber (Investigation [supporting], Methodology [supporting]), Elisabetta Boeri-Erba (Investigation [supporting], Methodology [supporting], Writing—review & editing [supporting]), and Rebekka Wild (Conceptualization [lead], Data curation [equal], Formal analysis [equal], Funding acquisition [lead], Investigation [supporting], Methodology [supporting], Project administration [lead], Resources [lead], Supervision [lead], Validation [lead], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Funding

This work was supported by the French National Center for Scientific Research in form of the ATIP-Avenir program [to R.W.] and the Schlumberger Foundation for Education and Research [to R.W.].

Conflict of interest statement. None declared.

Data availability

Source data are provided with this paper.

References

- Almeida R, Levery SB, Mandel U, Kresse H, Schwientek T, Bennett EP, Clausen H. Cloning and expression of a proteoglycan UDP-galactose:beta-xylose beta1,4-galactosyltransferase I. A seventh member of the human beta4-galactosyltransferase gene family. *J Biol Chem.* 1999;274(37):26165–26171.
- Annaval T, Wild R, Créton Y, Sadir R, Vivès RR, Lortat-Jacob H. Heparan sulfate proteoglycans biosynthesis and post synthesis mechanisms combine few enzymes and few core proteins to generate extensive structural and functional diversity. *Molecules.* 2020;25(18):E4215.
- Bai X, Zhou D, Brown JR, Crawford BE, Hennes T, Esko JD. Biosynthesis of the linkage region of glycosaminoglycans: cloning and activity of galactosyltransferase II, the sixth member of the beta 1,3-galactosyltransferase family (beta 3GalT6). *J Biol Chem.* 2001;276(51):48189–48195.
- Behiels E, Elegheert J. High-level production of recombinant eukaryotic proteins from mammalian cells using lentivirus. *Methods Mol Biol.* 2021;2305:83–104.

- Bishop JR, Schuksz M, Esko JD. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*. 2007;446(7139):1030–1037.
- Bludau I, Willems S, Zeng W-F, Strauss MT, Hansen FM, Tanzer MC, Karayel O, Schulman BA, Mann M. The structural context of posttranslational modifications at a proteome-wide scale. *PLoS Biol*. 2022;20(5):e3001636.
- Briggs DC, Hohenester E. Structural basis for the initiation of glycosaminoglycan biosynthesis by human Xylosyltransferase 1. *Structure*. 2018;26(6):801–809.e3.
- Carey DJ, Evans DM, Stahl RC, Asundi VK, Conner KJ, Garbes P, Cizmeci-Smith G. Molecular cloning and characterization of N-syndecan, a novel transmembrane heparan sulfate proteoglycan. *J Cell Biol*. 1992;117(1):191–201.
- Chatron-Colliet A, Brusa C, Bertin-Jung I, Gulberti S, Ramalanjaona N, Fournel-Gigleux S, Brézillon S, Muzard M, Plantier-Royon R, Rémond C, et al. ‘Click’-xylosides as initiators of the biosynthesis of glycosaminoglycans: comparison of mono-xylosides with xylobiosides. *Chem Biol Drug Des*. 2017;89(3):319–326.
- Chernousov MA, Carey DJ. N-syndecan (syndecan 3) from neonatal rat brain binds basic fibroblast growth factor. *J Biol Chem*. 1993;268(22):16810–16814.
- Connell BJ, Lortat-Jacob H. Human immunodeficiency virus and heparan sulfate: from attachment to entry inhibition. *Front Immunol*. 2013;4:385.
- Connell BJ, Sadir R, Baleux F, Laguri C, Kleman J-P, Luo L, Arenzana-Seisdedos F, Lortat-Jacob H. Heparan sulfate differentially controls CXCL12 α - and CXCL12 γ -mediated cell migration through differential presentation to their receptor CXCR4. *Sci Signal*. 2016;9(452):ra107.
- van Damme T, Pang X, Guillemin B, Gulberti S, Syx D, de Rycke R, Kaye O, de Die-Smulders CEM, Pfundt R, Kariminejad A, et al. Biallelic B3GALT6 mutations cause spondylodysplastic Ehlers-Danlos syndrome. *Hum Mol Genet*. 2018;27(20):3475–3487.
- Delbaere S, De Clercq A, Mizumoto S, Noborn F, Bek JW, Alluyn L, Gistelincx C, Syx D, Salmon PL, Coucke PJ, et al. B3galt6 knockout zebrafish recapitulate β 3GalT6-deficiency disorders in human and reveal a Trisaccharide proteoglycan linkage region. *Front Cell Dev Biol*. 2020;8:597857.
- Elegheert J, Behiels E, Bishop B, Scott S, Woolley RE, Griffiths SC, Byrne EFX, Chang VT, Stuart DI, Jones EY, et al. Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat Protoc*. 2018;13(12):2991–3017.
- Gao J, Lin P-H, Nick ST, Huang J, Tykesson E, Ellervik U, Li L, Huang X. Chemoenzymatic synthesis of Glycopeptides bearing galactose-xylose disaccharide from the proteoglycan linkage region. *Org Lett*. 2021;23(5):1738–1741.
- Gondelaud F, Bouakil M, Le Fèvre A, Miele AE, Chirof F, Duclos B, Liwo A, Ricard-Blum S. Extended disorder at the cell surface: the conformational landscape of the ectodomains of syndecans. *Matrix Biol Plus*. 2021;12:100081.
- Gopal S. Syndecans in inflammation at a glance. *Front Immunol*. 2020;11:227.
- Götting C, Kuhn J, Zahn R, Brinkmann T, Kleesiek K. Molecular cloning and expression of human UDP-d-xylose:proteoglycan core protein beta-d-xylosyltransferase and its first isoform XT-II. *J Mol Biol*. 2000;304(4):517–528.
- Gschwandtner M, Strutzmann E, Teixeira MM, Anders HJ, Diedrichs-Möhrling M, Gerlza T, Wildner G, Russo RC, Adage T, Kungl AJ. Glycosaminoglycans are important mediators of neutrophilic inflammation in vivo. *Cytokine*. 2017;91:65–73.
- Gulberti S, Lattard V, Fondeur M, Jacquinet J-C, Mulliert G, Netter P, Magdalou J, Ouzzine M, Fournel-Gigleux S. Phosphorylation and sulfation of oligosaccharide substrates critically influence the activity of human beta1,4-galactosyltransferase 7 (GalT-I) and beta1,3-galacturonosyltransferase I (GlcAT-I) involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J Biol Chem*. 2005;280(2):1417–1425.
- Gulberti S, Jacquinet J-C, Chabel M, Ramalanjaona N, Magdalou J, Netter P, Coughtrie MWH, Ouzzine M, Fournel-Gigleux S. Chondroitin sulfate N-acetylgalactosaminyltransferase-1 (CSGalNAcT-1) involved in chondroitin sulfate initiation: impact of sulfation on activity and specificity. *Glycobiology*. 2012;22(4):561–571.
- Guo H-B, Perminov A, Bekele S, Kedziora G, Farajollahi S, Varaljay V, Hinkle K, Molinero V, Meister K, Hung C, et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci Rep*. 2022;12(1):10696.
- Iozzo RV, Sanderson RD. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J Cell Mol Med*. 2011;15(5):1013–1031.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–589.
- Kellokumpu S, Hassinen A, Glumoff T. Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. *Cell Mol Life Sci*. 2016;73(2):305–325.
- Kim BT, Kitagawa H, Tamura J, Saito T, Kusche-Gullberg M, Lindahl U, Sugahara K. Human tumor suppressor EXT gene family members EXTL1 and EXTL3 encode alpha 1,4- N-acetylglucosaminyltransferases that likely are involved in heparan sulfate/heparin biosynthesis. *Proc Natl Acad Sci USA*. 2001;98(13):7176–7181.
- Kimura F, Suzu S, Nakamura Y, Wakimoto N, Kanatani Y, Yanai N, Nagata N, Motoyoshi K. Structural analysis of proteoglycan macrophage colony-stimulating factor. *J Biol Chem*. 1994;269(31):19751–19756.
- Kitagawa H, Shimakawa H, Sugahara K. The tumor suppressor EXT-like gene EXTL2 encodes an alpha1, 4-N-acetylhexosaminyltransferase that transfers N-acetylgalactosamine and N-acetylglucosamine to the common glycosaminoglycan-protein linkage region. The key enzyme for the chain initiation of heparan sulfate. *J Biol Chem*. 1999;274(20):13933–13937.
- Kitagawa H, Taoka M, Tone Y, Sugahara K. Human glycosaminoglycan glucuronyltransferase I gene and a related processed pseudogene: genomic structure, chromosomal mapping and characterization. *Biochem J*. 2001;358(3):539–546.
- Koike T, Izumikawa T, Sato B, Kitagawa H. Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *J Biol Chem*. 2014;289(10):6695–6708.
- Li J-P, Kusche-Gullberg M. Heparan Sulfate: biosynthesis, structure, and function. *Int Rev Cell Mol Biol*. 2016;325:215–273.
- López-Casillas F, Payne HM, Andres JL, Massagué J. Betaglycan can act as a dual modulator of TGF-beta access to signaling receptors: mapping of ligand binding and GAG attachment sites. *J Cell Biol*. 1994;124(4):557–568.
- Merry CLR, Lindahl U, Couchman J, Esko JD. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, et al., editors. *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY 11724: Cold Spring Harbor Laboratory Press; 2022. <http://www.ncbi.nlm.nih.gov/books/NBK579925/>
- Morla S. Glycosaminoglycans and glycosaminoglycan Mimetics in cancer and inflammation. *Int J Mol Sci*. 2019;20(8):1963.
- Nadanaka S, Zhou S, Kagiya S, Shoji N, Sugahara K, Sugihara K, Asano M, Kitagawa H. EXTL2, a member of the EXT family of tumor suppressors, controls glycosaminoglycan biosynthesis in a xylose kinase-dependent manner. *J Biol Chem*. 2013;288(13):9321–9333.
- Noborn F, Nilsson J, Larson G. Site-specific glycosylation of proteoglycans: a revisited frontier in proteoglycan research. *Matrix Biol*. 2022;111:289–306.
- Ozsan McMillan I, Li J-P, Wang L. Heparan sulfate proteoglycan in Alzheimer’s disease: aberrant expression and functions in molecular

- pathways related to amyloid- β metabolism. *Am J Physiol Cell Physiol.* 2023;324(4):C893–C909.
- Pedersen LC, Tsuchida K, Kitagawa H, Sugahara K, Darden TA, Negishi M. Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J Biol Chem.* 2000;275(44):34580–34585.
- Pedersen LC, Darden TA, Negishi M. Crystal structure of beta 1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *J Biol Chem.* 2002;277(24):21869–21873.
- Persson A, Nilsson J, Vorontsov E, Noborn F, Larson G. Identification of a non-canonical chondroitin sulfate linkage region trisaccharide. *Glycobiology.* 2019;29(5):366–371.
- Pönighaus C, Ambrosius M, Casanova JC, Prante C, Kuhn J, Esko JD, Kleesiek K, Goötting C. Human xylosyltransferase II is involved in the biosynthesis of the uniform tetrasaccharide linkage region in chondroitin sulfate and heparan sulfate proteoglycans. *J Biol Chem.* 2007;282(8):5201–5206.
- Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins. *J Mol Biol.* 2021;433(20):167208.
- Sammon D, Krueger A, Busse-Wicher M, Morgan RM, Haslam SM, Schumann B, Briggs DC, Hohenester E. Molecular mechanism of decision-making in glycosaminoglycan biosynthesis. *Nat Commun.* 2023;14(1):6425.
- Sato T, Gotoh M, Kiyohara K, Akashima T, Iwasaki H, Kameyama A, Mochizuki H, Yada T, Inaba N, Togayachi A, et al. Differential roles of two N-acetylgalactosaminyltransferases, CSGalNAcT-1, and a novel enzyme, CSGalNAcT-2. Initiation and elongation in synthesis of chondroitin sulfate. *J Biol Chem.* 2003;278(5):3063–3071.
- Signor L, Boeri Erba E. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometric analysis of intact proteins larger than 100 kDa. *J Vis Exp.* 2013;79(79):50635.
- Takeuchi K, Yoshioka N, Higa Onaga S, Watanabe Y, Miyata S, Wada Y, Kudo C, Okada M, Ohko K, Oda K, et al. Chondroitin sulphate N-acetylgalactosaminyl-transferase-1 inhibits recovery from neural injury. *Nat Commun.* 2013;4(1):2740.
- Theocharis AD, Manou D, Karamanos NK. The extracellular matrix as a multitasking player in disease. *FEBS J.* 2019;286(15):2830–2869.
- Toledo AG, Pihl J, Spliid CB, Persson A, Nilsson J, Pereira MA, Gustavsson T, Choudhary S, Zarni Oo H, Black PC, et al. An affinity chromatography and glycoproteomics workflow to profile the chondroitin sulfate proteoglycans that interact with malarial VAR2CSA in the placenta and in cancer. *Glycobiology.* 2020;30(12):989–1002.
- Tsutsui Y, Ramakrishnan B, Qasba PK. Crystal structures of β -1,4-galactosyltransferase 7 enzyme reveal conformational changes and substrate binding. *J Biol Chem.* 2013;288(44):31963–31970.
- Vallet SD, Berthollier C, Ricard-Blum S. The glycosaminoglycan interactome 2.0. *Am J Physiol Cell Physiol.* 2022;322(6):C1271–C1278.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022;50(D1):D439–D444.
- Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology.* 2015;25(12):1323–1324.
- Wen J, Xiao J, Rahdar M, Choudhury BP, Cui J, Taylor GS, Esko JD, Dixon JE. Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc Natl Acad Sci USA.* 2014;111(44):15723–15728.
- Wilson LFL, Dendooven T, Hardwick SW, Echevarría-Poza A, Tryfona T, Krogh KBRM, Chirgadze DY, Luisi BF, Logan DT, Mani K, et al. The structure of EXTL3 helps to explain the different roles of bi-domain exostosins in heparan sulfate synthesis. *Nat Commun.* 2022;13(1):3314.
- Yamada S, Sugahara K, Özbek S. Evolution of glycosaminoglycans. *Commun Integr Biol.* 2011;4(2):150–158.
- Zhang L, David G, Esko JD. Repetitive Ser-Gly sequences enhance heparan sulfate assembly in proteoglycans. *J Biol Chem.* 1995;270(45):27127–27135.