# Discovering Biological Roles of Glycosaminoglycans and Protein *O*-GlcNAcylation Using Chemical Tools

Thesis by

**Matthew Everett Griffin**

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy


CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2017

(Defended June 1, 2017)

Matthew Everett Griffin
ORCID: 0000-0001-9549-4418

# ACKNOWLEDGEMENTS

I would first and foremost like to express my gratitude to my advisor, Linda Hsieh-Wilson, for providing me with the encouragement, resources, and expertise to tackle such interesting and exciting chemical biology questions. I would also like to thank my committee members, Dennis Dougherty, Rustem Ismagilov, and Paul Sternberg, for their insightful comments and suggestions both on my research and my future career plans. I would like to thank my collaborators Edward Driggers, Dan Mason, Eric Peters, Mona Shahgholi, and Jost Vielmetter as well for their thoughtful contributions.

I would also like to thank the current and former members of the Hsieh-Wilson lab for their intelligence and patience while also making the lab such an enjoyable work environment. In particular, I would like to acknowledge Abigail Pulsipher, whose brilliant and kindhearted nature has been such an integral and cherished part of my formal education and circle of friends. I would also like to especially thank all of the lifelong friends that I have made in the lab: Greg Miller, Elizabeth Jensen, Yao Xiao, John Thompson, Alex Sorum, Yelena Koldobskaya, Cori Jenkins, Andrew Wang, Jean-Luc Chaubard, Sheldon Cheung, Kuang Wei Yang, Young In Oh, Chithra Krishnamurthy, Gloria Sheng, and Song Gil Lee. In addition to your unquestionable intelligence, you have all been a beacon of happiness throughout my time here for which I am indescribably grateful.

In addition to the scientific peers who have helped to shape my time here, I would never have been able to get off the proverbial ground without such a hardworking infrastructure whose work goes far too often unrecognized. In particular, I am very grateful to Elisa Brink, Susan Mucha, Joe Drew, and Agnes Tong, who have always had an open door and open ear for my requests and hopefully infrequent grievances.

I am also indebted to my friends outside of lab who have brightened my life during my time here. Although too many to name, I would like to especially thank Kelvin Bates, Christopher Haley, LeeAnn Suen, Matt Rienzo, Annet Blom, Matt Smarte, Marcus Low, Noah Duffy, Renee Arias, Keith Beadle, Belinda Wenke, Andrew Buller, Maddi Kieffer, Chris Marotta, Doug Duquette, Beau Pritchett, Ben Stagg, Alexis Komor, Myles Herbert, and anyone who I have forgotten in my old age.

Team 151 – Helen Segal, Eric Lubeck, and Connie Wang – you have become my California family and have given me some of my happiest memories here.

My Louisiana family – Stephen, Lynn, Andrew, and Lauren – I cannot describe how much I appreciate your encouragement when I have been down, your patience when I have been overbearing, and your love at all times. I love you.

# ABSTRACT

Carbohydrates surround nearly every cell in the human body. Glycosaminoglycans like chondroitin sulfate and heparan sulfate on the cell surface regulate protein ligand engagement and receptor activation to control a variety of biological processes including development, angiogenesis, and neuronal growth. These polysaccharides exert activity through protein binding to their diverse chemical structures. Therefore, the development of methods to tailor glycosaminoglycan populations at the cell surface with defined structures could provide novel approaches to control biological activity. Herein, two new methods to engineer the cell surface glycocalyx with known glycosaminoglycans are reported. Together, these methods provide complementary short- and long-term approaches to change carbohydrate structures at the cell surface to guide neuronal growth and stem cell differentiation. It is also critical to identify unknown protein-carbohydrate interactions that underlie biological phenomena. Novel GAG interactions with the orphan receptor Tie1 and angiopoietin ligands Ang1 and Ang4 are reported herein. These interactions provide the first evidence for a physiological ligand for Tie1 since its discovery over 25 years ago. Moreover, interactions between Ang1 and Ang4 and heparan sulfate are shown to potentiate Tie2 survival signaling, providing novel insights into an important but poorly understood signaling axis.

Within the cell, thousands of proteins are modified by *O*-GlcNAc glycosylation, a process that is uniquely catalyzed by a single transferase and hydrolase pair unlike many other post-translational modifications. *O*-GlcNAcylation functions in many biological contexts including transcription, translation, proteostasis, and metabolism. Key to understanding its effects on these physiological phenomena is the discovery of *O*-GlcNAc modification sites. However, due to a number of technical challenges, *O*-GlcNAc proteomics has not

progressed nearly as quickly as phosphoproteomics. Thus, developing new methods to enrich *O*-GlcNAcylated substrates and map modification sites is critical to unravel the myriad functions of *O*-GlcNAc. Herein, a labeling approach using a chemically cleavable tag is reported as an improved method to capture and release *O*-GlcNAcylated substrates. Unlike other methods, the cleavable Dde tag used herein is quantitatively removed under mild, neutral conditions and leaves a minimal residual tag on the *O*-GlcNAcylated peptide to be analyzed. Moreover, the Dde linker outcompetes a UV-cleavable tag that has been previously utilized in the field, identifying 414 unique *O*-GlcNAcylated peptides. Together, these results highlight the potential usefulness of the new approach developed herein to illuminate novel roles of *O*-GlcNAcylation in diverse systems.

# PUBLISHED CONTENT AND CONTRIBUTIONS

Pulsipher A, **Griffin ME**, Stone SE, Brown JM, Hsieh-Wilson LC. "Directing neuronal signaling through cell-surface glycan engineering." *J. Am. Chem. Soc.* **2014**, *136* (19): 6794-6797. doi: 10.1021/ja5005174. Research article.

M.E.G. participated in the conception and execution of experiments, including chemical synthesis of glycosaminoglycan probes and surface longevity experiments, and in writing of the manuscript. This article, including figures, is reproduced in part within Chapter 2 with permission according to the ACS AuthorChoice terms of use.

Pulsipher A*, **Griffin ME***, Stone SE, Hsieh-Wilson LC. "Long-lived glycan engineering to direct stem cell fate." *Angew. Chem. Int. Ed.* **2015**, *54* (5): 1466-1470. doi: 10.1002/anie.201409258; * denotes equal contribution. Research article.

M.E.G. participated in the conception and execution of experiments, including chemical synthesis of glycosaminoglycan probes, phosphorylation assays, and microarray analysis, and in writing of the manuscript. The article, including figures, is reproduced in part within Chapter 2 with permission from the publisher.

**Griffin ME**, Hsieh-Wilson LC. "Glycan engineering for cell and developmental biology." *Cell Chem. Biol.* **2016**, *23* (1): 108-121. doi: 10.1016/j.chembiol.2015.12.007. Review article.

M.E.G. participated in the conception and writing of the manuscript. The article, including figures, is reproduced in part within Chapter 1 with permission from the publisher.

**Griffin ME**, Jensen EH, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. "Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag." *Mol. Biosys.* **2016**, *12* (6): 1756-1759. doi: 10.1039/c6mb00138f. Research article.

M.E.G. participated in the conception and execution of all experiments, data analysis, and in writing of the manuscript. The article, including figures, is reproduced in part within Chapter 5 with permission under a Creative Commons Attribution 3.0 Unported License.

# TABLE OF CONTENTS

# LIST OF FIGURES AND TABLES

# NOMENCLATURE

| | |
|---|---|
| Ang (Angpt) | Angiopoietin |
| AF*n* | AlexaFluor *n* |
| Ala | alanine |
| B | biotin/biotinylated |
| BCA | bicinchoninic acid |
| BEMA(D) | β-elimination Michael addition (DTT) |
| BMP | bone morphogenic protein |
| Boc | *tert*-butyloxycarbonyl |
| Bp | base pair |
| BSA | bovine serum albumin |
| CBP | CREB-binding protein |
| CCD | charge coupled device |
| ChABC | chondroitinase ABC |
| CHO | Chinese hamster ovary |
| Chst | carbohydrate sulfotransferase |
| CID | collision-induced dissociation |
| CL | chlorohexyl linker |
| c-fms | colony stimulating factor 1 |
| CREB | cAMP response element binding |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| CS-(A/C/E) | chondroitin sulfate (A/C/E) |
| CuAAC | copper-catalyzed azide-alkyne cycloaddition |
| Cys | cysteine |
| Cy3 | cyanine 3 |
| DAPI | 4’,6-diamidino-2-phenylindole |
| DCM | dichloromethane |

| | |
|---|---|
| Dde | 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl |
| Ddv/ivDde | 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)isovaleryl |
| de*X* | de-*X*-sulfated |
| DIPEA | *N*,*N*-diisopropylethylamine |
| DLS | dynamic light scattering |
| DMEM | Dulbecco's modified Eagle medium |
| DMF | *N*,*N*-dimethylformamide |
| DNA | deoxyribonucleic acid |
| DOPE | 1,2-dioleoyl-*sn*-glycero-3-phosphatidylethanolamine |
| DOTAP | 1,2-dioleoyl-3-trimethylammoniumpropane chloride |
| DS | dermatan sulfate |
| dsDNA | double stranded DNA |
| DTSSP | 3,3'-dithiobis(sulfosuccinimidyl propionate) |
| DTT | dithiothreitol |
| EDC | 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| EDTA | ethylenediaminetetraacetic acid |
| EGTA | ethylene glycol-bis(β-aminoethyl ether)tetraacetic acid |
| Efn | ephrin |
| EGF | epidermal growth factor |
| E*n* | embryonic day *n* |
| ELISA | enzyme-linked immunosorbent assay |
| Eph | erythropoietin-producing human hepatocellular receptor |
| ERK | extracellular signal-regulated kinase |
| (m)ESC | (murine) embryonic stem cell |
| ETD | electron transfer dissociation |
| Ext1 | exostosis type 1 |
| F | 5-(((2-(carbohydrazino)methyl)thio)acetyl)aminofluorescein |
| FBS | fetal bovine serum |
| Fc | fraction crystallizable |

| | |
|---|---|
| FACS | fluorescence-assisted cell sorting |
| FGF(R) | fibroblast growth factor (receptor) |
| FNIII | fibronectin-like III |
| GAG | glycosaminoglycan |
| Gal | galactose |
| GALE | UDP-galactose 4'-epimerase |
| GalN | galactosamine |
| GalNAc | *N*-acetylgalactosamine |
| GalNAz | *N*-azidoacetylgalactose |
| (Y289L) GalT | (Y289L) β-1,4-galactosyltransferase |
| gDNA | genomic DNA |
| GFP | green fluorescent protein |
| Glc | glucose |
| GlcA | glucuronic acid |
| GlcN | glucosamine |
| GlcNAc | *N*-acetylglucosamine |
| GlcNAz | *N*-azidoacetylglucosamine |
| Gly | glycine |
| HA | hyaluronic acid OR hemagglutinin tag |
| HBSS | Hank's buffered saline solution |
| HCD | higher-energy collisional dissociation |
| HDR | homology-directed repair/recombination |
| HEK | human embryonic kidney |
| hep | heparin |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HLB | hydrophilic-lipophilic-balanced |
| HS | heparan sulfate OR horse serum |
| HTP | HaloTag protein |
| HUVEC | human umbilical vein endothelial cell |

| | |
|---|---|
| ICC | immunocytochemistry |
| IdoA | iduronic acid |
| IF | immunofluorescence |
| Ig | immunoglobulin domain |
| Ile | isoleucine |
| IRDye800 | infrared dye 800 |
| ketoGal | 2-acetonyl-2-deoxygalactose |
| KS | keratin sulfate |
| Leu | leucine |
| LTQ | linear trap quadrupole |
| LSM | laser scanning microscope |
| LV | lentivirus |
| LWAC | lectin weak affinity chromatography |
| ManNAz | *N*-azidoacetylmannosamine |
| MAPK | mitogen-activated protein kinase |
| MEF | murine embryonic fibroblasts |
| MeOH | methanol |
| MES | 2-(4-morpholino)ethanesulfonic acid |
| Met | methionine |
| MOE | metabolic oligosaccharide engineering |
| MS | mass spectrometry |
| NaAsc | sodium ascorbate |
| NANOG | homeobox transcription factor Nanog |
| Ndst | *N*-deacetylase *N*-sulfotransferase |
| NGF | nerve growth factor |
| NHEJ | non-homologous end joining |
| NHS | *N*-hydroxysuccinimide |
| NMR | nuclear magnetic resonance |
| Nrp1 | neuropilin 1 |

| | |
|---|---|
| nt | nucleotide |
| OGA | *O*-GlcNAcase |
| (s)OGT | (short form) *O*-GlcNAc transferase |
| PAM | protospacer motif |
| PARP1 | poly-ADP ribose polymerase 1 |
| PBS(T) | phosphate buffered saline (Tween 20) |
| PC | photocleavable |
| PC12 | pheochromocytoma 12 (cell line) |
| PCR | polymerase chain reaction |
| PDGFR | platelet-derived growth factor receptor |
| PEG | polyethylene glycol |
| PFK1 | phosphofructokinase I |
| (CS/HS)PG | proteoglycan |
| Phe | phenylalanine |
| PI3K | phosphatidylinositol 3-kinase |
| PLA | proximity ligation assay |
| P/S | penicillin/streptomycin |
| PTM | post-translational modification |
| qRT-(PCR) | quantitative reverse transcription / real time |
| RNA | ribonucleic acid |
| RNAi | RNA interference |
| RTK | receptor tyrosine kinase |
| SCX | strong cation exchange |
| SDHA | succinic dehydrogenase A |
| SDS-PAGE | sodium dodecyl sulfate protein acrylamide gel electrophoresis |
| S.E.M. | standard error of the mean |
| Ser | serine |
| SOX1/2 | (sex determining region Y)-box 1/2 |
| SPR | surface plasmon resonance |

| | |
|---|---|
| ssODN | single stranded oligonucleotide |
| Stat5 | signal transducer and activator of transcription 5 |
| TBS(T) | Tris buffered saline (Tween 20) |
| TEM | transmission electron microscopy |
| TFA | trifluoroacetic acid |
| THPTA | tris(3-hydroxypropyltriazolylmethyl)amine |
| Thr | threonine |
| Tie1/2 | tyrosine kinase with immunoglobulin-like and EGF-like domains 1/2 |
| TNFα | tumor necrosis factor α |
| Trk | tropomyosin receptor kinase |
| Trp | tryptophan |
| TUJ1 | β-tubulin III |
| Tyr | tyrosine |
| UDP | uridine diphosphate |
| UV | ultraviolet |
| VEGF(R) | vascular endothelial growth factor (receptor) |
| (s)WGA | (succinylated) wheat germ agglutinin |
| Wnt | wingless-type MMTV integration site family member |
| Xyl | xylose |

*Chapter 1*

# Glycosaminoglycans as Active Signaling Components of the Extracellular Matrix

Portions of this chapter are published as:

**Griffin ME**, Hsieh-Wilson LC. "Glycan engineering for cell and developmental biology." *Cell Chem. Biol.* **2016**, *23*: 108-121. doi: 10.1016/j.chembiol.2015. 12.007. Review article.

**1.1 Glycosaminoglycan Structures and Biosynthesis**

Carbohydrates are generally thought of as a fuel source for life. However, these molecules also function in many other roles necessary for survival including development, angiogenesis, and neuronal growth.[1-5] In particular, carbohydrates at the cell surface can strongly regulate signal transduction and cellular activity. This feat is achieved in large part through their structural diversity, which allows them to selectively bind to a variety of different proteins and in turn modulate their functions. Unsurprisingly, the dysregulation of cell-surface carbohydrate production and presentation can contribute to a variety of diseases including inflammation and cancer progression.[6, 7] Therefore, discovering relationships between the chemical structures of carbohydrates, the proteins to which they bind, and the resulting biological functions is critical both for the basic understanding of many physiological processes and for the prevention and treatment of various pathologies.

Cell-surface carbohydrates exist in a variety of forms and are classified based on their overall size, membrane anchor, monosaccharide composition, glycosidic connections, and further modifications of the monosaccharide residues.[8] Of particular interest is the class of carbohydrates known as glycosaminoglycans (GAGs), which exist as linear polysaccharides of generally 20 to 200 repeating disaccharide units.[8] GAGs and their attached proteins, known collectively as proteoglycans, are almost ubiquitously present at the cell surface either anchored in the cell membrane or secreted into the extracellular matrix; however, the

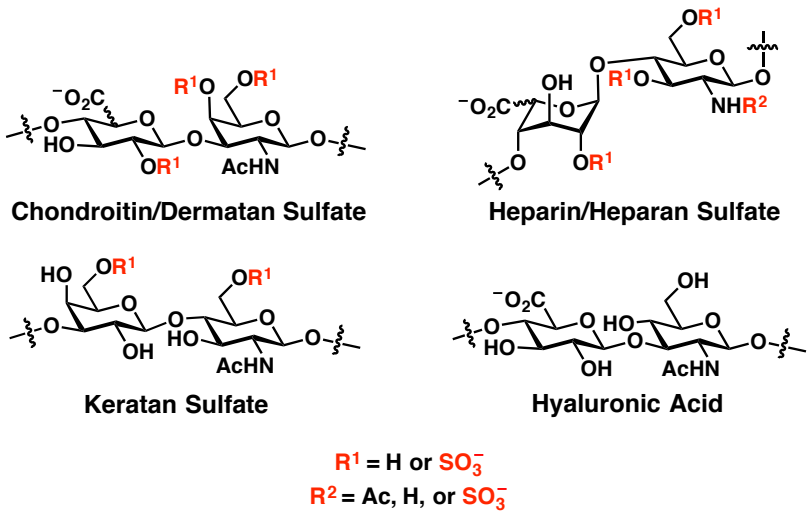



**Figure 1.1 Structures of GAG family members**. GAGs are made of repeating disaccharide units. Three of the four families can be sulfated at different hydroxyl groups along the polysaccharide.

chemical composition of GAG structures varies significantly between cell types. GAGs can be subdivided into four main classes based on their monosaccharide components: (1) chondroitin sulfate and dermatan sulfate, (2) heparin and heparan sulfate, (3) keratan sulfate, and (4) hyaluronan (Figure 1-1). Chondroitin sulfate and dermatan sulfate (CS/DS) contain a repeating disaccharide unit of a hexuronic acid (GlcA for CS and both GlcA and IdoA for DS) and GalNAc. Heparin and heparan sulfate (hep/HS) are made of a mixture of GlcA and IdoA (~10% GlcA for heparin and 10-50% GlcA for HS) along with GlcNAc, which can be deacetylated as GlcN. Keratan sulfate (KS) exists as repeats of Gal and GlcNAc. These three GAG structures are synthesized in the Golgi apparatus on their proteoglycan cores and can also be differentially sulfated on their hydroxyl groups. The fourth class of GAGs, hyaluronan or hyaluronic acid (HA), is unique in that it is not attached to a protein structure and is generally much larger than

the other three classes, existing upwards of 5 MDa or more in mass and 20 μm in length.[9] Furthermore, the repeating unit of HA (GlcA and GlcNAc) is assembled at the cell membrane by hyaluronic acid synthases and is completely unsulfated. For the purpose of brevity, only CS/DS and hep/HS will be discussed further.

The repeating patterns of CS/DS and hep/HS can be classified further by their sulfation patterns. In mammals, CS and DS can be sulfated on the GlcA/IdoA-C2, GalNAc-C4, and GalNAc-C6 hydroxyl positions. Hep and HS can be modified at the GlcA/IdoA-C2, GlcN-C3, and GlcN-C6 hydroxyl positions as well as the free amine on GlcN. These modifications are catalyzed by carbohydrate sulfotransferases that reside in the Golgi apparatus.[10] Sulfate groups can be removed by sulfatases; however, this generally occurs in lysosomes after GAG internalization during degradation.[11] Only two enzymes, HS 6-*O*-endosulfatases Sulf1 and Sulf2, are known to modify GAG structures once they are secreted to the cell surface.[12]

CS/DS and hep/HS biosynthesis follow similar pathways.[13, 14] First, a common tetrasaccharide (Xyl-Gal-Gal-GlcA) is appended to Ser residues of various proteins. Usually, this occurs at Ser-Gly/Ala-X-Gly sequences;[15] however, this motif is not universally found at all modification sites. Next, the first *N*-acetylhexosamine residue is attached to the core tetrasaccharide. This step commits the growing strand to either CS/DS (GalNAc) or hep/HS (GlcNAc). The identity of the attached GAG is dependent on a number of factors including the protein core.[16] Although certain proteins like versican and glypicans are modified

only by CS or HS, respectively, others like syndecans can contain both structures simultaneously, whereas some proteins like neuropilin-1 have different structures attached to the same Ser residue on different copies of the protein.[17, 18] The chains are then elongated by a number of polymerizing enzymes. For CS, GlcA can be epimerized to IdoA at this point to produce regions of DS. For hep/HS, GalNAc is first de-*N*-acetylated and *N*-sulfated, after which GlcA residues can be epimerized to IdoA. Finally, sulfotransferases modify hydroxyl groups along the length of the carbohydrate backbone to produce the mature sulfated polysaccharide. Importantly, these modification reactions do not proceed to completion, increasing heterogeneity of the final structure. Moreover, other than substrate preferences exhibited by the modifying enzymes, it is relatively unclear how the cell orchestrates structural heterogeneity of the produced GAG polysaccharides. It is hypothesized that regions of high and low sulfation density exist along the oligosaccharide, but little structural information is directly available due to the difficulties in GAG sequencing.[19-21] Nevertheless, it is through these different structures that GAGs exert their biological activity by binding to proteins. Thus, to fully understand the biological activity of GAGs, it is critical (1) to discover the interactions between proteins and specific GAG structures and (2) to design methods to change GAG structures at the cell surface to alter and perhaps control biological function.

**1.2 GAG Binding to the Receptor Tyrosine Kinase Superfamily**

To control intracellular activity, GAGs must interact with transmembrane proteins on the cell surface to transfer external information across the cell membrane. One of the largest groups of these proteins is the receptor tyrosine kinase (RTK) superfamily.[22, 23] Made up of 58 receptors organized into 20 subfamilies, RTKs are characterized by the presence of an intracellular tyrosine kinase domain that is activated by receptor dimerization, causing cross-phosphorylation of tyrosine residues on opposite receptors and leading to downstream activation of signaling pathways.[23] Their extracellular domains are much more structurally diverse, allowing RTKs to function in a variety of fundamental biological processes including cell survival and motility.[23-26] Moreover, their dysregulation has been linked to a variety of disease states including cancer, diabetes, and atherosclerosis.[27-29]

RTKs generally function through binding to extracellular ligands. Interestingly, many of the RTK subfamilies have been linked to interactions with GAGs – either through binding to soluble protein ligands or to the receptor itself.[30-41] One of the most famous examples of the involvement of GAGs is the ternary complex formed by fibroblast growth factor 1 (FGF1), FGF receptor 1 (FGFR1), and hep/HS.[30] As illustrated by crystallographic studies, hep/HS binds to the receptor-ligand complex using a binding site that spans both proteins in a 2:2:2 stoichiometric ratio, cooperating to stabilize the active receptor dimer. Carbohydrate microarray studies have also illustrated that FGFR1 binding to

hep/HS is facilitated by FGF2 binding,[36] providing mechanistic evidence for ternary complex assembly. Other examples of GAG binding to ligands and receptors include the vascular endothelial growth factor 1 (VEGF1) and VEGF receptor 1 (VEGFR1) system.[32, 42] Surface plasmon resonance (SPR) analyses have shown that VEGFR1 but not VEGFR2 can directly interact with hep/HS. As seen before with FGF2/FGFR1, VEGF1 binding to hep/HS facilitates the formation of a ternary complex containing VEGFR2. More recently, CS-E has been demonstrated to bind to members of the erythropoietin-producing human hepatocellular (Eph) family of receptors.[40, 41] CS-E binding to EphA4 and EphB3 can facilitate receptor activation without the canonical ephrin (Efn) ligands, highlighting a novel mechanism of action for RTK signal transduction mediated by GAG binding.

Together, these results illustrate only a small portion of the knowledge gained from the discovery of GAG-RTK interactions and their biological consequences. However, our understanding of the connections between GAG binding and RTK signaling is far from complete. Given the sheer size of the RTK superfamily and the diverse biological settings where they function, it is quite possible that many new GAG-RTK interactions with significant biological ramifications remain still undiscovered.

### 1.3 Altering Cell-Surface GAG Populations

A variety of approaches have been developed to alter GAG structures at the cell surface and observe the resulting biological phenotype. The majority of these methods are reductive, meaning that they remove GAG structures through biosynthetic inhibition or degradation, and occur through genetic manipulation. These approaches include gene deletion or knockout, gene knockdown by RNAi, and gene overexpression. Genetic methods offer excellent spatial and temporal control, enabling the precise manipulation of specific genes in a cell-specific and inducible manner. However, because of the linear synthesis of the GAG backbone and the substrate specificity of individual sulfotransferases, the genetic disruption of a single enzyme may lead to dramatic changes in GAG populations and are generally unsuitable to probe in importance of individual sulfation epitopes. For example, *N*-deacetylation and *N*-sulfation of GlcNAc in HS biosynthesis is critical for further sulfation reactions. Therefore, knockout of the responsible enzyme *N*-deacetylase and *N*-sulfotransferase 1 (Ndst1) leads not only to decreases in *N*-sulfation but also *O*-sulfation at all other positions.[2, 43] Similarly, the production of the CS-E epitope by carbohydrate sulfotransferase 15 (Chst15) requires the activity of Chst11 to first add a sulfate group to the GalNAc-C4 hydroxyl position and produce the CS-A epitope. Deletion of Chst11 leads to the loss of both the CS-A and CS-E motifs.[44] Moreover, knocking out GAG enzymes can lead to developmental defects or embryonic lethality, which can hinder the identification of functions in adult organisms.[45] Nonetheless, important discoveries have been

made regarding the necessity of GAGs for proper development using genetic approaches. For instance, the importance of HS in bone maturation has been demonstrated by the production of a mouse model deficient in HS biosynthesis due to a hypomorphic mutation in the HS polymerase Ext1.[46] These mice exhibited improper endochondral ossification during development, and this phenotype was attributed to the importance of HS binding with the growth factor Indian hedgehog.

An alternative reductive approach is the use of GAG degrading enzymes to selectively remove carbohydrates at the cell surface. As with genetic approaches, direct delivery or transgenic expression of the enzyme can be finely controlled to provide spatiotemporal selectivity. However, these enzymes lack fine substrate specificity and will at least partially degrade all CS or HS GAGs depending on the enzyme used. Furthermore, the longevity of this approach depends greatly on the stability of the enzyme, and long-term experiments may require multiple deliveries of the enzyme to avoid complications from newly synthesized GAGs. However, GAG degrading enzymes are invaluable tools to quickly and effectively remove nearly all GAGs of a specific subpopulation. One promising application of this method has been the delivery of chondroitinase ABC (ChABC) derived from the bacterium *Proteus vulgaris* to sites of spinal cord injury.[47-49] Reactive astrocytes produce large quantities of CS after injury to inhibit axonal regeneration,[50] and direct injection or viral delivery of ChABC has been associated with neuronal regrowth and functional recovery.[47, 48]

Recently, complementary additive approaches have been developed to overcome the obstacle of structural selectivity found in reductive methods. In these methods, known collectively as *de novo* glycan display, carbohydrate or glycomimetic structures are directly inserted into plasma membranes using approaches such as lipid insertion, liposomal fusion, or protein conjugation.[51-56] These techniques provide excellent control over glycan structure, allowing known epitopes to be displayed for functional analysis. However, exogenous sugars are typically displayed alongside the native glycan population, which could obscure the biological effects of the newly added carbohydrates. To address this complication, *de novo* glycan display methods can be used in combination with reductive approaches to minimize the contributions of interfering endogenous carbohydrates. The versatility of the technique also allows for the display of a wide range of carbohydrate-based structures, including glycomimetics such as synthetic glycopolymers, glycans appended to simplified proteins, or even the glycan component of glycoproteins alone. As a relatively new field, *de novo* glycan display has only been applied to a limited number of biological contexts. For example, anchoring of lactosyl or cellobiosyl-containing glycopolymers in the cell membrane by passive lipid insertion was used to examine galectin-mediated crosslinking and aggregation.[52] However, prior to the work outlined in Chapter 2, the ability to elicit biological activity as a function of glycan structures at the cell surface was unknown. Our work[53, 54], along with other, simultaneous publications in the field[55, 56], demonstrated that the display of defined carbohydrate structures

at the cell surface could be used to drive multiple biological processes including immunoevasion, neuronal outgrowth, and stem cell differentiation. Together, these results highlight the utility of *de novo* glycan display as a novel tool to directly connect carbohydrate structure and biological function unlike other existing methods.

## 1.4 References


1. X. Lin. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development*. **2004**, *131*: 6009-6021.
2. J. R. Bishop, M. Schuksz, J. D. Esko. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*. **2007**, *446*: 1030-1037.
3. R. V. Iozzo, J. D. San Antonio. Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena. *J Clin Invest*. **2001**, *108*: 349–355.
4. K. Sugahara, T. Mikami. Chondroitin/dermatan sulfate in the central nervous system. *Curr Opin Struct Biol*. **2007**, *17*: 536-545.
5. G. M. Miller, L. C. Hsieh-Wilson. Sugar-dependent modulation of neuronal development, regeneration, and plasticity by chondroitin sulfate proteoglycans. *Exp Neurol*. **2015**, *274*: 115-125.
6. K. R. Taylor, R. L. Gallo. Glycosaminoglycans and their proteoglycans: host-associated molecular patterns for initiation and modulation of inflammation. *FASEB J*. **2006**, *20*: 9-22.

7. R. Sasisekharan, Z. Shriver, G. Venkataraman, U. Narayanasami. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat. Rev. Cancer.* **2002**, *2*: 521-528.

8. J. D. Esko, K. Kimata, U. Lindahl. *Proteoglycans and Sulfated Glycosaminoglycans*. In: A. Varki *et al.*, Eds., *Essentials of Glycobiology*. 2nd Ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

9. D. Vigetti *et al.* Hyaluronan: biosynthesis and signaling. *Biochim Biophys Acta.* **2014**, *1840*: 2452-2459.

10. N. Gamage *et al.* Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci.* **2006**, *90*: 5-22.

11. H. H. Freeze. *Genetic Disorders of Glycan Degradation*. In: A. Varki *et al.*, Eds., *Essentials of Glycobiology*. 2nd Ed. Cold Spring Harbor (NY): 2009.

12. W. C. Lamanna *et al.* Heparan sulfate 6-*O*-endosulfatases: discrete in vivo activities and functional co-operativity. *Biochem J.* **2006**, *400*: 63-73.

13. T. Mikami, H. Kitagawa. Biosynthesis and function of chondroitin sulfate. *Biochim Biophys Acta.* **2013**, *1830*: 4719-4733.

14. K. Sugahara, H. Kitagawa. Heparin and heparan sulfate biosynthesis. *IUBMB Life.* **2002**, *54*: 163-175.

15. R. Raman, V. Sasisekharan, R. Sasisekharan. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem Biol.* **2005**, *12*: 267-277.

16. K. Prydz. Determinants of Glycosaminoglycan (GAG) Structure. *Biomolecules.* **2015**, *5*: 2003-2022.

17. R. V. Iozzo, L. Schaefer. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol.* **2015**, *42*: 11-55.

18. Y. Shintani *et al.* Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2 signaling. *EMBO J.* **2006**, *25*: 3045-3055.

19. J. T. Gallagher. Multiprotein signalling complexes: regional assembly on heparan sulphate. *Biochem Soc Trans.* **2006**, *34*: 438-441.

20. M. Ly *et al.* The proteoglycan bikunin has a defined sequence. *Nat Chem Biol.* **2011**, *7*: 827-833.

21. N. Volpi, R. J. Linhardt. High-performance liquid chromatography-mass spectrometry for mapping and sequencing glycosaminoglycan-derived oligosaccharides. *Nat Protoc.* **2010**, *5*: 993-1004.

22. J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell.* **2000**, *103*: 211-225.

23. M. A. Lemmon, J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell.* **2010**, *141*: 1117-1134.

24. T. N. Sato *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature.* **1995**, *376*: 70-74.

25. P. Duchek, P. Rorth. Guidance of cell migration by EGF receptor signaling during *Drosophila* oogenesis. *Science.* **2001**, *291*: 131-133.

26. P. Duchek, K. Somogyi, G. Jekely, S. Beccari, P. Rorth. Guidance of cell migration by the Drosophila PDGF/VEGF receptor. *Cell.* **2001**, *107*: 17-26.

27. E. B. Pasquale. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat Rev Cancer.* **2010**, *10*: 165-180.

28. J. Boucher, A. Kleinridders, C. R. Kahn. Insulin receptor signaling in normal and insulin-resistant states. *Cold Spring Harb Perspect Biol.* **2014**, *6*.

29. K. V. Woo *et al.* Tie1 attenuation reduces murine atherosclerosis in a dose-dependent and shear stress-specific manner. *J Clin Invest*. **2011**, *121*: 1624-1635.

30. J. Schlessinger *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Molecular Cell*. **2000**, *6*: 743-750.

31. C. Rolny, D. Spillmann, U. Lindahl, L. Claesson-Welsh. Heparin amplifies platelet-derived growth factor (PDGF)-BB-induced PDGF alpha-receptor but not PDGF beta-receptor tyrosine phosphorylation in heparan sulfate-deficient cells. Effects on signal transduction and biological responses. *J Biol Chem*. **2002**, *277*: 19315-19321.

32. D. Xu, M. M. Fuster, R. Lawrence, J. D. Esko. Heparan sulfate regulates $VEGF_{165}$- and $VEGF_{121}$-mediated vascular hyperpermeability. *J Biol Chem*. **2011**, *286*: 737-745.

33. L. E. Kemp, B. Mulloy, E. Gherardi. Signalling by HGF/SF and Met: the role of heparan sulphate co-receptors. *Biochem Soc Trans*. **2006**, *34*: 414-417.

34. V. Koprivica *et al.* EGFR activation mediates inhibition of axon regeneration by myelin and chondroitin sulfate proteoglycans. *Science*. **2005**, *310*: 106-110.

35. M. W. Barnett, C. E. Fisher, G. Perona-Wright, J. A. Davies. Signalling by glial cell line-derived neurotrophic factor (GDNF) requires heparan sulphate glycosaminoglycan. *J Cell Sci*. **2002**, *115*: 4495-4503.

36. C. J. Rogers *et al.* Elucidating glycosaminoglycan-protein-protein interactions using carbohydrate microarray and computational approaches. *Proc Natl Acad Sci USA*. **2011**, *108*: 9747-9752.

37. P. B. Murray *et al*. Heparin is an activating ligand of the orphan receptor tyrosine kinase ALK. *Sci Signal*. **2015**, *8*: ra6.

38. G. Bezakova, M. A. Ruegg. New insights into the roles of agrin. *Nat Rev Mol Cell Biol*. **2003**, *4*: 295-308.

39. T. Arai, A. Parker, W. Busby, Jr., D. R. Clemmons. Heparin, heparan sulfate, and dermatan sulfate regulate formation of the insulin-like growth factor-I and insulin-like growth factor-binding protein complexes. *J Biol Chem*. **1994**, *269*: 20388-20393.

40. C. J. Rogers *et al.* Chondroitin sulfate E influences retinotopic mapping via EphB3. *in preparation*. **2017**.

41. G. J. Miller *et al.* Chondroitin sulfate E mediates axonal inhibition via EphA4 activation. *in preparation*. **2017**.

42. M. Teran, M. A. Nugent. Synergistic binding of vascular endothelial growth factor-A and its receptors to heparin selectively modulates complex affinity. *J Biol Chem*. **2015**, *290*: 16451-16462.

43. E. Forsberg, L. Kjellen. Heparan sulfate: lessons from knockout mice. *J Clin Invest*. **2001**, *108*: 175-180.

44. M. Klüppel, T. N. Wight, C. Chan, A. Hinek, J. L. Wrana. Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. *Development*. **2005**, *132*: 3989-4003.

45. S. Mizumoto, S. Yamada, K. Sugahara. Human genetic disorders and knockout mice deficient in glycosaminoglycan. *Biomed Res Int*. **2014**, *2014*: 495764.

46. L. Koziel, M. Kunath, O. G. Kelly, A. Vortkamp. Ext1-dependent heparan sulfate regulates the range of Ihh signaling during endochondral ossification. *Developmental Cell*. **2004**, *6*: 801-813.

47. E. J. Bradbury *et al.* Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature.* **2002**, *416*: 636-640.

48. R. R. Zhao *et al.* Lentiviral vectors express chondroitinase ABC in cortical projections and promote sprouting of injured corticospinal axons. *J Neurosci Methods.* **2011**, *201*: 228-238.

49. G. García-Alías, S. Barkhuysen, M. Buckle, J. W. Fawcett. Chondroitinase ABC treatment opens a window of opportunity for task-specific rehabilitation. *Nat Neurosci.* **2009**, *12*: 1145-1151.

50. L. L. Jones, R. U. Margolis, M. H. Tuszynski. The chondroitin sulfate proteoglycans neurocan, brevican, phosphacan, and versican are differentially regulated following spinal cord injury. *Exp Neurol.* **2003**, *182*: 399-411.

51. D. Rabuka, M. B. Forstner, J. T. Groves, C. R. Bertozzi. Noncovalent cell surface engineering: incorporation of bioactive synthetic glycopolymers into cellular membranes. *J Am Chem Soc.* **2008**, *130*: 5947-5953.

52. B. Belardi, G. P. O'Donoghue, A. W. Smith, J. T. Groves, C. R. Bertozzi. Investigating cell surface galectin-Mediated cross-linking on glycoengineered cells. *J Am Chem Soc.* **2012**, *134*: 9549-9552.

53. A. Pulsipher, M. E. Griffin, S. E. Stone, J. M. Brown, L. C. Hsieh-Wilson. Directing neuronal signaling through cell-surface glycan engineering. *J Am Chem Soc.* **2014**, *136*: 6794-6797.

54. A. Pulsipher, M. E. Griffin, S. E. Stone, L. C. Hsieh-Wilson. Long-lived engineering of glycans to direct stem cell fate. *Angew Chem Int Ed.* **2015**, *54*: 1466-1470.

55. J. E. Hudak, S. M. Canham, C. R. Bertozzi. Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat Chem Biol.* **2014**, *10*: 69-75.

56. M. L. Huang, R. A. A. Smith, G. W. Trieger, K. Godula. Glycocalyx remodeling with proteoglycan mimetics promotes neural specification in embryonic stem cells. *J Am Chem Soc.* **2014**, *136*: 10565-10568.

*Chapter 2*

# Methods for Short- and Long-Term Glycan Engineering at the Cell Surface to Control Biological Outcomes

Portions of this chapter are published as:

Pulsipher A, **Griffin ME**, Stone SE, Brown JM, Hsieh-Wilson LC. "Directed neuronal signaling through cell-surface glycan engineering." *J. Am. Chem. Soc.* **2014**, *136*: 6794-6797. doi: 10.1021/ja5005174. Research article.

and

Pulsipher A*, **Griffin ME***, Stone SE, Hsieh-Wilson LC. "Long-lived glycan engineering to direct stem cell fate." *Angew. Chem. Int. Ed.* **2015**, *54*: 1466-1470. doi: 10.1002/anie.201409258; * denotes equal contribution. Research article.

## 2.1 Abstract

Cell-surface GAGs are mediators of a variety of critical signaling events and can dictate the biological activity of cell populations through their interactions with cell-surface receptors and soluble ligands. Many approaches to connect structural determinants of the cellular glycocalyx with biological function employ reductive approaches by reducing or eliminating glycan structures from the extracellular matrix. However, a forward approach in which specific glycans are anchored onto plasma membranes would allow more control over the exact carbohydrate structures being displayed at the cell surface. Moreover, this tactic would provide direct evidence for biological activity as a result of a carbohydrate structure. Here, two methods are presented to modify cell surfaces with defined carbohydrate structures. The first method employs the functionalization of CS polysaccharides with an aminooxy group and conjugation to ketone-displaying liposomes to directly fuse with the cell membrane. The second approach describes the transgenic expression of a transmembrane HaloTag construct, which can covalently bond with chlorohexyl-functionalized HS polysaccharides. These two complementary methods provide both short- and long-term display of known carbohydrates, which allow for the direct control of cellular fate in both neuronal growth and stem cell differentiation.

## 2.2 Liposomal Delivery of CS GAGs to Control Neuronal Growth

### 2.2.1 Approach and Synthesis

Although some strategies for engineering cell-surface glycans have been reported,[1-8] the remodeling of cell membranes with complex polysaccharides such as GAGs has not been demonstrated. Moreover, the application of these methodologies prior to our work has been largely limited to imaging[2-5] or studying cell-surface phenomena, such as receptor clustering.[6] Here we developed a method to display specific sulfated GAG structures on cell surfaces using a liposomal fusion strategy. Tailoring membranes with the CS-E sulfation epitope activated growth factor-mediated signaling pathways and enabled the fine-tuned modulation of neuronal growth. These findings demonstrate that chemically controlling the presentation of exogenous glycans on cell surfaces can induce sustained effects on cellular signaling and function. Our studies also highlight the potential for glycan engineering to modulate complex cellular events, and they provide a powerful, new tool for remodeling cell membranes with a wide variety of important biomolecules.

We chose to utilize liposomes as glycan carriers due to their biocompatibility, ease of preparation, low cytotoxicity, and tunable biophysical properties.[9] Elegant studies have used glycan-presenting vesicles for intracellular antigen delivery[10, 11] or sugar-encapsulated vesicles with folate receptors for cell-specific metabolic labeling.[4] However, only two liposomal methods have been developed to our

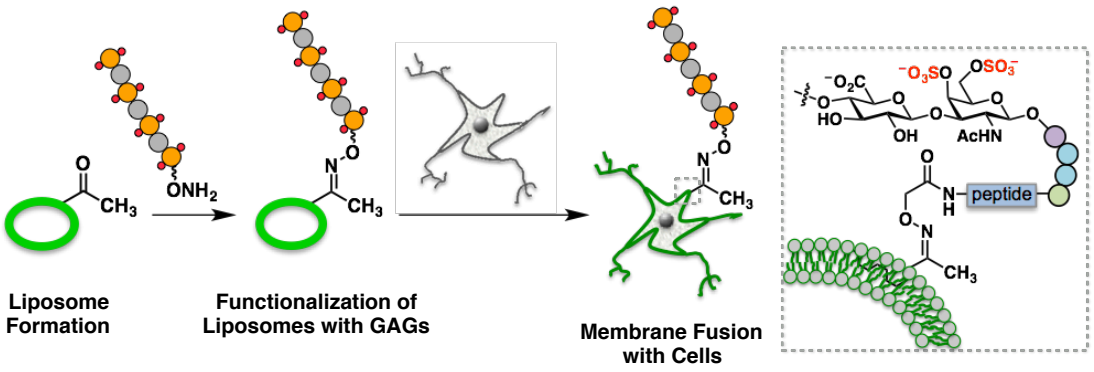



**Figure 2-1. Schematic of liposomal delivery method for glycan engineering**. Liposomes containing a ketone functional handle are produced and then functionalized with GAGs. Liposomes are then added to cells, where they fuse with the cell membrane to display carbohydrates on the cell surface.

knowledge for the cell-surface display of exogenous molecules, namely fluorophores.[12, 13] We expanded on these methods in an effort to incorporate large, sulfated GAGs into cell membranes (Figure 2-1). To promote membrane fusion and surface presentation of the glycans rather than intracellular uptake, we used cationic 1,2-dioleoyl-3-trimethylammoniumpropane chloride (DOTAP) and neutral 1,2-dioleoyl-*sn*-glycero-3-phosphatidylethanolamine (DOPE) as our primary lipids. Phosphatidylethanolamine lipids are an abundant, natural component of the neuronal cell membrane, and DOPE-based liposomes containing a lipid-functionalized fluorophore have been used to label a variety of cell types, including primary neurons.[12] We also incorporated 2-dodecanone into the liposomes to add a ketone handle for appending the glycans via oxime chemistry. CS polysaccharides containing a peptide fragment with an *N*-terminal amine were readily derivatized with an aminooxy group by coupling them to

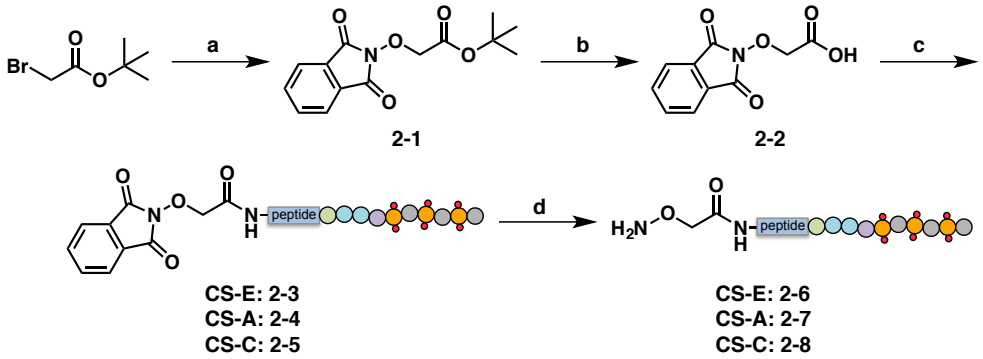



**Figure 2-2. Synthesis of aminooxy-functionalized CS GAGs (2-6, 2-7, and 2-8)**.(a) *N*-hydroxyphthalimide (1.2 eq), $NaHCO_3$ (1.2 eq), DMF, 60 °C, 2 h, 88%; (b) TFA (3 eq), DCM, 1 h, 99%; (c) EDC (1.1 eq), sulfo-NHS (1.1 eq), 100 mM MES pH 6.0, 300 mM NaCl, RT, 30 min, then polysaccharide-$NH_2$, 50 mM $NaHCO_3$, RT, 12, 60%. (d) hydrazine monohydrate (3 eq), $ddH_2O$, RT, 3 h, 92%.

phthaloyl-protected aminooxyacetic acid followed by cleavage of the phthaloyl group with hydrazine monohydrate (Figure 2-2). The polysaccharides were then incubated with dodecanone-containing liposomes at 25 °C for 3 h to produce GAG-displaying vesicles. The synthetic ease and versatility of this approach represent advantages compared to existing methods, which require the radical-mediated synthesis of polymers end-functionalized with lipids.[5-8] With our approach, the lipid reagents are commercially available, and many biomolecules can be derivatized with aminooxy groups, including various glycans, peptides, lipids, nucleic acids, and proteins,[14-16] thus providing a general strategy for displaying a diverse range of bioactive molecules.

### 2.2.2 Validation

Preliminary optimization of liposomal membrane fusion was performed on rat pheochromocytoma (PC12) cells using liposomes functionalized with a

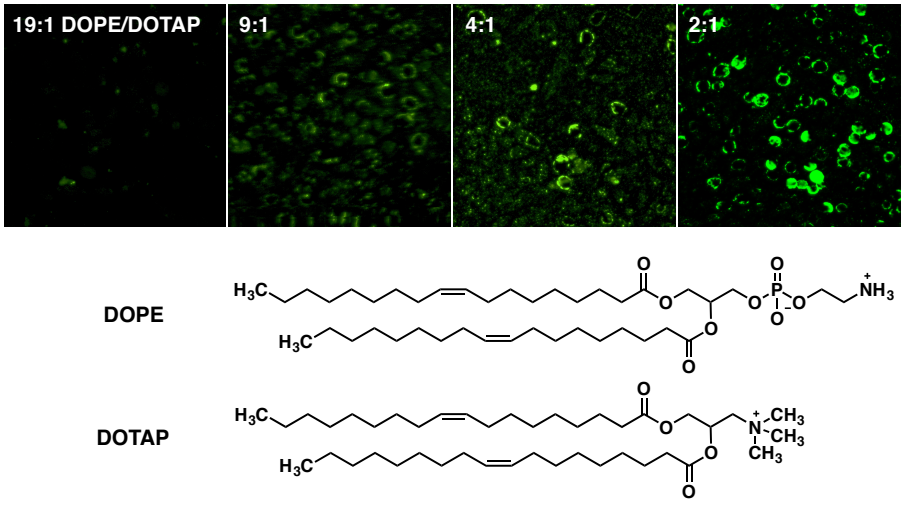



**Figure 2-3. Optimization of lipid composition of liposomes for membrane fusion.** Liposomes were produced with different w/w ratios of DOPE and DOTAP and 10% w/w dodecanone. AF488-hyd was incubated with the liposomes (dd$H_2O$, RT, 3h) to spontaneously react with the presented ketone moiety. PC12 cells were then incubated with the fluorescent liposomes (37 °C, 30 min) and imaged to visualize fusion efficiency.

hydrazide-conjugated fluorophore (AF488-hyd). We found that a 2:1 w/w ratio of DOPE:DOTAP was optimal for membrane fusion, as visualized by fluorescence microscopy (Figure 2-3). To approximate the relative levels of fluorophore incorporation at the cell surface, we incubated liposomes containing varying concentrations of AF488-hyd with PC12 cells on ice for 30 min. Cells labeled with liposomes containing 10 mol % AF488-hyd displayed similar fluorescence signal profiles by fluorescence-assisted cell sorting (FACS) analysis as cells labeled with an anti-CS-E monoclonal antibody[17] that detected endogenous CS-E levels (Figure 2-4). These results suggest that this liposomal strategy can incorporate exogenous molecules into cell membranes at levels roughly similar to those of endogenous CS polysaccharides. We next examined whether this approach could be used to

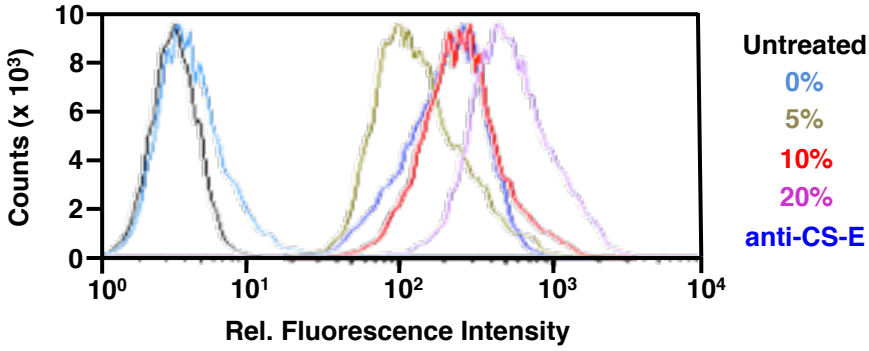
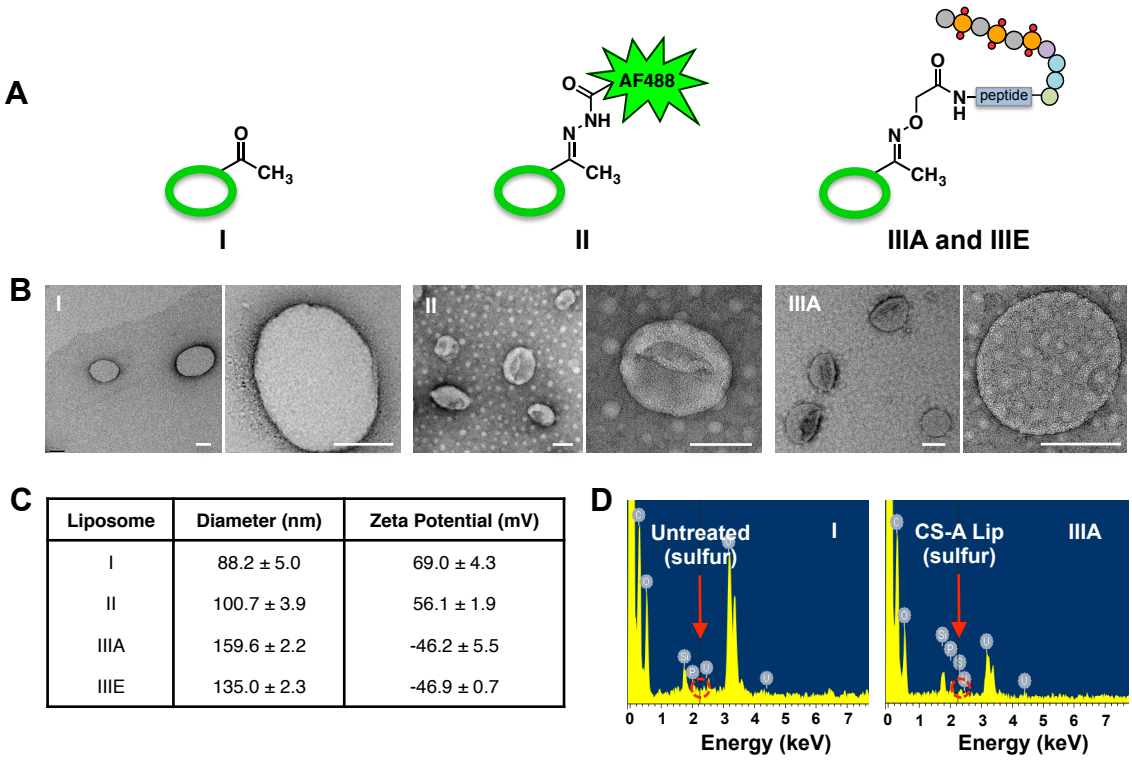



**Figure 2-4. Quantifying cell surface labeling by liposomal delivery.** PC12 cells were functionalized with liposomes containing different amounts of 2-dodecanone (w/w) that had been reacted with AF-488 and analyzed by flow cytometry. As an approximation, these data were compared to PC12 cells labeled with a CS-E monoclonal antibody.




| Liposome | Diameter (nm) | Zeta Potential (mV) |
|---|---|---|
| I | 88.2 ± 5.0 | 69.0 ± 4.3 |
| II | 100.7 ± 3.9 | 56.1 ± 1.9 |
| IIIA | 159.6 ± 2.2 | -46.2 ± 5.5 |
| IIIE | 135.0 ± 2.3 | -46.9 ± 0.7 |

**Figure 2-5. Biophysical characterization of conjugated liposomes.** (A) Schematic of different liposomes used for characterization. (B) Liposomes were negatively stained with uranyl acetate and imaged using TEM (scale bar = 50 nm). (c) DLS was used to measure average liposome hydrodynamic diameters, which showed increases after functionalization. Zeta potential analysis showed a negative shift in potential after functionalization with sulfated GAGs. (d) Energy dispersion spectroscopy showed the incorporation of sulfur after functionalization with sulfated GAGs.

display large GAG polysaccharides on cell surfaces. Liposomes containing 2:1 w/w DOPE:DOTAP and 10% w/w dodecanone were functionalized with CS-E-enriched polysaccharides (~70 kDa). To characterize their biophysical properties, we used transmission electron microscopy (TEM), dynamic light scattering (DLS), and zeta potential measurements. The liposomes exhibited parameters predicted to favor membrane fusion, including spherical morphologies, average diameters ranging between 132.6 and 159.6 nm, and good stabilities (zeta potentials of ±46–69 mV; Figure 2-5b). The change from positive to negative electrokinetic potential (69 to −46 mV; Figure 2-5c) confirmed successful conjugation of the sulfated polysaccharides. Furthermore, energy dispersive spectroscopy (EDS) verified the presence of sulfur on CS-functionalized liposomes after CS conjugation (Figure 2-5d). To test for membrane fusion, PC12 cells were treated with chondroitinase to remove endogenous CS, incubated with CS-E-modified liposomes, and immunostained with an anti-CS-E antibody. Importantly, strong immunostaining for CS-E was observed on the surfaces of cells treated with CS-E-functionalized liposomes compared to chondroitinase-treated cells without liposome addition, indicating efficient incorporation of the polysaccharides (Figure 2-6).

### 2.2.3 Controlling Intracellular Signaling

Having validated the method, we investigated whether the approach could be used to control cellular signaling pathways. Previous studies from our laboratory

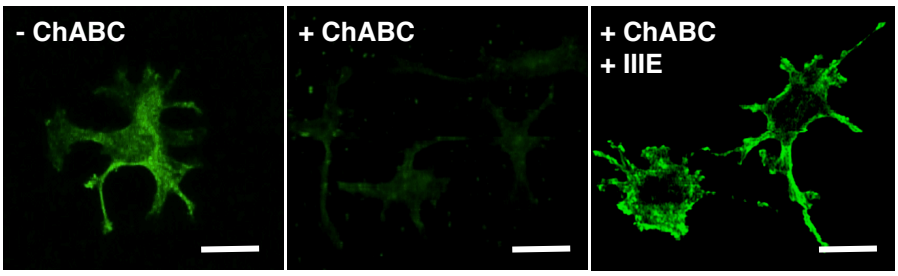



**Figure 2-6. Validation of cell functionalization with CS-E via liposomal delivery.** PC12 cells were left untreated, treated with chondroitinase ABC (ChABC), or treated with ChABC and then CS-E displaying liposomes (**IIIE**) and then immunostained with a CS-E monoclonal antibody.

have demonstrated that CS-E polysaccharides can recruit nerve growth factor (NGF) to the cell surface and promote the assembly of NGF-tropomyosin receptor kinase (Trk) complexes.[18] Complex formation, in turn, activates protein kinase B (Akt) and phosphatidylinositol 3-kinase (PI3K) signaling pathways and thereby enhances neurite outgrowth.[19] Thus, we reasoned that cell-surface presentation of exogenous CS-E polysaccharides might recruit NGF to the membrane and assist in the formation of activated signaling complexes to induce these biological responses. Embryonic day 18 (E18) rat cortical neurons were cultured *in vitro* for 7 days and then treated with liposomes displaying CS-E- or CS-C-enriched polysaccharides (**IIIE** or **IIIC**, 30 min, 37 °C). Neurons were stimulated with the neurotrophin NGF for 0, 10, 30, or 60 min, and Akt activation was monitored using a phospho-Ser473 Akt antibody. Remarkably, neurons remodeled with CS-E polysaccharides showed approximately a 3-fold increase in Akt activation relative to untreated neurons at each time point (Figure 2-7), consistent with increased recruitment of NGF to the cell surface and robust

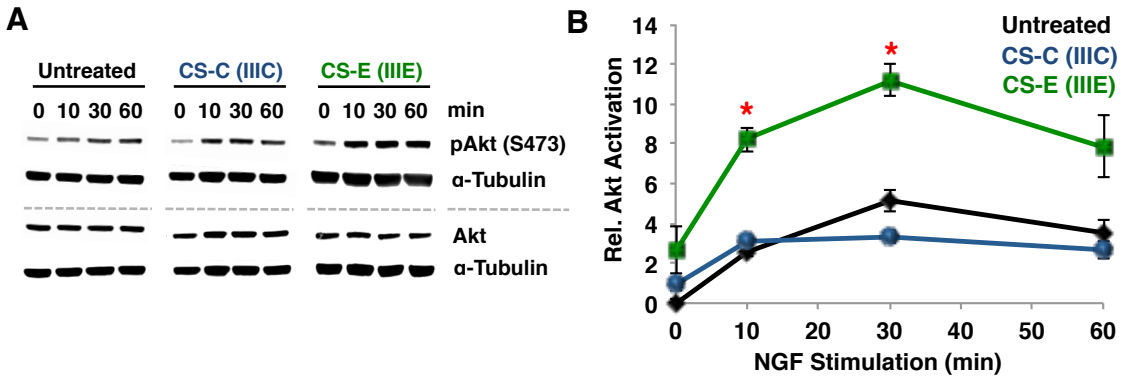



**Figure 2-7. Potentiation of Akt signaling via cell surface glycan engineering.** Embryonic cortical neurons were left untreated or functionalized with CS-C- (**IIIC**) or CS-E- (**IIIE**) displaying liposomes. Cells were then stimulated with NGF, and Akt activation was monitored by (A) Western blotting with (B) quantification. * $P < 0.05$.

activation of Trk receptors. In contrast, neurons remodeled with CS-C polysaccharides showed phospho-Akt levels similar to those of untreated neurons. These results support the importance of the CS-E motif in NGF-stimulated Trk activation.[18] Moreover, they show that liposomal-mediated presentation of specific, sulfated CS polysaccharides on cell surfaces can activate important neuronal signaling pathways.

### 2.2.4 Controlling Neuronal Growth

We next examined whether the exogenous CS-E-mediated activation of Trk pathways could direct neuronal growth. Rat E18 hippocampal neurons were cultured *in vitro* for 2 days and subsequently treated for 30 min with liposomes prefunctionalized with CS-A-, CS-C-, or CS-E-enriched polysaccharides. Neurons were then incubated in media lacking liposomes for an additional 24 h and immunostained with an anti-α-tubulin antibody to image the processes using

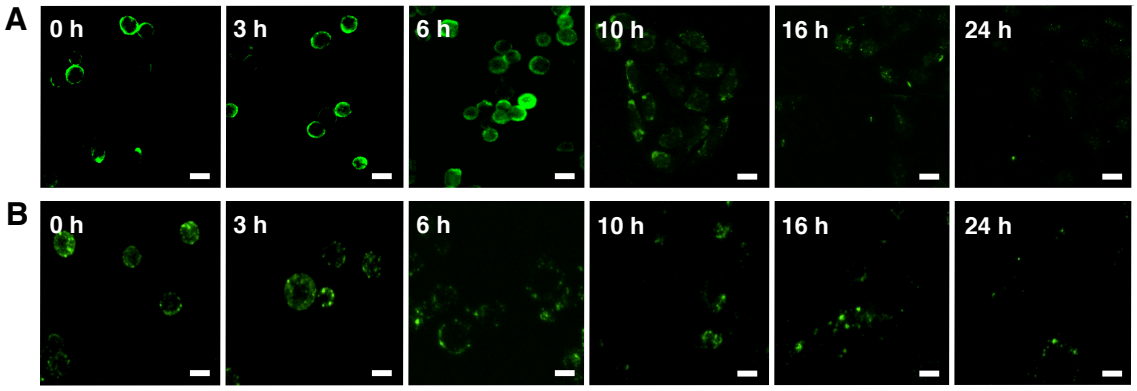



**Figure 2-8. Surface lifetime of lipid-anchored molecules.** PC12 cells were functionalized with (A) AF488- (**II**) or (B) fluorescent CS-E- (**IIIE**) functionalized liposomes and monitored by microscopy. The membrane lifetime of both molecules lasts only a few hours, with the signal from the small molecule dye being slightly longer lived. Scale bar = 50 μm for (A) and 30 μm for (B).

confocal fluorescence microscopy. Initial studies revealed no appreciable difference in neurite outgrowth between neurons displaying the different sulfation motifs. We postulated that the membrane lifetime of the exogenous CS GAGs might be too short to elicit functional responses such as neurite outgrowth, which requires *de novo* protein and lipid biosynthesis. Therefore, we assayed the membrane lifetime of the exogenous lipids by treating PC12 cells with liposomes bearing fluorophore-conjugated CS-E polysaccharides and monitoring the fluorescence signal over 24 h (Figure 2-8). A decrease in signal was observed within 6 h, and loss of the signal progressed over the course of 10 h. After 16 h, weak fluorescence was detected around the cell periphery, suggesting that the CS-conjugated lipids had been internalized or had diffused into the medium. To circumvent this problem, we repeated the outgrowth assays with multiple additions of liposomes every 8 h over a 24-h period. Under these conditions, we

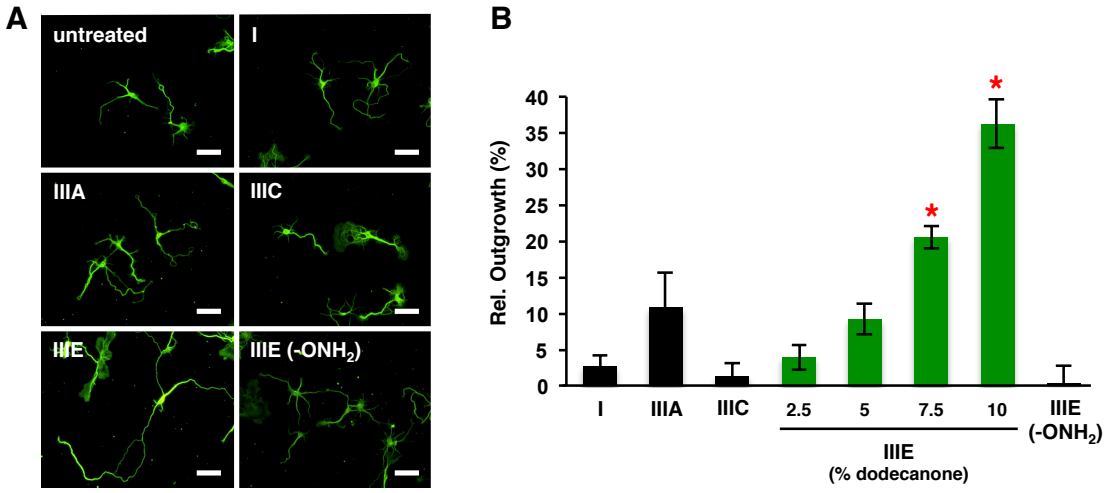



**Figure 2-9. Controlling neurite outgrowth via cell surface glycan engineering.** Embryonic hippocampal neurons were left untreated or treated with CS-A- (**IIIA**), CS-C- (**IIIC**), or CS-E- (**IIIE**) functionalized liposomes once every 8 hours for 24 h. Samples were (A) visualized by immunofluorescence and (B) quantified. For all samples, liposomes were produced with 10% w/w 2-dodecanone except for the panel of **IIIE** liposomes. Neurite outgrowth could be finely tuned by the amount of CS-E added onto the surface. Treatment with liposomes that had been incubated with CS-E without the aminooxy group showed no difference in outgrowth compared to control samples. Scale bar = 50 μm. * $P < 0.05$.

found that cell-surface presentation of CS-E polysaccharides significantly enhanced neurite outgrowth by 36.3 ± 3.3% relative to untreated neurons (Figure 2-9). As expected, neurons displaying CS-A or CS-C polysaccharides showed minimal neurite outgrowth when compared to untreated neurons (10.8 ± 4.8% and 1.3 ± 1.8%, respectively). As a further control, dodecanone-containing liposomes were reacted with CS-E polysaccharides lacking the aminooxy functionality and then incubated with the cells. No difference in neurite outgrowth was observed relative to untreated neurons. Remarkably, the extent of neurite outgrowth could be finely tuned by controlling the concentration of CS-E

polysaccharides at the cell surface. Liposomes containing 0% to 10% dodecanone were conjugated with aminooxy-functionalized CS-E and then incubated with E18 rat hippocampal neurons as above. Notably, we observed a dose-dependent increase in neurite outgrowth from 4.0% to 36.3% as the dodecanone concentration was increased from 2.5% to 10% (Figure 2-9). Together, these studies demonstrate that this approach for engineering glycans on cell surfaces can be used to finely modulate both the signaling and functional responses of neurons.

## 2.3 HaloTag Anchoring of HS GAGs to Control Stem Cell Differentiation

### 2.3.1 Approach and Synthesis

Heparan sulfate (HS) glycosaminoglycans (GAGs) have recently been identified as important regulators of stem cell differentiation.[20-24] HS GAGs exert this control by selectively interacting with proteins involved in the differentiation process, such as fibroblast growth factors (FGFs), bone morphogenic proteins (BMPs), and wingless-type MMTV integration site family members (Wnts).[22, 25] Furthermore, specific sulfation patterns of HS have been implicated in the progression of ESCs from self-renewal to a differentiated state. For example, undersulfated HS is found on pluripotent cells,[23] whereas highly sulfated HS is associated with differentiated cells and has been proposed to promote interactions between soluble FGF and BMP factors and their receptors.[22] However, the precise sulfated epitopes and mechanisms involved in the generation of specific cell lineages remain unclear. We postulated that the presentation of particular HS GAG structures on ESC surfaces might enable the selective activation of signaling pathways and thereby induce desirable cell fates. Such an approach would also provide novel insights into the structure-function relationships of HS GAGs and their roles in stem cell biology.

Elegant studies have recently shown that the short-term display of synthetic HS glycopolymers can promote stem cell specification to form intermediate neural rosettes.[26] However, directing the generation of fully differentiated,

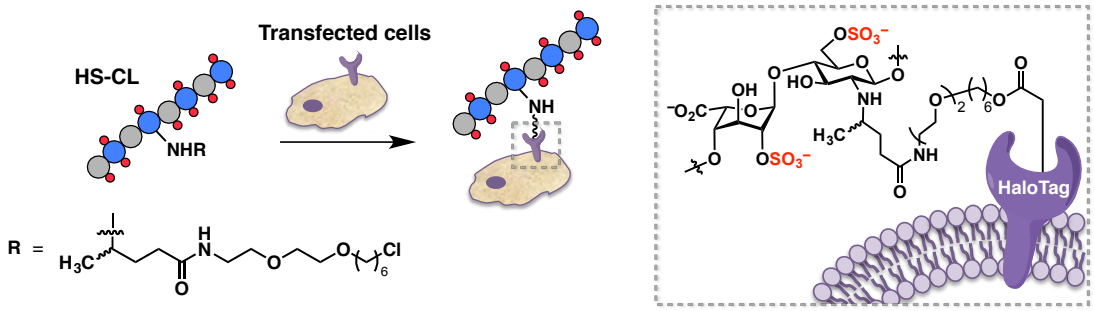



**Figure 2-10. Schematic of HTP glycan engineering approach.** HS GAGs modified with the CL linker are added directly to cells expressing the transmembrane HTP construct, which covalently bonds to the CL linker.

mature cell types will likely require the development of new methods to enable the long-term, stable presentation of defined HS GAGs. Although our liposomal method as well as the passive lipid insertion technique can elicit short-term cellular responses,[8, 27] the lipid tail anchor limits the membrane lifetime of the exogenous glycans to several hours. To address these shortcomings, we developed a method to tailor cell surfaces with specific HS derivatives using membrane-bound HaloTag proteins (HTPs) as anchors (Figure 2-10). Molecules covalently attached to HTPs displayed prolonged cell-surface lifetimes of more than one week, circumventing the temporal limitation of lipid anchors. Moreover, mouse ESCs remodeled with heparin/trisulfated HS underwent accelerated exit from selfrenewal and commitment to a neural lineage through early activation of extracellular signal-regulated kinase (ERK)/mitogen-activated protein kinase (MAPK) signaling pathways. These results highlight the potential to elucidate the functional roles of HS GAGs and direct cell differentiation by remodeling the glycocalyx of stem cells.

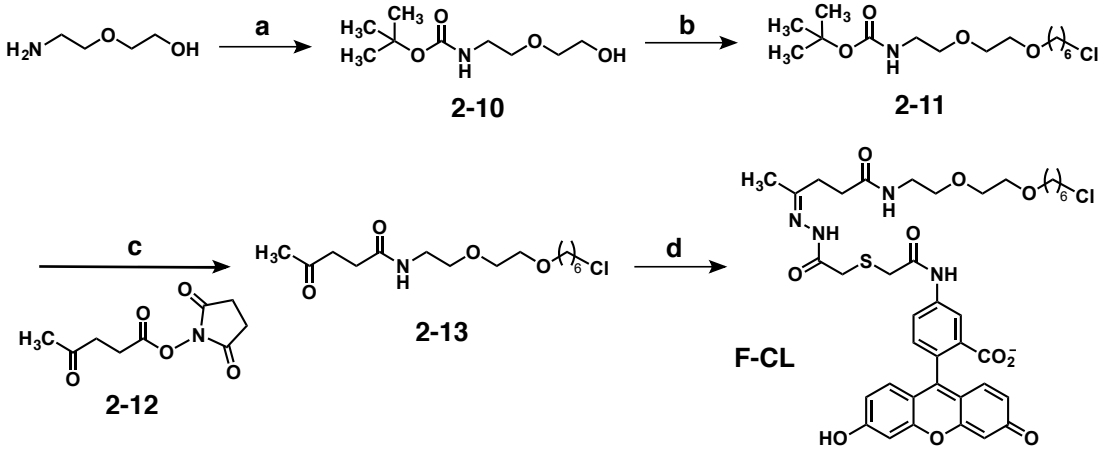



**Figure 2-11. Synthesis of 2-13 and F-CL.** (a) $Boc_2O$ (2 eq), MeOH, 0 °C, 2 h, 91%; b) 60% NaH (1.5 eq), DMF, 0 °C, 1 h, then 1-chloro-6-iodohexane (1.2 eq), 24 h, 37%; c) TFA (6 eq), DCM, 4 h, then **2-12** (1 eq), DCM, 24 h, 67%; d) 5-(((2-(carbohydrazino)methyl) thio)acetyl)-aminofluorescein (**F**) (0.5 eq), MeOH, 16 h, 75%.

HTP is a modified alkane dehalogenase that forms a covalent adduct with chloroalkane substrates.[28] Strategies based on HTP have been adapted for diverse applications, ranging from cancer diagnostics to chemical proteomics.[29-31] However, most reported applications have used HTP methods to append molecules that serve as detection or capture agents. We chose to exploit the HTP platform to modulate biological processes in living cells. We first investigated the membrane lifetime of molecules conjugated to HTPs. *N*-*tert*-Butyloxycarbonyl (Boc)-protected 1-(2-(2-amino-ethoxy)ethoxy)-6-chlorohexane **2-11**[31] was deprotected with trifluoroacetic acid and reacted with *N*-hydroxysuccinimidyl levulinate **2-12** to obtain chloroalkane linker (CL) **2-13** (Figure 2-11). Condensation of **2-13** with a fluorescein-hydrazide derivative gave CL-conjugated fluorescein (**F-CL**).

### 2.3.2 Validation

To test the approach, hemagglutinin (HA)-tagged HTP was stably expressed in Chinese hamster ovary (CHO) cell membranes by fusing it to the platelet-derived growth factor receptor (PDGFR) transmembrane domain. Cells were then incubated with F-CL for one hour at 37 °C, and individual wells were fixed and imaged every twelve hours for eight days. Remarkably, we observed a strong fluorescence signal that persisted for at least eight days after only a single F-CL treatment (Figure 2-12). In contrast, no fluorescence signal was observed when cells were treated with the fluorescein-hydrazide derivative alone or with cells lacking HTP. These results indicate that the display of HTP conjugates is specific and long-lived despite membrane turnover, highlighting the potential of this approach to exert long-lasting effects on cellular function. We next examined whether this method could be used to present HS GAGs on cell surfaces. Although the glucosamine sugars in heparin/HS are mostly *N*-acetylated or *N*-sulfated, the free amine is also present in low abundance (1–3%), which provides a convenient functional handle for attaching the chloroalkane linker in a single step. We biotinylated HS (B-HS) and conjugated it to **1** by reductive amination chemistry (**B-HS-CL**; Figure 2-13). CHO cells stably expressing HTP were incubated with **B-HS** or **B-HS-CL** at 37 °C for six hours. Cells were lysed and subjected to blotting analysis using an anti-HA antibody and streptavidin IRDye800. Importantly, biotinylated HS was detected only from HTP-expressing

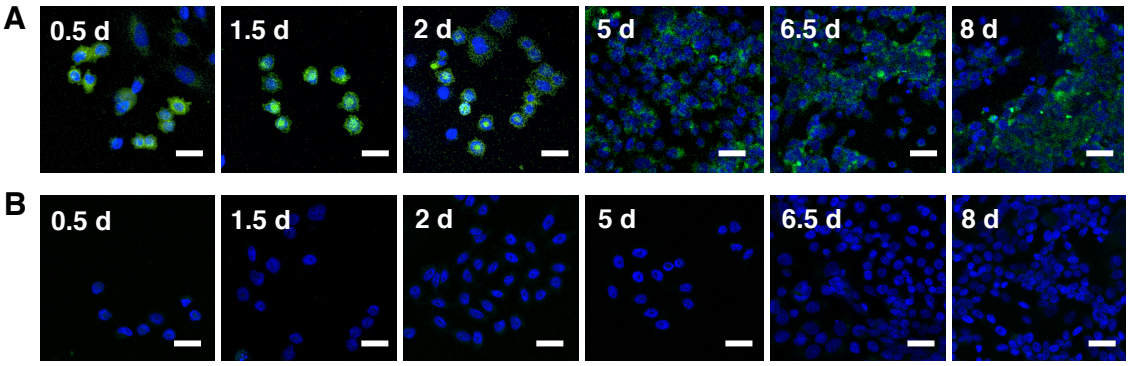



**Figure 2-12. Surface lifetime of HTP-anchored molecules.** CHO cells expressing HTP were treated with (A) **F-CL** or (B) **F** and monitored by microscopy for over one week. Surface labeling was observed at 8 days with no nonspecific labeling from unlabeled fluorescein (**F**). Scale bar = 50 μm.

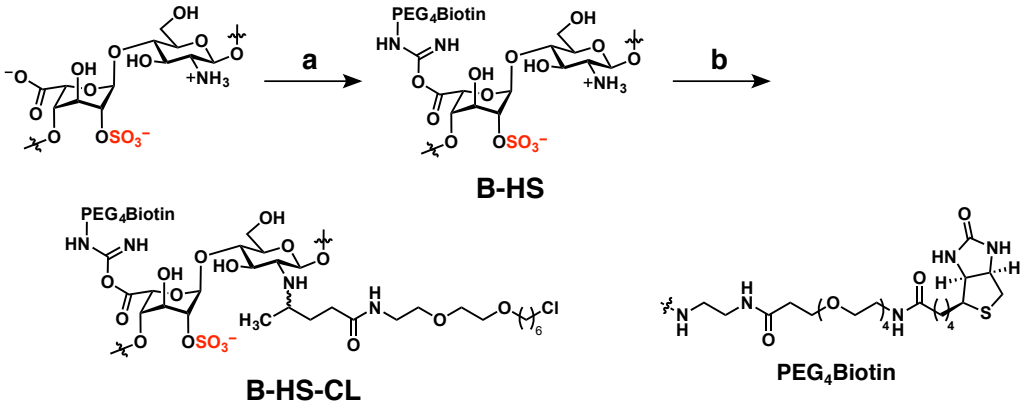



**Figure 2-13. Synthesis of B-HS and B-HS-CL**. (a) BrCN (excess), 0.2 M NaOH pH 11, RT, 10 min, then Biotin$PEG_4NH_4$ (excess), 0.2 M sodium borate pH 8, RT, 12 h; (b) **2-13** (excess), $NaBH_3CN$ (excess), MeOH, $ddH_2O$, RT, 16 h.

cells treated with **B-HS-CL**, but not **B-HS**, which lacks the chloroalkane moiety (Figure 2-14). Immunocytochemical analysis of cells further confirmed the presence of biotinylated HS on HTP-expressing cells incubated with **B-HS-CL**, but not **B-HS**. Together, these studies demonstrate the selective display of HS GAGs on cell surfaces using HTP anchors.

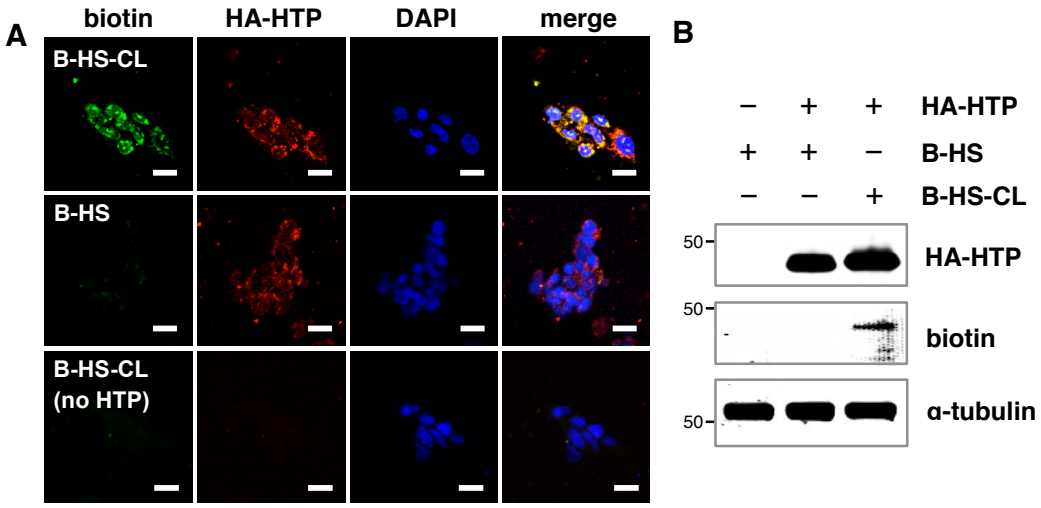
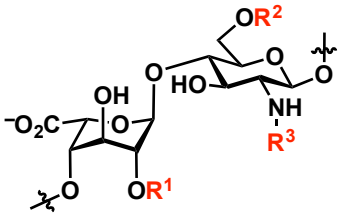



**Figure 2-14. Glycan engineering with HS GAGs**. CHO cells expressing HTP or not were incubated with **B-HS-CL** or **B-HS**. Incorporation of HS onto the cell surface was validated by (A) immunofluorescence or (B) Western blotting. Scale bar = 50 μm.

### 2.3.3 Deconvoluting the Role of HS GAG Sulfation Epitopes in FGF Activation

With a method for the long-lived presentation of GAGs in hand, we examined the ability of our non-natural presentation of HS GAGs to stimulate stem cell signaling pathways. HS GAGs regulate FGF-FGF receptor (FGFR) signaling events involved in stem cell differentiation by assisting in the assembly of active growth factor/receptor complexes.[32, 33] Activation of FGF2-FGFR1, in turn, initiates several intracellular signaling pathways, including the ERK/MAPK cascade.[34] A distinct period of increased ERK1/2 phosphorylation through FGF signaling triggers ESCs to exit self-renewal and the transition to a neural cell fate.[35, 36] To display particular sulfated HS structures on ESCs and stimulate the FGF2-FGFR1 pathway, we conjugated **2-13** to various heparin/HS polysaccharides. The commercially available polysaccharides (12–13 kDa) were derived from a single




**HS-CL**: $R^1 = SO_3^-$, $R^2 = SO_3^-$, $R^3 = SO_3^-$, Ac, or CL
**de-HS-CL**: $R^1$ = H, $R^2$ = H, $R^3$ = H, Ac, or CL
**deO-HS-CL**: $R^1$ = H, $R^2$ = H, $R^3 = SO_3^-$, Ac, or CL
**6-deO-HS-CL**: $R^1 = SO_3^-$, $R^2$ = H, $R^3 = SO_3^-$, Ac, or CL
**2-deO-HS-CL**: $R^1$ = H, $R^2 = SO_3^-$, $R^3 = SO_3^-$, Ac, or CL

**Figure 2-15. Chemical structures of HS GAGs**.

natural source and chemically desulfated to produce heparin/HS with defined sulfation motifs. Specifically, CL-functionalized heparin/trisulfated HS (**HS-CL**), fully desulfated heparin/HS (**de-HS-CL**), *O*-desulfated heparin/HS (**deO-HS-CL**), 6-*O*-desulfated heparin/HS (**6-deO-HS-CL**), and 2-*O*-desulfated heparin/HS (**2-deO-HS-CL**) were readily generated in one step under standard reductive amination conditions (Figure 2-15). A homogeneous population of pluripotent mouse ESCs were obtained commercially and transiently transfected with the HTP construct. The ESCs were treated with heparinase II to remove endogenous HS GAGs and incubated overnight in serum-depleted medium containing the various HS-CL derivatives. Cells were then stimulated with FGF2 and assayed for ERK1/2 activation by immunoblotting with antibodies against phosphorylated (pThr202/pTyr204) and total ERK1/2. To our delight, ESCs engineered to display heparin/trisulfated HS exhibited a 1.5-fold increase in ERK1/2 phosphorylation compared to untreated cells (Figure 2-16a). **HS-CL** had no effect on ESCs lacking HTP, confirming that ERK1/2 activation required the HTP anchors (Figure 2-16b). Moreover, ERK1/2 activation was dependent on the sulfation pattern, as

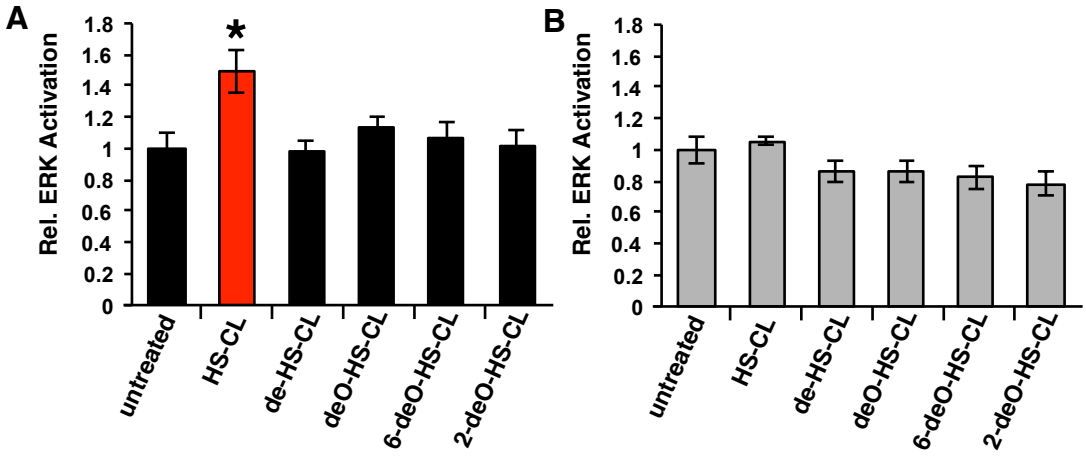



**Figure 2-16. ERK1/2 activation by cell surface engineering.** (A) mESCs were transfected with the HTP construct, functionalized with different **HS-CL** structures and then stimulated with FGF2. ERK1/2 phosphorylation was measured by Western blotting and quantified. (B) The same experiment was conducted with untransfected mESCs. * $P < 0.05$.

cells engineered with other selectively desulfated heparin/HS variants showed no significant increase in phospho-ERK1/2 levels.

Structural and biochemical studies have led to conflicting reports on the pivotal determinants of HS-FGF2-FGFR1 complexation and ERK1/2 activation.[33, 37, 38] For example, crystallographic studies suggested that 6-*O*-sulfation is critical for both ligand and receptor binding,[33] whereas certain biochemical studies indicated that downstream ERK1/2 activation elicited by the complex is not significantly attenuated by loss of the 6-*O*-sulfate group.[38] Interestingly, the same crystallographic studies showed that the other sulfate groups are also important for interactions with FGF2, suggesting that all three of the sulfate groups may be required for formation of the ternary HS-FGF2-FGFR1 complex.[33] To clarify the role of sulfation, we used GAG microarrays[18] to probe the ternary interaction

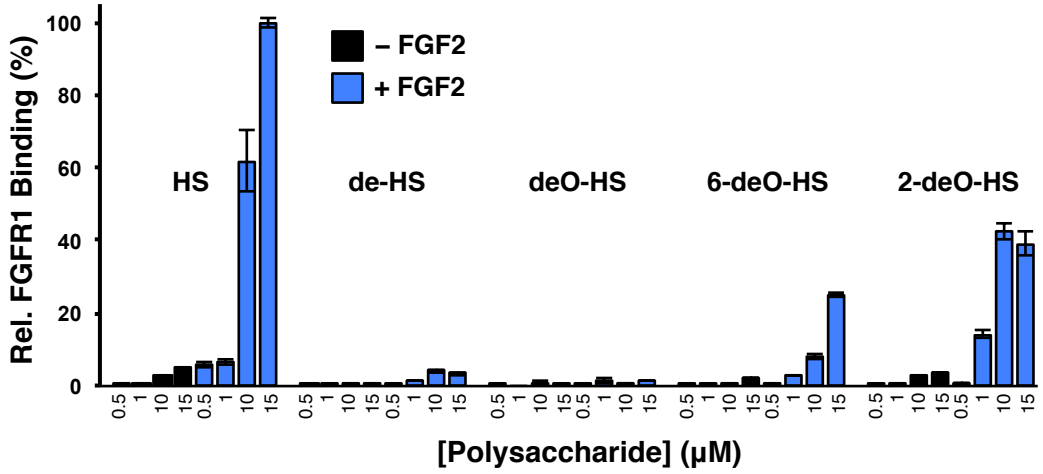



**Figure 2-17. Ternary complex formation between FGF2 and FGFR1 with HS.** Microarrays were incubated with FGFR1-Fc with and without FGF2. Bound FGFR1-Fc was visualized by an anti-human Fc antibody.

more closely. Microarrays printed with various concentrations of the HS derivatives were incubated with an FGFR1-Fc fusion protein in the presence or absence of FGF2. Binding of FGFR1-Fc was visualized using an anti-Fc antibody conjugated to AlexaFluor 647. We found that FGFR1 bound preferentially to heparin/trisulfated HS only in the presence of FGF2, suggesting the formation of a ternary complex (Figure 2-17). Consistent with our ERK1/2 activation studies, binding of FGFR1 was significantly attenuated by either 2-*O*- or 6-*O*-desulfation of HS. Taken together, our studies suggest that both 2-*O*- and 6-*O*-sulfation of HS are critical for FGF2-FGFR1-mediated ERK1/2 activation in ESCs. Thus, an HTP-based approach can be used to hijack endogenous signaling pathways and deconvolute the sulfation requirements of complex GAG-mediated processes.

### 2.3.4 Accelerating Stem Cell Differentiation

Finally, we investigated whether the HTP-dependent presentation of specific heparin/HS epitopes could promote stem cell differentiation into mature cell types. Dopaminergic neurons from fetal sources have shown long-term cell survival and preliminary clinical benefits when implanted into patients with Parkinson's disease.[39, 40] As such, the development and refinement of *in vitro* systems to generate neuronal cell populations could have widespread biomedical implications for the treatment of neurological diseases. The differentiation of ESCs into dopaminergic neurons is promoted by FGF8b.[41] Interestingly, our microarray studies indicated that FGF8b and its receptor FGFR3c are capable of forming ternary complexes with heparin/trisulfated HS, whereas complex assembly was attenuated with 2-*O*-desulfated or 6-*O*-desulfated heparin/HS and was not observed with fully *O*-desulfated heparin/HS (Figure 2-18). Pluripotent mouse ESCs transiently expressing HTP were treated with heparinase II and then grown in neural induction medium supplemented with FGF2 and **HS-CL**. For comparison, ESCs were treated identically, but in the absence of **HS-CL** (untreated) or replacing **HS-CL** with **de-HS-CL**. After two days, the FGF2-containing medium was substituted with a neural induction medium containing FGF8b. We monitored the differentiation process at specific time points by profiling the gene expression levels of specific markers using real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) analyses

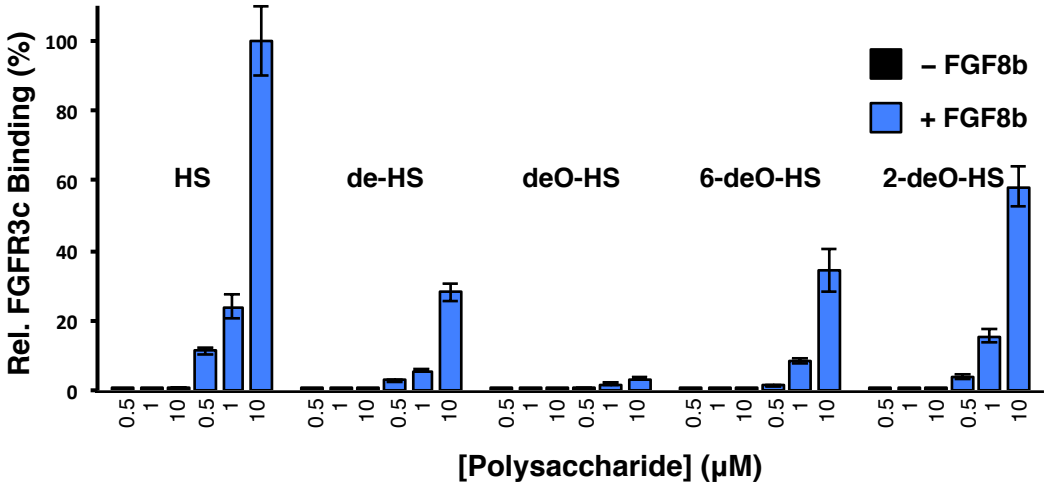



**Figure 2-18. Ternary complex formation between FGF8b and FGFR3c with HS.** Microarrays were incubated with FGFR3c-Fc with and without FGF8b. Bound FGFR3c-Fc was visualized by an anti-human Fc antibody.

(Figure 2-19). The transcription factors NANOG and SOX1 are well-established markers for pluripotency and self-renewal exit/neural lineage commitment, respectively, whereas TUJ1 is widely used as a specific marker for mature, fully differentiated neurons.[42] In parallel, the presence of NANOG, SOX1, and TUJ1 in cells was imaged by immunocytochemistry (Figure 2-20). We found that mRNA NANOG levels declined most rapidly in ESCs treated with **HS-CL**, which is consistent with an accelerated loss of pluripotency. In contrast, the levels remained high until day nine in untreated cells and cells treated with **de-HS-CL**. The decrease in NANOG levels at day nine for untreated and **de-HS-CL** treated cells is most likely due to *de novo* HS biosynthesis following the single heparinase treatment. Consistent with an accelerated loss of pluripotency, the decline in NANOG levels in **HS-CL**-treated cells was accompanied by a corresponding

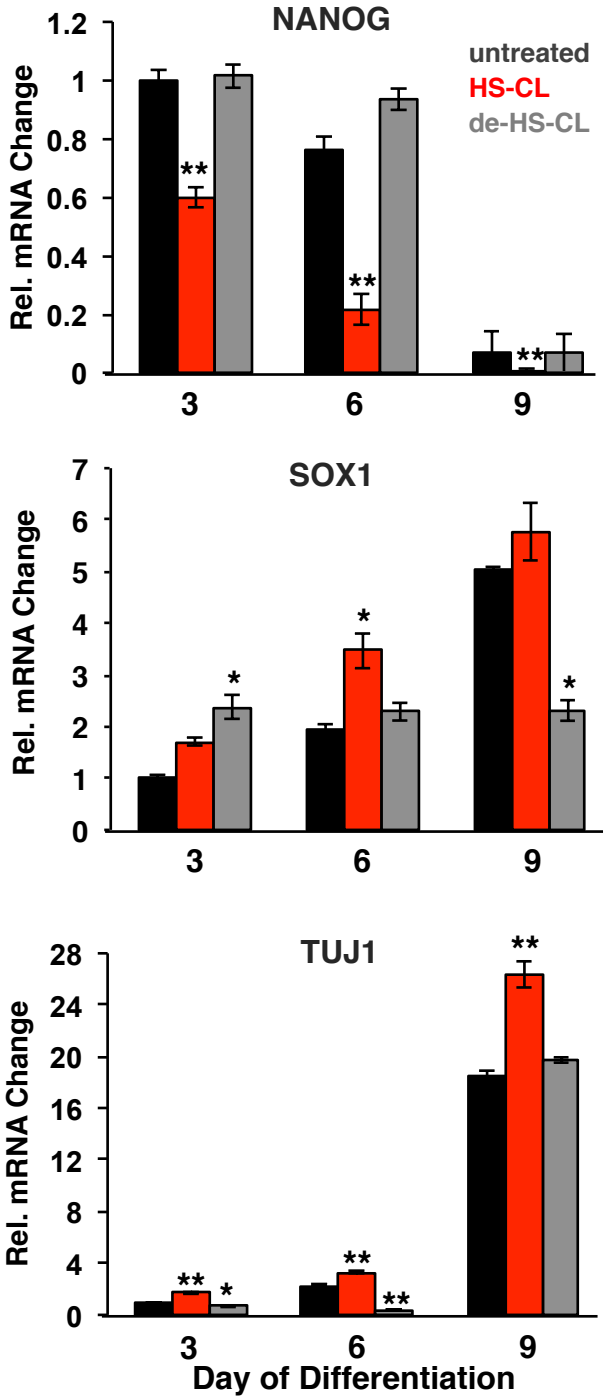



**Figure 2-19. Accelerated stem cell differentiation with glycan engineering (qRT-PCR).** mESCs expressing HTP were incubated with nothing, **HS-CL**, or **de-HS-CL** and cultured in differentiation medium as described. Differentation was monitored by qRT-PCR of NANOG, SOX1, and TUJ1. * $P < 0.05$, ** $P < 0.01$.

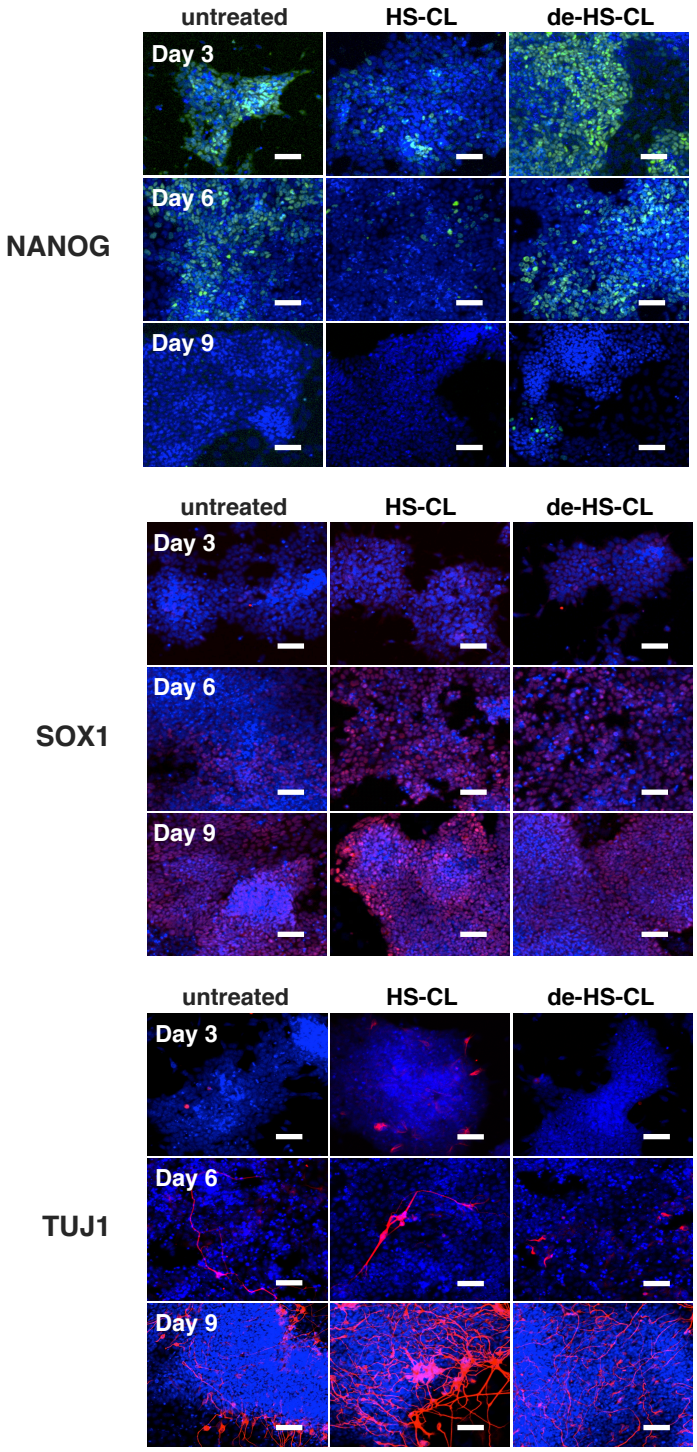



**Figure 2-20. Accelerated stem cell differentiation with glycan engineering (immunofluorescence).** mESCs from the experiment described in Figure 2-18 were stained for NANOG (green), SOX1 (red), or TUJ1 (red) and DAPI (blue). Scale bar = 50 μm.

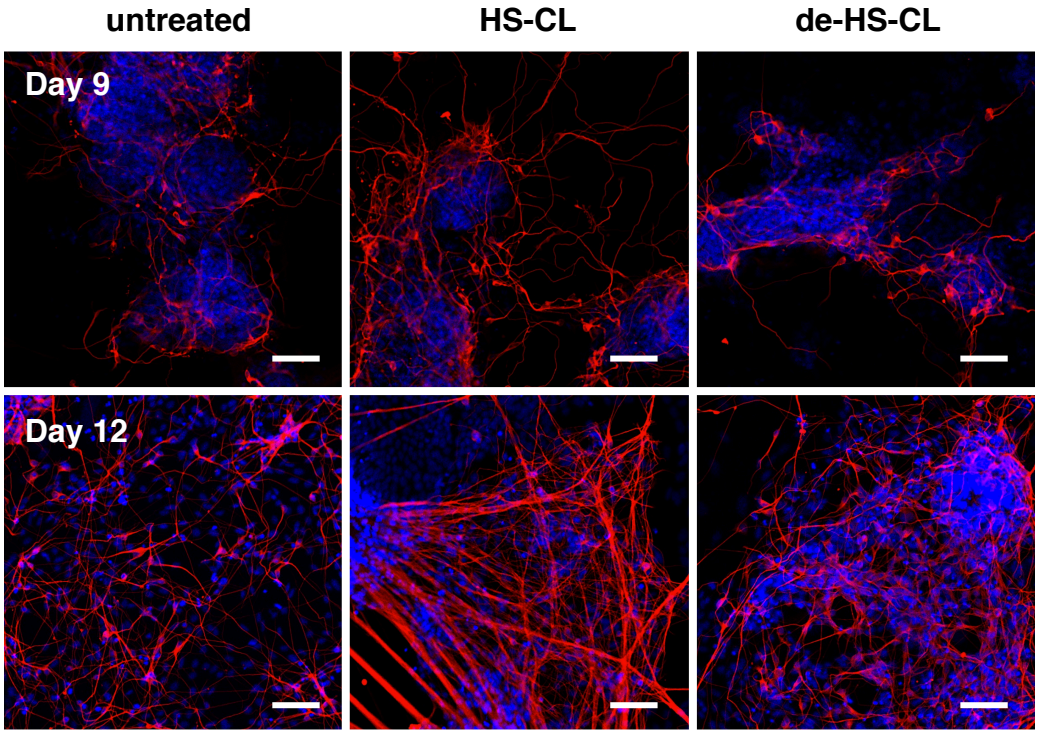



**Figure 2-21. Mature neuronal processes on differentiated mESCs.** mESCs from the experiment described in Figure 2-18 were allowed to grow to 9 or 12 days and were stained for TUJ1 (red) and DAPI (blue). Scale bar = 50 μm.

increase in the neuroectoderm-specific marker SOX1. At day six, SOX1 levels were significantly higher in **HS-CL**-treated cells than in untreated or **de-HS-CL**-treated cells, suggesting that **HS-CL**-treated cells had undergone accelerated exit from self-renewal and commitment to an intermediate, neuroectoderm state. Importantly, cells remodeled with **HS-CL** also showed significantly higher levels of TUJ1 than untreated or **de-HS-CL**-treated cells at all time points, which is indicative of faster progression to a mature, differentiated neuronal phenotype. As expected, the cells also developed more elaborate neurite processes than the untreated cells and **de-HS-CL** treated cells (Figure 2-21). Interestingly, cells

remodeled with **de-HS-CL** showed significantly lower amounts of SOX1 at day nine and TUJ1 at days three and six compared to untreated cells, which is consistent with studies suggesting that undersulfated HS may help to maintain pluripotency and restrict differentiation.[23] Together, our results illustrate the power of using long-term cell-surface displays of HS GAGs to activate specific signaling events and drive the differentiation of stem cells into mature neuronal populations.

## 2.4 Materials and Methods for Liposomal Experiments

### 2.4.1 Chemicals, Biochemical Reagents, Cell Lines, and Animals

All chemicals and reagents were of analytical grade, obtained from Sigma Aldrich, and used without further purification unless specified. Chondroitin sulfate (CS)-A, -C, and -E polysaccharides and chondroitinase ABC (ChABC) were purchased from Seikagaku Corporation. 1,2-Dioleoyl-*sn*-glycero-3-phosphoethanolaime (DOPE) and 1,2-dioleoyl-3-trimethylammoniumpropane chloride (DOTAP), were purchased from Avanti Polar Lipids, Inc. All secondary antibodies (AlexaFluor (AF) 680 goat anti-rabbit, AF680 goat anti-mouse, and AF488 goat anti-mouse), AF488 hydrazide sodium salt (AF488-hydrazide), and cell culture media and reagents were obtained from Life Technologies Corporation. Adherent pheochromocytoma cells (PC12 cells; CRL-1721.1), horse serum (HS), and F-12K media (30-2004) were obtained from American Type Culture Collection. Accutase (AM105) was received from Innovative Cell Technologies, Inc. Nerve growth factor (NGF) was obtained from R&D Systems, Inc. Primary antibodies against phospho-Akt (Ser473) (4060) and total Akt (4691) were purchased from Cell Signaling Technology, Inc. Sprague-Dawley rats were obtained from Charles River Laboratories, Inc.

### 2.4.2 Synthetic Methods

tert-*Butyl 2-((1,3-dioxoisoindolin-2-yl)oxy)acetate* (**2-1**). To a solution of *N*-hydroxyphthalimide (3.03 g, 18.6 mmol, 1.20 eq) in DMF (30 mL) was added

$NaHCO_3$ (1.56 g, 18.6 mmol, 1.20 eq). The mixture was stirred at 60 °C for 1 h, after which *tert*-butyl 2-bromoacetate (2.77 mL, 15.5 mmol) was added. The reaction mixture was stirred for 16 h at 60 °C. The mixture was then diluted with DCM (30 mL), extracted with 1 M $NaHCO_3$ until the aqueous layer became colorless, washed with $ddH_2O$ (2 x 25 mL) and brine (1 x 25 mL), dried over $MgSO_4$, and concentrated to afford a yellow solid **2-1**. The product was purified by silica flash chromatography (Hex:EtOAc, 85:15) to afford a white solid **2-1** (3.76 g, 88%). $^1$H NMR (500 MHz, $CDCl_3$): δ 1.48 (s, 9H, $CH_3$), 4.71 (s, 2H, $CH_2$), 7.75-7.76 (m, 2H, J = 5 Hz, Ar-H), 7.84-7.85 (m, 2H, J = 5 Hz, Ar-H).

*2-((1,3-Dioxoisoindolin-2-yl)oxy)acetic acid* (**2-2**). Compound **2-2** was synthesized as previously described.[43] Briefly, TFA (2 mL) was slowly added to a solution of 4 (0.51 g, 1.8 mmol) in DCM (6.0 mL). The mixture was stirred at RT for 2 h under Ar. The reaction was then concentrated and azeotroped with DCM (3 x 5 mL) to afford a white solid 5 (0.40 g, 99%). $^1$H NMR (500 MHz, $CDCl_3$): δ 4.88 (s, 2H, $CH_2$), 7.85-7.87 (m, 2H, J = 10 Hz, Ar-H), 7.93-7.94 (m, 2H, J = 10 Hz, Ar-H).

*2-((1,3-Dioxoisoindolin-2-yl)oxy)-conjugated CS polysaccharides* (**2-3, 2-4, 2-5**). To a buffered solution containing 0.1 M MES and 0.5 M NaCl (pH 6.0), **2-2** (5.0 mg, 0.023 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC; 4.2 mg, 0.027 mmol, 1.2 eq), sulfo-*N*-hydroxysuccinimide (sulfo-NHS; 7.5 mg, 0.035 mmol, 1.5 eq) were added, and the mixture was stirred at RT for 30 min. Concurrently, CS-E, CS-A, or CS-C (4.0 mg, ~70 μmol, ~0.5 eq) was stirred in

0.05 M $NaHCO_3$ for 30 min. The polysaccharide amine solution was then added to the small molecule acid mixture and allowed to react for 12 h at RT under Ar. The mixture was then purified via dialysis against $ddH_2O$ for 48 h (6 x 8 h) using a molecular weight membrane of 35,000 Da, followed by gel filtration chromatography (G-25). Fractions containing the polysaccharide were lyophilized to afford white solids. For **2-3** CS-E: $^1H$ NMR (500 MHz, $CDCl_3$): δ 2.04 (s, 3H, $CH_3$), 3.28-3.49 (m, 2H, CH), 3.51-4.35 (m, 9H, CH, $CH_2OSO_3$-), 4.42-4.67 (m, 1H, CH), 7.48 (m, 2H, Ar-H), 7.58 (m, 2H, Ar-H). 64% yield. For **2-4** CS-A: 1H NMR (500 MHz, $CDCl_3$): δ 2.03 (s, 3H, $CH_3$), 3.40 (br s, 1H, CH), 3.64 (br s, 1H, CH), 3.72-3.95 (m, 6H, CH), 4.03 (br s, 2H, CH), 4.51-4.63 (m, 2H, CH), 7.62 (m, 1H, Ar-H), 7.77 (m, 1H, Ar-H). 60% yield. For **2-5** CS-C: 1H NMR (500 MHz, $CDCl_3$): δ 2.02 (s, 3H, $CH_3$), 3.39 (br s, 1H, CH), 3.64 (br s, 1H, CH), 3.72-4.06 (m, 6H, CH, $CH_2OSO_3$-), 4.12-4.27 (m, 3H, CH), 4.57 (br s, 1H, CH), 7.62 (m, 2H, Ar-H), 7.75 (m, 2H, Ar-H). 78% yield.

*Aminooxy-conjugated CS polysaccharides* (**2-6**, **2-7**, **2-8**). Hydrazine monohydrate (1 µL, 0.031 mmol, 4.3 eq) was added dropwise to a solution of **2-3**, **2-4**, or **2-5** (4.0 mg, ~70 µmol) in $ddH_2O$ (4 mL), and the mixture was stirred for 3 h at RT under Ar. The solution was then purified via dialysis against $ddH_2O$ for 48 h (6 x 8 h) using a molecular weight membrane of 35,000 Da, followed by gel filtration chromatography (G-25). Fractions containing the polysaccharide were lyophilized to afford white solids. For **2-6** CS-E: $^1H$ NMR (600 MHz, $CDCl_3$): δ

2.04 (s, 3H, $CH_3$), 3.38 (br s, 1H, CH), 3.56-3.85 (m, 5H, CH), 3.86-4.33 (m, 4H, CH, $CH_2OSO_3$-), 4.42-4.67 (m, 2H, CH). For **2-7** CS-A: $^1$H NMR (500 MHz, $CDCl_3$): δ 2.04 (s, 3H, $CH_3$), 3.39 (br s, 1H, CH), 3.61 (br s, 1H, CH), 3.69-3.87 (m, 6H, CH), 4.08 (br s, 2H, CH), 4.46-4.64 (m, 2H, CH). For **2-8** CS-C: $^1$H NMR (500 MHz, $CDCl_3$) δ 2.03 (s, 3H, $CH_3$), 3.38 (br s, 1H, CH), 3.61 (br s, 1H, CH), 3.69-4.06 (m, 5H, CH, $CH_2OSO_3$-), 4.15-4.27 (m, 3H, CH), 4.46-4.65 (m, 2H, CH).

*Fluorophore-conjugated aminooxy-CS-E* (**2-9**). The protocol for conjugating aminooxy-CS-E with fluoresceinamine was adapted from Glabe *et al.*[44] To a stirring solution of **2-6** (1.00 mg, 14.29 μmol) in $ddH_2O$ (1 mL) was added cyanogen bromide (1.00 mg, 0.00094 mmol, excess) in $ddH_2O$ (0.3 mL). The pH of the reaction was adjusted to 11.0 using 0.2 M NaOH, which was stirred for 10 min at RT. The mixture was then desalted on a disposable PD-10 Sephadex column (17-0851-01, GE Healthcare Biosciences) using 0.2 M sodium borate buffer (pH 8.0). The CS-E fractions were pooled (3 mL) and immediately stirred with fluoresceinamine (1.00 mg, 0.0029 mmol, excess) in the dark overnight (12 h). The mixture was concentrated and purified on a G-25 Sephadex column in the dark. The CS-E fractions were again pooled, flash frozen, and lyophilized to afford an orange solid 2-**9**.

### 2.4.3 Liposomal Formation and Characterization

*Liposome formation*. A solution of 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE, 120 μL, 10 mg/mL in $CHCl_3$), dodecanone (12 μL, 10 mg/mL in $CHCl_3$),

and 1,2-dioleoyl-3-trimethylammoniumpropane chloride salt (DOTAP, 60 µL, 10 mg/mL in $CHCl_3$) was thoroughly mixed and then dried with a stream of argon followed by high vacuum for 12 h. The dried lipids were reconstituted in dd$H_2O$ (2 mg/mL total lipids) and hydrated for 30 min. The mixture was then sonicated for 2 min and extruded (11 passes through the membrane) to form large unilaminar vesicles using a mini-extruder (Avanti Polar Lipids, Inc.) according to the manufacturer's instructions.

*Liposome conjugation*. AF488-hydrazide-presenting liposomes: Liposomes (100 µL, 2.24 mM dodecanone) were incubated with AF488-hydrazide (12.8 µL, 1.75 mM, 10:1 ketone/hydrazide) for 3 h at RT in the dark overnight at a stoichiometric ratio to label approximately 10% of the available ketones. This ratio was chosen after method optimization to conserve carbohydrate material but maintain a significant amount of cell labeling. CS-aminooxy-tethered liposomes: Liposomes (100 µL, 2.24 mM dodecanone) were reacted with CS-A/C/E-aminooxy-polysaccharides **2-6**, **2-7**, or **2-8** (~20 µL, 33.3/31.7/25.6 µM) for 3 h at RT in the dark overnight. Mixed liposomes: Liposomes (100 µL, 2.24 mM dodecanone) were reacted with CS-A/C/E-aminooxy polysaccharides **2-6**, **2-7**, or **2-8** (~20 µL, 33.3/31.7/25.6 µM) for 4 h at RT in the dark, followed by AF488-hydrazide at a stoichiometric ratio to label approximately 10% of the available ketones.

*Dynamic light scattering (DLS).* Ketone-presenting (**I**) and AF488-hydrazide- (**II**), CS-A-$ONH_2$- (**IIIA**), CS-E-$ONH_2$- (**IIIE**), and mixed AF488/CS-A-$ONH_2$- (**IV**) conjugated liposomes were generated as described above and diluted (1:20) with 1.5 mM KCl. The size distributions of these liposomal samples were then determined using a DynaPro NanoStar (Wyatt Technologies) DLS instrument. The data are represented as the mean liposomal hydrodynamic diameter ± S.E.M. with eight to ten measurements conducted per experiment ($n$ = 3).

*Zeta potential analyses.* Ketone-presenting (**I**) and AF488-hydrazide- (**II**), CS-A-$ONH_2$- (**IIIA**), and CS-E-$ONH_2$- (**IIIE**), mixed AF488/CS-A-$ONH_2$- (**IV**) conjugated liposomes were generated as described above, diluted (1:20) with 1.5 mM KCl, and subjected to zeta potential analyses using a ZetaPALS (Brookhaven Instruments Corp.) analyzer. The data are represented as the mean zeta potential ± S.E.M. with three to ten measurements conducted per experiment ($n$ = 3).

*Transmission electron microscopy (TEM) and energy dispersion spectroscopy (EDS).* Dodecanone-containing liposomes were prepared and conjugated with AF488-hydrazide and CSA-aminooxy polysaccharide **2-7** as described above. A drop of each liposome sample was placed on a TEM copper/carbon grid for 30 s, which was then blotted using filter paper. A drop of 2% uranyl acetate was then added to each sample for 5 min followed by blotting with filter paper. The grid was then rinsed with $ddH_2O$ and dried. TEM and elemental analyses of each liposome sample were performed using an FEI Tecnai F30UT (300 kV) equipped with a

high angle annular dark field detector, an Oxford energy dispersion X-ray detector, and an AMT CCD camera.

### 2.4.4 Biological Procedures

*Cell culture*. PC12 cells were cultured on collagen-coated (10 µg/mL) tissue culture plates in F-12K media supplemented with 6.5% HS, 6.5% FBS, and 1x penicillin/streptomycin (P/S) at 37 °C with 5% CO2. Hippocampal and cortical neurons were cultured in Neurobasal media supplemented with B27, 20 mM GlutaMAX, and 1x P/S at 37 °C with 10% $CO_2$.

*Cell-surface remodeling*. AF488-hydrazide-bearing liposomes were diluted 1:100 with the appropriate cell culture medium and incubated with PC12 cells or hippocampal neurons for 30-45 min. The medium was removed, and the cells were washed three times with 0.5 mL of sterile 1x phosphate buffered saline (PBS). Fresh medium (0.5 mL) was then added, and the cells were imaged live under a Zeiss LSM 5 Pascal inverted confocal microscope (Carl Zeiss Microscopy). For labeling differentiated PC12 cells with CS-E, PC12 cells were cultured on collagen-coated tissue culture plates and stimulated with NGF (100 µM; 256-GF, R&D Systems) to induce differentiation. After 4 d of NGF treatment, the cells were treated with 1 U/mL ChABC (1:1 activation buffer (0.4 M Tris/HCl pH 8.0, 0.4 M NaOAc, 0.1% BSA)/complete medium) for 2 h (37 °C, 5% CO2). The cells were then washed with PBS (4 x 0.5 mL), and medium that was supplemented with CS-E-functionalized liposomes (diluted 1:100) was added for

30-45 min. After fusion, the cells were fixed with 4% paraformaldehyde (PFA) in PBS for 20 min, stained with a monoclonal mouse anti-CS-E antibody (1:250 in 1% BSA/0.1% Triton X-100/PBS; 4 °C, 12 h), followed by an AF488-conjugated goat anti-mouse antibody (1:1000 in 1% BSA/0.1% Triton X-100/PBS; RT, 1 h), and imaged using a Zeiss LSM-5 Pascal inverted confocal microscope.

*Immunochemistry*. For imaging, cells were fixed with 4% PFA in PBS (15 min), rinsed with PBS twice, permeated, and blocked with a solution containing 1% BSA and 0.1% Triton X-100 (20 min), incubated with a monoclonal mouse anti-CS-E antibody (1:250; 4 °C, 12 h), washed with PBS (2 x 5 min), incubated with an anti-mouse AF488 antibody (1:1000; RT, 1 h), washed with PBS (2 x 5 min), and imaged under a Zeiss LSM 5 Pascal inverted confocal microscope.

*Flow cytometry*. Dodecanone-containing liposomes were prepared as mentioned above with the exception that the ratio of dodecanone to DOPE was varied (0, 5, 10, 15, and 20% w/w). PC12 cells were cultured until ~80% confluent and then detached from the tissue culture plate using TrypLE. The cells were twice washed with cold PBS (500 µL) and spun down (1,400 x *g*, 2.5 min). For AF488-hydrazide labeling, the cells with resuspended and incubated on ice in a buffered solution (45 mM BSA, 10 mM HEPES, and DNase I (0.5 U/mL) in HBSS at pH 7.4) containing 0, 5, 10, 20% w/w ketone-hydrazide conjugated liposomes for 30 min (diluted 1:100). The dye concentration was adjusted per condition according to the ketone concentration in the liposomes at a stoichiometric ratio to label

approximately 10% of the available ketones. After labeling, the cells were again twice washed with cold PBS (500 µL) and spun down (1,400 x *g*, 2.5 min). After resuspension in the buffer listed above, the cells were filtered through a 40 µm-filtered FACS tube, and subjected to flow cytometric analyses using a FACS-Calibur flow cytometer (BD Biosciences). For CS-E labeling, PC12 cells were incubated with a monoclonal mouse anti-CS-E antibody (1:250) for 1 h followed by two wash cycles, incubated with an AF488-conjugated goat anti-mouse antibody (1:1000) for 1 h on ice, washed again twice, and subjected to flow cytometric analyses. The labeled samples were referenced to untreated cells. As controls, PC12 cells were separately incubated with liposomes without the ketone and hydrazide-AF488 or the AF488-conjugated goat anti-mouse antibody. Each sample was run in duplicate, and each sample set was performed twice.

*Cell internalization assay*. Liposomes were reacted with 9 as mentioned above for 3 h at RT in the dark. The liposome stock solution was diluted 1:100 with HBSS (45 mM BSA, 10 mM HEPES, pH 7.4) and added (400 µL) to PC12 cells for 30 min on ice in the dark. After labeling, the cells were diluted with PBS (500 µL), washed (3 x 500 µL wash/centrifuge cycles), and added to RPMI medium without phenol red. The cells were imaged live at 0, 3, 6, 10, 16, and 24 h after labeling under a Zeiss 700 LSM confocal microscope.

*Neurite outgrowth assays*. Hippocampal neurons from E18 Sprague-Dawley rats were dissected as previously described[45] and plated at 2,000 cells/well on poly-

lysine-coated 96-well plates (BD Biosciences) in 50 µL of complete DMEM (10% FBS, 1% P/S) for 30 min, followed by 150 µL of complete Neurobasal (B27, 20 mM GlutaMAX, 1% P/S). After 48 h, CS-A, -C, and -E-conjugated liposomes (10% w/w dodecanone; 2.5-10% ketone for CS-E dose-dependent studies) were diluted 1:100 in 200 µL of complete Neurobasal and added. The medium was replaced after 30 min to remove the remaining liposomes and unreacted CS. Due to endocytosis of CS after approximately 6-10 h, cells were treated with polysaccharide-displaying liposomes again at 8 and 16 h, followed by medium replacement 30 min after addition. At each time point, the cells were fixed with 4% PFA (20 min), washed with PBS (2 x 5 min), permeabilized with 1% BSA and 0.1% Triton X-100 in PBS (15 min), washed with PBS (2 x 5 min), incubated with a mouse α-tubulin antibody (Sigma-Aldrich; 1:1000) at 4 °C for 12 h, washed with PBS (2 x 5 min), incubated with a AF488-conjugated goat anti-mouse antibody (1:1000) at RT for 1 h, washed with PBS (2 x 5 min), and imaged under a Zeiss LSM 5 Pascal inverted confocal microscope as previously described.[17] Each well was reconstructed in Adobe Illustrator after capturing a 9 x 9 grid of 2034-pixel images. Four to five wells were blindly subjected to MetaMorph (Molecular Devices) neurite outgrowth software analyses to calculate the mean outgrowth per neuron. As controls, untreated neurons, neurons subjected to ketone liposomes without polysaccharide, and neurons treated with ketone liposomes (+CS-E without an aminooxy group) were assayed for outgrowth. The data are

represented as the mean outgrowth percentage (± S.E.M.) with $n$ = 7 (200 cells measured per assay).

*Stimulation assays.* Cortical neurons from E18 Sprague-Dawley rats were dissected as previously described[46] and plated at 5 x $10^6$ cells per poly-ornithine-coated (Sigma Aldrich) 10-cm plate in complete DMEM (10% FBS, 1% P/S) for 30 min, followed by complete Neurobasal (B27, 20 mM GlutaMAX, 1% P/S). After 7 d, CS-C- and CS-E-conjugated ketone liposomes (10% w/w dodecanone) were added to the complete Neurobasal for 30 min. The medium was then replaced with fresh complete Neurobasal that contained 50 ng/mL NGF (256-GF; R&D Systems), and the neurons were stimulated for 0, 10, 30, and 60 min. At each time point, the medium was evacuated, the cells were scraped and collected in cold PBS (1 mL), centrifuged (4 °C, 3,000 x *g*, 2.5 min), and homogenized in lysis buffer (50 mM Tris/HCl pH 7.4, 250 mM mannitol, 50 mM NaF, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM $Na_2VO_4$, 1% Triton X-100, and 1x complete protease inhibitors (Roche) by 10 passages through a 26-gauge 3/8" needle on ice. The lysates were then cleared (4 °C, 16,000 x *g*, 15 min), and the total protein concentration was determined for each sample using a BCA assay kit (Thermo Scientific Pierce). Western blot analyses were performed as previously described,[17] blotting separately with a rabbit anti-phospho-Ser473 Akt antibody and a rabbit anti-Akt antibody at 1:1000 in 5% BSA. Akt activation was calculated by determining the ratio of phospho-Akt to total Akt for each time point and

condition normalized to α-tubulin staining. Relative Akt activation was reported with respect to the untreated values (+NGF, no liposomes).

## 2.5 Materials and Methods for HaloTag Experiments

### 2.5.1 Synthetic Methods

*N-Boc-2-(2-hydroxyethoxy)ethylamine* (**2-10**). Compounds **2-10** and **2-11** were synthesized as previously described.[31] A solution of di-*tert*-butyl dicarbonate ($Boc_2O$) (2.18 g, 10.0 mmol, 1 eq) in anhydrous methanol (2 mL) was added dropwise to a stirring solution of 2-(2-aminoethoxy)ethanol (1.05 g, 9.99 mmol) in anhydrous methanol (20 mL) at 0 °C. The mixture was stirred for 30 min at 0 °C and then warmed to RT. After 2 h, the solution was diluted with DCM (30 mL), washed with $ddH_2O$ (3 x 25 mL) and brine (1 x 25 mL), dried over $MgSO_4$, filtered, and concentrated to afford a colorless oil **2-10** (1.85 g, 91%). $^1$H NMR (500 MHz, $CDCl_3$): δ 1.42 (s, 9H, $CH_3$), 3.30 (t, $J$ = 5.1 Hz, 2H, $CH_2N$), 3.49-3.58 (m, 4H, $CH_2OCH_2$), and 3.68-3.75 (m, 2H, $CH_2OH$). MS (ESI) calcd. for $C_4H_{11}NO_2^+$ [M + $H^+$] 206.14, found 206.10.

*N-Boc-1-(2-(2-aminoethoxy)ethoxy)-6-chlorohexane* (**2-11**). A solution of **2-10** (0.34 g, 1.7 mmol) in DMF (2 mL) was added dropwise to a stirring solution of 60% NaH (0.095 g, 2.4 mmol, 1.4 eq) in DMF (30 mL) at 0 °C. The mixture was stirred for 1 h at 0 °C, after which 1-chloro-6-iodohexane (0.30 mL, 2.0 mmol, 1.2 eq) was added. Stirring was continued for 24 h at RT, and then the reaction was

quenched with 1 M HCl and extracted with DCM, $ddH_2O$ (6 x 25 mL) and brine (1 x 25 mL), dried over $MgSO_4$, filtered, and concentrated to afford a yellow oil. Purification by silica gel flash chromatography using 3:1 Hex/EtOAc gave **2-11** as a colorless oil (0.20 g, 37%). $^1$H NMR (500 MHz, $CDCl_3$): δ 1.33-1.40 (m, 2H, $CH_2$), 1.43 (s, 9H, $CH_3$), 1.44-1.48 (m, 2H, $CH_2$), 1.55-1.64 (m, 2H, $CH_2$), 1.73-1.80 (m, 2H, $CH_2$), 3.30 (t, *J* = 5.2 Hz, 2H, $CH_2N$), 3.45 (t, *J* = 6.7 Hz, 2H, $CH_2Cl$), 3.49-3.57 (m, 6H, $CH_2O$), and 3.57-3.61 (m, 2H, $CH_2O$). MS (ESI) calcd. for $C_{14}H_{29}ClNO_5^+$ [M + $H^+$] 324.20, found 324.21.

*2,5-Dioxopyrrolidin-1-yl 4-oxopentanoate* (**2-12**). *N*-Hydroxysuccinimide (NHS; 0.55 g, 4.8 mmol, 1.1 eq) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) hydrochloride (0.91 g, 4.8 mmol, 1.1 eq) were added to a stirring solution of levulinic acid (0.50 g, 4.3 mmol) in anhydrous DCM at 0 °C. The mixture was warmed to RT and stirred for 2 h, at which time the reaction was diluted with DCM, extracted with $ddH_2O$ (3 x 25 mL) and brine (1 x mL), dried over $MgSO_4$, filtered, and concentrated to afford a white fluffy solid **2-12** (0.69 g, 75%). $^1$H NMR (500 MHz, $CDCl_3$): δ 2.21 (s, 3H, $CH_3$), 2.83 (s, 4H, $CH_2CH_2$-succinimide), and 2.86-2.92 (m, 4H, $CH_2CH_2$-levulinate). MS (ESI) calcd. for $C_9H_{11}NO_5^+$ [M + $H^+$] 214.08, found 214.10.

*N-(2-(2-(6-Chlorohexyloxy)ethoxy)ethyl)-4-oxopentanamide* (**2-13**). Trifluoroacetic acid (TFA) (0.14 mL, 1.8 mmol, 6 eq) was added dropwise to a stirring solution of **2-11** (0.10 g, 0.31 mmol) in anhydrous DCM (5 mL). The mixture was stirred at

RT for 4 h, then azeotroped with toluene, and concentrated to afford the free amine as a yellow oil, which was used without further purification. This intermediate (0.025 g, 0.074 mmol) was stirred with *N*,*N*-diisopropylethylamine (DIPEA; 0.076 mL, 0.436 mmol, 6 eq) in anhydrous DCM (3 mL) for 20 min at RT, to which **2-12** (0.016 g, 0.075 mmol, 1 eq) was added and stirred overnight. After 24 h, the reaction was diluted with DCM (20 mL) and extracted with dd$H_2O$ (3 x 25 mL) and brine (1 x 25 mL), dried over $MgSO_4$, filtered, and concentrated to afford a yellow oil. Purification by silica gel flash chromatography using 97:3 DCM/MeOH gave **2-13** as a colorless oil (0.016 g, 67%). $^1$H NMR (500 MHz, $CDCl_3$): δ 1.33-1.41 (m, 2H, $CH_2$), 1.42-1.49 (m, 2H, $CH_2$), 1.56-1.65 (m, 2H, $CH_2$), 1.73-1.81 (m, 2H, $CH_2$), 2.18 (s, 3H, $CH_3$), 2.43 (t, *J* = 6.6 Hz, 2H, $CH_2C(O)$), 2.79 (t, *J* = 6.6 Hz, 2H, $CH_2C(O)$), 3.41-3.49 (m, 4H, $CH_2Cl$ and $CH_2N$), 3.51-3.56 (m, 4H, $CH_2O$), 3.56-3.63 (m, 4H, $CH_2O$), 6.15 (bs, 1H, NH). MS (ESI) calcd. for $C_{15}H_{28}ClNO_4^+$ [M + $H^+$] 322.18, found 322.20.

*Fluorescein-chloroalkane linker conjugate* (**F-CL**). To a solution of **2-13** (2.6 mg, 8.1 μmol) in 0.5 mL anhydrous MeOH was added 5-(((2-(carbohydrazino)methyl)thio)acetyl)aminofluorescein (F) in dd$H_2O$ (4.0 mg, 8.1 μmol, 1 eq; C-356, Life Technologies). The mixture was stirred for 12 h at RT in the dark, concentrated, and purified by silica gel flash chromatography in the dark using a 97:3 DCM/MeOH mixture to afford the desired compound as an orange solid (4.8 mg, 75%). $^1$H NMR (500 MHz, $D_2O$): δ 1.26-1.33 (m, 2H, $CH_2$), 1.33-

1.39 (m, 2H, $CH_2$), 1.40 (s, 3H, $CH_3$), 1.44-1.52 (m, 2H, $CH_2$), 1.63-1.70 (m, 2H, $CH_2$), 3.32-3.55 (m, 20H, $CH_2$), 6.42-6.48 (m, 2H, Ar-H) 6.49-6.60 (m, 4H, Ar-H), 7.07 (dd, *J* = 8.3, 2.9 Hz, 1H, Ar-H), 7.73-7.81 (m, 1H, Ar-H), and 8.21-8.30 (m, 1H, Ar-H). MS (ESI) calcd. for $C_{39}H_{43}ClN_4O_{10}S^{-2}Na^+$ [$M^{-2}$ + $Na^+$] 817.23, found 817.25.

*Biotinylated heparan sulfate* (**B-HS**). To a stirring solution of EZ-Link NHS-$PEG_4$-biotin (0.10 g, 0.17 mmol; 21363, Thermo Scientific) in dry DCM (5 mL) was added a solution of ethylenediamine (0.022 mL, 0.33 mmol, 2 eq) and triethylamine (TEA; 0.15 mL, 1.1 mmol, 6 eq) in dry DCM (1 mL). The mixture immediately turned cloudy upon addition and was then stirred for 1.5 h at RT. The precipitate was filtered, and the reaction was concentrated to afford the conjugated amine as a white solid (0.068 g, 76%), which was used without further purification. De-6-*O*-sulfated HS (7.0 mg, 0.58 μmol) was dissolved in dd$H_2O$ (1 mL), followed by the addition of cyanogen bromide in dd$H_2O$ (5.0 mg, 47 μmol, excess). The pH was adjusted to 11.0 using 0.2 M NaOH and stirred for 10 min. The mixture was then desalted on a PD-10 Sephadex column using 0.2 M sodium borate (pH 8.0). The HS fractions were pooled (3 mL) and immediately stirred with the biotin-conjugated amine (5.0 mg, 9.4 μmol, excess) overnight (12 h). The mixture was then flash frozen, lyophilized, redissolved in dd$H_2O$ (0.5 mL), and purified using a G-25 Sephadex column. The desired fractions were pooled, flash frozen, and lyophilized to afford **B-HS** as a white solid (99% recovery). $^1$H NMR

(600 MHz, $D_2O$): δ 3.29 (bs, 1H), 3.72-3.91 (m, 5H), 4.04-4.13 (m, 1H), 4.18-4.32 (m, 1H), 4.33-4.44 (m, 1H), 4.84 (bs, 1H), 5.26 (bs, 1H, IdoA C-1), 5.42 (bs, 1H, GlcN C-1); substoichiometric: 1.30 (bs, 2H, $CH_2$-biotinPEG$_4$), 1.43 (bs, 2H, $CH_2$-biotinPEG$_4$), 1.50-1.52 (m, 2H, $CH_2$-biotinPEG$_4$), 2.05 (bs, 3H, NHAc), 2.29 (t, 2H, $CH_2$-biotinPEG$_4$), 2.52-2.60 (m, 2H, $CH_2$-biotinPEG$_4$), 2.78-2.82 (m, 2H, C(O)$CH_2$-biotinPEG$_4$) 2.99-3.03 (m, 1H, CH-biotinPEG$_4$), 3.32-3.54 (m, 22H, $CH_2O$- and $CH_2N$-biotinPEG$_4$). Biotin(PEG)$_4$ incorporation was estimated to be 0.9 molecules per polysaccharide.

*Biotinylated heparan sulfate with chloroalkane linker* (**B-HS-CL**). A solution of **2-13** (4.0 mg, 12 μmol, excess) and $NaBH_3CN$ (2.0 mg, 32 μmol, excess) in 1:1 MeOH:dd$H_2O$ (400 μL) was added to **B-HS** (2.0 mg, 0.17 μmol) in dd$H_2O$ (500 μL). MeOH (approximately 300 μL) was added until the reaction turned from cloudy to colorless, and the mixture was stirred for 12 h at RT, concentrated, and purified by gel filtration chromatography using G-25 Sephadex resin. The pooled fractions containing the polysaccharide were lyophilized to yield a white powder (99% recovery). $^1$H NMR (600 MHz, $D_2O$): δ 3.24 (bs, 1H), 3.62-3.89 (m, 5H), 4.04-4.14 (m, 1H), 4.17-4.32 (m, 1H), 4.33-4.44 (m, 1H), 5.24-5.45 (m, 2H, IdoA C-1, GlcN C-1); substoichiometric: 1.07 (d, $J$ = 6.3 Hz, 3H, $CH_3$-CL), 1.50-1.54 (m, 2H), 2.01 (s, 3H, NHAc), 2.18-2.23 (m, 6H), 2.34-2.45 (m, 2H), 2.75 (s, 2H), 2.99-3.10 (m, 2H). CL incorporation was estimated to be 1.8 CL molecules per polysaccharide.

*General procedure for chloroalkane heparin/heparan sulfate derivatives.* In a typical procedure, **2-13** (4.0 mg, 12 μmol, excess) was dissolved in MeOH and $ddH_2O$ (1:1, 400 μL) with $NaBH_3CN$ (2.0 mg, 32 μmol, excess) and added to HS, de-HS, deO-HS, 6-deO-HS, or 2-deO-HS (2.0 mg, 0.17 μmol) in $ddH_2O$ (500 μL). MeOH (approximately 300 μL) was added until the reaction turned from cloudy to colorless, and the mixture was stirred for 12 h at RT, concentrated, and purified by gel filtration chromatography using G-25 Sephadex resin. The pooled fractions containing the polysaccharides were lyophilized to yield white powders (99% recovery of polysaccharide in all cases). CL incorporation was estimated to be between 0.4 to 4 molecules per polysaccharide. **HS-CL**: $^1$H NMR (600 MHz, $D_2O$): δ 3.29 (bs, 1H), 3.78-3.91 (m, 2H), 4.06 (bs, 1H), 4.12 (bs, 1H), 4.20-4.31 (m, 2H), 4.33-4.44 (m, 2H), 4.91 (bs, 1H), 5.27-5.46 (m, 2H, IdoA C-1, GlcN C-1); substoichiometric: 1.19 (dd, *J* = 6.3, 0.9 Hz, 3H, $CH_3$-CL), 1.67-1.77 (m, 2H), 2.06 (bs, NHAc), 2.20-2.31 (m, 6H), 2.38-2.45 (m, 2H), 2.73 (s, 2H), 2.77-2.86 (m, 2H). **de-HS-CL**: $^1$H NMR (600 MHz, $D_2O$): δ 3.77-3.94 (m, 4H), 3.96-4.12 (m, 2H), 4.19-4.38 (m, 2H), 4.92 (bs, 1H), 5.08-5.23 (m, 2H, IdoA C-1), 5.40 (bs, 1H, GlcN C-1); substoichiometric: 1.19 (d, *J* = 6.2 Hz, 3H, $CH_3$-CL), 1.68-1.77 (m, 2H), 2.02 (bs, 3H, NHAc), 2.41 (t, *J* = 6.8 Hz, 2H), 2.66 (s, 2H), 2.79 (t, *J* = 6.8 Hz, 2H). **deO-HS-CL**: $^1$H NMR (600 MHz, $D_2O$): δ 3.26 (bs, 1H), 3.62-3.77 (m, 4H), 3.78-3.93 (m, 3H), 4.06 (bs, 1H), 4.12 (bs, 1H), 4.95 (bs, 1H, IdoA C-1), 5.38 (bs, 1H, GlcN C-1); substoichiometric: 1.19 (d, *J* = 6.3 Hz, 3H, $CH_3$-CL), 1.70-1.75 (m, 2H), 2.04 (bs, 3H, NHAc), 2.21-2.26 (m, 6H), 2.40-2.43 (m, 2H), 2.66 (s, 2H),

2.78 (t, *J* = 6.9 Hz, 2H). **6-deO-HS-CL**: $^1$H NMR (600 MHz, $D_2O$): δ 3.26 (bs, 1H), 3.64-3.95 (m, 5H), 4.05 (bs, 1H), 4.26 (bs, 1H), 4.36 (bs, 1H), 4.93 (bs, 1H), 5.28-5.41 (m, 2H, IdoA C-1, GlcN C-1); substoichiometric: 1.19 (dd, *J* = 6.3, 1.7 Hz, 3H, $CH_3$-CL), 1.69-1.76 (m, 2H), 2.06 (bs, 3H, NHAc), 2.18-2.30 (m, 6H), 2.41 (td, *J* = 6.9, 1.7 Hz, 3H), 2.66 (d, *J* = 1.7 Hz, 2H), 2.79 (td, *J* = 6.9, 1.5 Hz, 2H). **2-deO-HS-CL**: $^1$H NMR (600 MHz, $D_2O$): δ 3.26 (bs, 1H), 3.60-3.93 (m, 5H), 3.94-4.29 (m, 3H), 4.35 (bs, 1H), 5.30-5.45 (m, 2H, GlcN C-1, IdoA C-1); substoichiometric: 1.19 (d, *J* = 6.3 Hz, 3H, $CH_3$-CL), 1.69-1.77 (m, 2H), 2.05 (bs, 3H, NHAc), 2.21-2.34 (m, 6H), 2.44 (t, *J* = 6.8 Hz, 2H), 2.72-2.83 (m, 4H).

### 2.5.2 Biological Procedures

*Transmembrane HaloTag protein DNA construct*. The sequence encoding the HaloTag protein (HTP) was PCR amplified from the pFC14K HT7 CMV Flexi Vector (Promega) to introduce a 5' SacII restriction site and a 3' SalI restriction site using the following primers:

```
HTP-upstream:   TTATCCGCGGTGGATCCGAAATCGGTACTGGCTTT
HTP-downstream: ACTAGTCGACACCGGAAATCTCCAGAGTAGACAG
```

The PCR product was digested with SalI and SacII according to standard procedures and gel purified. The pDisplay vector (pD; Life Technologies) was linearized with SalI and SacII according to standard procedures and purified using a QIAquick PCR purification kit (Qiagen). The insert was ligated into the linearized vector using T4 DNA ligase. The HTP gene was inserted between the

sequence encoding the N-terminal murine Ig κ-chain leader and the C-terminal PDGFR transmembrane domain to generate the final HTP construct. The plasmid was confirmed by Sanger sequencing (Laragen Inc.).

*Cell culture*. All cell cultures were maintained at 37 °C and 5% $CO_2$ unless otherwise indicated. CHO cells were cultured on tissue culture plates in F12 medium supplemented with 10% FBS and 1% P/S and were passaged after reaching ~80% confluency. mESCs were cultured strictly according to ATCC protocols to maintain a homogeneous population of pluripotent cells. Confluent MEFs were treated with mitomycin C (10 μg/mL in DMEM with 10 % FBS and 1% P/S) for 3 h at 37 °C to provide a feeder cell layer for mESCs. mESCs were cultured directly on the MEF feeder layer in DMEM (ATCC) with 10% non-heat-inactivated FBS and 0.1 mM β-mercaptoethanol (defined henceforth as complete DMEM). To passage or isolate mESCs for transfection, collagenase IV (10 μg/mL) in DPBS (with $CaCl_2$ and $MgCl_2$) was equilibrated to 37 °C and 5% $CO_2$ and added to mESCs (3 mL) and allowed to detach the cells for 3-5 min at 37 °C. The cells were gently agitated to remove the large colonies, leaving the majority of remaining MEFs attached to the plate. The released mESCs were then added to 6 mL of pre-warmed complete DMEM and centrifuged (1,200 rpm for 2.5 min). The collagenase IV-containing medium was aspirated, and the cells were resuspended in 10 mL of fresh pre-warmed complete DMEM. At this time, the cells were either transferred to new tissue culture plates on top of a confluent layer of mitomycin

C-treated MEFs for further culture or to plates pre-coated with 0.2% gelatin for transfection. mESCs were passaged every 3 to 5 days, depending on their colony sizes and proximity to one another, and the medium was replaced every 2 days.

*HTP transfections*. CHO cells were plated and transfected with the HTP plasmid using Lipofectamine LTX (Life Technologies) according to the manufacturer's protocol. For transfecting mESCs, the protocol for Xfect mESC Transfection Reagent was followed. Briefly, mESCs were plated on 0.2% gelatin-coated 6-well plates at 1 x $10^6$ cells/well in complete DMEM. After 5 h, the HTP plasmid-lipid transfection complexes were generated and added according to the manufacturer's instructions. The mESCs were then incubated for 3 h with the transfection complexes at 37 °C and 5% $CO_2$, and the medium was replaced with 2 mL of pre-warmed complete DMEM.

*Generation of an HTP-expressing stable cell line*. CHO cells lacking galactosyltransferase I (pgsB-618, ATCC) that had been transfected with HTP were selected using complete F12 medium supplemented with 700 μL of G418 solution (0.8 mg/mL). The medium was replaced every 2 d for 3 weeks. After 3 weeks, the cells were incubated with fresh F12 medium containing 5 μM **F-CL** for 1 h (37 °C, 5% $CO_2$) and then sorted via fluorescence-activated cell sorting (FACS) using a FACS-Aria Flow Cytometer Cell Sorter (BD Biosciences) with an excitation wavelength of 488 nm.

*HTP labeling for fluorescent imaging analyses.* CHO cells were plated and cultured on poly-ornithine-coated (10 μg/mL) glass coverslips in complete F12 medium (10% FBS, 1% P/S) at 37 °C and 5% $CO_2$. For labeling, the medium was replaced with fresh F12 medium that contained 5 μM **F** or **F-CL** for 1 h (37 °C, 5% $CO_2$). The cells were then washed with PBS (2 x 1 mL), fixed with 4% paraformaldehyde in PBS (15 min at RT), rinsed with PBS (2 x 1 mL), permeabilized with a solution containing 1% BSA and 0.1% Triton X-100 (20 min at RT), incubated with a monoclonal mouse anti-Myc antibody (1:1000; 12 h at 4 °C), washed with PBS (2 x 5 min), incubated with a goat anti-mouse AF568 secondary antibody (1:1000; 1 h at RT), and washed with PBS (2 x 5 min). The coverslips were mounted with VECTASHIELD mounting medium containing DAPI (Vector Laboratories Inc.) and imaged using a Zeiss LSM 700 inverted confocal microscope.

*HTP labeling for Western blot analyses.* CHO cells stably expressing HTP were cultured on 10-cm plates in complete F12 medium until approximately 85% confluent. The medium was then replaced with fresh F12 medium (4 mL) that contained 5 μg/mL **B-HS** or **B-HS-CL**, and the cells were incubated for 6 h at 37 °C. After labeling, the medium was removed, and the cells were scraped and collected in cold PBS (1 mL), centrifuged (3,000 rpm for 2.5 min at 4 °C), and homogenized in 150 μL of lysis buffer (50 mM Tris/HCl pH 7.4, 250 mM mannitol, 50 mM NaF, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM $Na_2VO_4$, 1%

Triton X-100, and 1x cOmplete protease inhibitor cocktail (Roche) via 10 passages through a 26g x 3/8" needle on ice. The lysates were then clarified by centrifugation (15,000 rpm for 15 min at 4 °C) and transferred to clean microcentrifuge tubes, after which the total protein concentration was determined for each sample using the Bradford assay (Bio-Rad Laboratories). Samples were boiled for 5 min with SDS-PAGE loading buffer and resolved by SDS-PAGE on 4-12% Bis-Tris gels (Life Technologies). Proteins were transferred to PVDF membranes and blocked with blocking buffer (5% milk in a solution of 50 mM Tris pH 7.6, 150 mM NaCl, and 0.5% Tween-20 (TBST)) for 1 h at RT. Blots were incubated with IRDye800 streptavidin (1:1000; LI-COR Biosciences) and an anti-HA rabbit monoclonal antibody (1:1000) overnight at 4 °C in blocking buffer, rinsed twice with TBST, and incubated again with an AF680-conjugated goat anti-rabbit IgG antibody (1:10000) for 1 h at RT. Membranes washed three times with TBST for 5 min at RT and were then imaged and quantified using an Odyssey scanner (LI-COR Biosciences).

*HTP-conjugate lifetime assay*. CHO cells were plated onto 15-mm coverslips, cultured, and labeled with 5 μM **F** or **F-CL** in F12 medium for 1 h (37 °C, 5% $CO_2$). Cells were rinsed once with prewarmed PBS and subsequently cultured with fresh F12 medium (37 °C, 5% $CO_2$). At each time point (between 0.5 and 8 d), the coverslips were removed from culture, submerged in PBS for 2 min, fixed with 4% paraformaldehyde in PBS for 15 min, and then washed in PBS for 5 min.

The coverslips were blotted dry, mounted onto glass microscope slides using VECTASHIELD mounting medium containing DAPI, and imaged using a Zeiss LSM 700 inverted confocal microscope.

*Microarray analyses*. Glycosaminoglycan microarrays were generated as previously described.[18] Slides were blocked with 10% FBS in PBS for 1 h at RT. Stock solutions of FGF2, FGF8b, FGFR1-Fc (2 μM), and FGFR3c-Fc were prepared in sterile 1x PBS containing 1% BSA. Blocked slides were washed once with PBS and slowly rocked with 150 μL of 1% BSA/PBS containing 1 μM receptor with or without 1 μM ligand (*i.e.*, 1 μM FGFR1-Fc ± 1 μM FGF2; 1 μM FGFR3-Fc ± 1 μM FGF8b) for 2 h at RT. Slides were washed three times with PBS and incubated with an AF647-conjugated goat anti-human Fc antibody (1:5000) in 1% BSA/PBS for 1 h at RT in the dark with gentle rocking. Slides were then washed three times with PBS and twice with dd$H_2O$ and blown dry under a stream of filtered air. Arrays were scanned using a G2565BA DNA Microarray Scanner (Agilent), and fluorescence was quantified using GenePix 5.0 software (Molecular Devices) with normalization against local background. The data represent the average of 10 spots per concentration of polysaccharide.

*ERK activation assays*. Pluripotent mESCs were transfected with the HTP plasmid as described above. After 48 h, the cells were detached with collagenase IV (10 μg/mL, 3 mL) in DPBS, pelleted, and resuspended in 2 mL of complete DMEM containing 1 U/mL heparinase II (HepII, Sigma-Aldrich) for 2 h at 37 °C with

gentle mixing every 15 min. HepII was quenched with 6 mL of complete DMEM, and the cells were pelleted and resuspended in complete DMEM. mESCs were then plated on six-well plates coated with 0.2% gelatin and incubated with **HS-CL**, **de-HS-CL**, **deO-HS-CL**, **6-deO-HS-CL**, or **2-deO-HS-CL** (5 μg/mL) overnight. Cells were serum-starved the following day with DMEM containing 0.5% FBS for 8 h and then stimulated for 15 m with DMEM containing 0.5% FBS and 10 ng/mL of FGF2. Cells were then quickly washed with 1 mL of ice-cold PBS, scraped, and pelleted. The pellet was lysed and homogenized as described above. Samples were boiled for 5 min with SDS-PAGE loading buffer, resolved by SDS-PAGE on 4-12% Bis-Tris gels, transferred to PVDF membrane, and blocked for 1 h at RT with Odyssey Blocking Buffer (LI-COR Biosciences). Membranes were immunoblotted with a rabbit anti-phospho-ERK monoclonal antibody (1:2000) overnight at 4 °C and then a mouse anti-ERK monoclonal antibody (1:2000) for 1 h at RT. Blots were then incubated with an AF680-conjugated goat anti-mouse IgG antibody (1:10000) and IRDye800-conjugated goat anti-rabbit IgG antibody (1:10000) for 1 h at RT. Membranes were washed three times with TBST for 5 min and then imaged and quantified using an Odyssey scanner (LI-COR Biosciences). Ratios of phospho-ERK to total ERK were calculated for each condition and normalized to an untreated control. Statistical analysis was performed using a one-way ANOVA with post-hoc analysis by Dunnett's test against the untreated control. Data represent the mean ± S.E.M. for three independent experiments (n = 3).

*Neuronal differentiation assays*. Pluripotent mESCs were transfected with the HTP plasmid and HepII treated as described above. The reaction was quenched with 6 mL of complete DMEM, pelleted, and resuspended in 1 mL of neural induction medium (1 part DMEM:F12 (1:1) 1% N-2 supplement, 1% P/S and 1 part Neurobasal, 1% GlutaMAX, 1% B-27, 1% P/S). HepII-treated cells were then added to 0.2% gelatin-coated plates in 2 mL of neural induction medium containing FGF2 (10 ng/mL) and HS-CL derivatives (untreated, **HS-CL**, or **de-HS-CL**; 5 μg/mL). The medium was replaced after 48 h with fresh neural induction medium containing FGF8b (10 ng/mL). At 3, 6, and 9 days after HepII treatment, the cells were subjected to RNA isolation for qRT-PCR and immunocytochemistry analyses.

*Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR).* mESCs were cultured and subjected to neuronal differentiation as described above. At different time points (Day 3, 6, and 9), RNA was extracted from untreated cells and cells treated with HS-CL or de-HS-CL using the RNeasy Mini Kit (Qiagen) per the manufacturer's instructions. RNA concentrations were obtained with a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific), and then the RNA samples were converted to cDNA using the iScript Advanced cDNA Synthesis Kit (Bio-Rad). qRT-PCR was performed using an Applied Biosciences 7300 Real-Time PCR System with primers for GAPDH, SDHA,

NANOG, SOX1, and TUJ1. Primers for amplification of genes of interest are listed below.

```
GAPDH-F: AACAGAAACTCCCACTCTTC
GAPDH-R: CCTGTTGCTGTAGCCGTATT (111 bp amplicon)


SDHA-F:  GCTGGAGAAGAATCGGTTATGA
SDHA-R:  GCATCGACTTCTGCATGTTTAG (97 bp)

NANOG-F: TTTGGAAGCCACTAGGGAAAG
NANOG-R: CCAGATGTTGCGTAAGTCTCATA (115 bp)

SOX1-F:  ACACACACACACACACTCTC
SOX1-R:  CCTCAAGATCTGGTCAGGAATG (101 bp)

TUJ1-F:  CCATTCTGGTGGACTTGGAA
TUJ1-R:  GCACCACTCTGACCAAAGATA (103 bp)
```

The following conditions were used for each 20-μL qRT-PCR reaction: 5 μL of cDNA (10 ng/μL), 1 μL of primer mixture (10 μM forward and 10 μM reverse), 10 μL of 2x SYBR master mix, and 4 μL of $ddH_2O$. The cycle threshold (Ct) values for NANOG, SOX1, and TUJ1 were first normalized against the geometric mean of GAPDH and SDHA at each condition and time point. Each data point was then normalized to the untreated value at day 3 to report relative changes. Statistical analysis was performed for each gene at each time point using a one-way ANOVA with post-hoc analysis by Dunnett's test against the untreated control. Data represent the mean ± S.E.M. from two samples run in triplicate ($n = 2$).

*Immunocytochemistry*. Untreated mESCs and mESCs treated with HS-CL or de-HS-CL were subjected to neuronal differentiation as described above. At 3, 6, and 9

days after HepII treatment, mESCs were fixed with 4% paraformaldehyde in PBS (15 min), rinsed with PBS (2 x 1 mL), permeabilized with a solution containing 1% BSA and 0.1% Triton X-100 (20 min), and incubated with a rabbit anti-NANOG monoclonal antibody (1:1000), a goat anti-SOX1 polyclonal antibody (1:250), or a rabbit anti-TUJ1 monoclonal antibody (1:1000) for 3 h at RT. Cells were washed with PBS (2 x 1 mL) and then incubated with an AF488-conjugated goat anti-rabbit antibody (1:5000), AF568-conjugated donkey anti-goat antibody (1:5000), or AF568-conjugated goat anti-mouse antibody (1:5000), respectively, for 1 h at RT. Cells were imaged using a Zeiss LSM 700 inverted confocal microscope.

## 2.6 References


1. C. D. Rillahan *et al.* Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat Chem Biol.* **2012**, *8*: 661-668.
2. S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi. *In vivo* imaging of membrane-associated glycans in developing zebrafish. *Science.* **2008**, *320*: 664-667.
3. T. Zheng *et al.* Tracking *N*-acetyllactosamine on cell-surface glycans *in vivo*. *Angew Chem Int Ed.* **2011**, *50*: 4113-4118.
4. R. Xie, S. Hong, L. Feng, J. Rong, X. Chen. Cell-selective metabolic glycan labeling based on ligand-targeted liposomes. *J Am Chem Soc.* **2012**, *134*: 9914-9917.

5. D. Rabuka, M. B. Forstner, J. T. Groves, C. R. Bertozzi. Noncovalent cell surface engineering: incorporation of bioactive synthetic glycopolymers into cellular membranes. *J Am Chem Soc*. **2008**, *130*: 5947-5953.

6. B. Belardi, G. P. O'Donoghue, A. W. Smith, J. T. Groves, C. R. Bertozzi. Investigating cell surface galectin-mediated cross-linking on glycoengineered cells. *J Am Chem Soc*. **2012**, *134*: 9549-9552.

7. J. Mauris *et al.* Modulation of ocular surface glycocalyx barrier runction by a galectin-3 *N*-terminal deletion mutant and membrane-anchored synthetic glycopolymers. *PLoS One*. **2013**, *8*: e72304.

8. J. E. Hudak, S. M. Canham, C. R. Bertozzi. Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat Chem Biol*. **2014**, *10*: 69-75.

9. V. P. Torchilin. Recent advances with liposomes as pharmaceutical carriers. *Nat Rev Drug Discovery*. **2005**, *4*: 145-160.

10. W. C. Chen *et al.* Antigen delivery to macrophages using liposomal nanoparticles targeting sialoadhesin/CD169. *PLoS One*. **2012**, *7*: e39039.

11. N. Kawasaki *et al.* Targeted delivery of lipid antigen to macrophages via the CD169/sialoadhesin endocytic pathway induces robust invariant natural killer T cell activation. *Proc Natl Acad Sci USA*. **2013**, *110*: 7826-7831.

12. A. Csiszar *et al.* Novel Fusogenic Liposomes for Fluorescent Cell Labeling and Membrane Modification. *Bioconjugate Chem*. **2010**, *21*: 537-543.

13. D. Dutta, A. Pulsipher, W. Luo, M. N. Yousaf. Synthetic chemoselective rewiring of cell surfaces: generation of three-dimensional tissue structures. *J Am Chem Soc*. **2011**, *133*: 8704-8713.

14. W. Luo, E. W. L. Chan, M. N. Yousaf. Tailored electroactive and quantitative ligand density microarrays applied to stem cell differentiation. *J Am Chem Soc*. **2010**, *132*: 2614-2621.

15. J. G. Hecker, G. O. Berger, K. A. Scarfo, S. Zou, M. H. Nantz. A flexible method for the conjugation of aminooxy ligands to preformed complexes of nucleic acids and lipids. *ChemMedChem*. **2008**, *3*: 1356-1361.

16. L. Yi *et al.* A highly efficient strategy for modification of proteins at the *C*-terminus. *Angew Chem Int Ed*. **2010**, *49*: 9417-9421.

17. J. M. Brown *et al.* A sulfated carbohydrate epitope inhibits axon regeneration after injury. *Proc Natl Acad Sci USA*. **2012**, *109*: 4768-4773.

18. C. J. Rogers *et al.* Elucidating glycosaminoglycan-protein-protein interactions using carbohydrate microarray and computational approaches. *Proc Natl Acad Sci USA*. **2011**, *108*: 9747-9752.

19. E. J. Huang, L. F. Reichardt. Trk receptors: roles in neuronal signal transduction. *Annu Rev Biochem.* **2003**, *72*: 609-642.

20. R. A. A. Smith, K. Meade, C. E. Pickford, R. J. Holley, C. L. R. Merry. Glycosaminoglycans as regulators of stem cell differentiation. *Biochem Soc Trans.* **2011**, *39*: 383-387.

21. D. C. Kraushaar, Y. Yamaguchi, L. Wang. Heparan sulfate is required for embryonic stem cells to exit from self-renewal. *J Biol Chem.* **2010**, *285*: 5907-5916.

22. D. C. Kraushaar *et al.* Heparan sulfate facilitates FGF and BMP signaling to drive mesoderm differentiation of mouse embryonic stem cells. *J Biol Chem*. **2012**, *287*: 22691-22700.

23. M. Forsberg *et al.* Undersulfation of heparan sulfate restricts differentiation potential of mouse embryonic stem cells. *J Biol Chem.* **2012**, *287*: 10853-10862.

24. C. E. Pickford *et al.* Specific Glycosaminoglycans modulate neural specification of mouse embryonic stem cells. *Stem Cells.* **2011**, *29*: 629-640.

25. X. Lin. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development*. **2004**, *131*: 6009-6021.

26. M. L. Huang, R. A. A. Smith, G. W. Trieger, K. Godula. Glycocalyx remodeling with proteoglycan mimetics promotes neural specification in embryonic stem cells. *J Am Chem Soc.* **2014**, *136*: 10565-10568.

27. A. Pulsipher, M. E. Griffin, S. E. Stone, J. M. Brown, L. C. Hsieh-Wilson. Directing neuronal signaling through cell-surface glycan engineering. *J Am Chem Soc*. **2014**, *136*: 6794-6797.

28. G. V. Los *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol.* **2008**, *3*: 373-382.

29. J. Jia *et al.* Development of a multiplex autoantibody test for detection of lung cancer. *PLoS One*. **2014**, *9*: e95444.

30. J. L. Brigham, B. G. K. Perera, D. J. Maly. A hexylchloride-based catch-and-release system for chemical proteomic applications. *ACS Chem Biol.* **2013**, *8*: 691-699.

31. M.-k. So, H. Yao, J. Rao. HaloTag protein-mediated specific labeling of living cells with quantum dots. *Biochem Biophys Res Commun.* **2008**, *374*: 419-423.

32. T. Spivak-Kroizman *et al.* Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation. *Cell.* **1994**, *79*: 1015-1024.

33. J. Schlessinger *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol Cell.* **2000**, *6*: 743-750.

34. F. Lanner, J. Rossant. The role of FGF/Erk signaling in pluripotent cells. *Development*. **2010**, *137*: 3351-3360.

35. M. P. Stavridis, J. S. Lunn, B. J. Collins, K. G. Storey. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development*. **2007**, *134*: 2889-2894.

36. T. Kunath *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*. **2007**, *134*: 2895-2902.

37. J. Kreuger, M. Salmivirta, L. Sturiale, G. Gimenez-Gallego, U. Lindahl. Sequence analysis of heparan sulfate epitopes with graded affinities for fibroblast growth factors 1 and 2. *J Biol Chem.* **2001**, *276*: 30744-30752.

38. N. Sugaya, H. Habuchi, N. Nagai, S. Ashikari-Hada, K. Kimata. 6-*O*-sulfation of heparan sulfate differentially regulates various fibroblast growth factor-dependent signalings in culture. *J Biol Chem.* **2008**, *283*: 10366-10376.

39. C. R. Freed *et al.* Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med.* **2001**, *344*: 710-719.

40. I. Mendez *et al.* Dopamine neurons implanted into people with Parkinson's disease survive without pathology for 14 years. *Nat Med.* **2008**, *14*: 507-509.

41. Y. P. Yan *et al.* Directed differentiation of dopaminergic neuronal subtypes from human embryonic stem cells. *Stem Cells.* **2005**, *23*: 781-790.

42. J. Kim *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc Natl Acad Sci USA*. **2011**, *108*: 7838-7843.

43. G. Clavé *et al.* A novel heterotrifunctional peptide-based cross-linking reagent for facile access to bioconjugates. Applications to peptide fluorescent labelling and immobilisation. *Org Biomol Chem.* **2008**, *6*: 3065-3078.

44. C. G. Glabe, P. K. Harty, S. D. Rosen. Preparation and properties of fluorescent polysaccharides. *Anal Biochem.* **1983**, *130*: 287-294.

45. S. E. Tully *et al.* A chondroitin sulfate small molecule that stimulates neuronal growth. *J Am Chem Soc*. **2004**, *126*: 7736-7737.

46. A. M. Colangelo *et al.* Recombinant human nerve growth factor with a marked activity *in vitro* and *in vivo*. *Proc Natl Acad Sci USA*. **2005**, *102*: 18658-18663.

*Chapter 3*

# Discovery of Orphan Receptor Tie1 and Angiopoietin Ligands Ang1 and Ang4 as Novel GAG-Binding Partners

### 3.1 Abstract

The Tie/Ang signaling axis is necessary for proper vascular development and remodeling. However, the mechanisms that modulate signaling through this receptor tyrosine kinase pathway are relatively unclear. In particular, the role of the orphan receptor Tie1 is highly disputed. Although this protein is required for survival, Tie1 has been found both to inhibit and yet be necessary for Tie2 signaling. While differing expression levels have been put forth as an explanation for its context-specific activity, the lack of known endogenous ligands for Tie1 has severely hampered understanding its molecular mode of action. Here we describe the discovery of orphan receptor Tie1 and angiopoietin ligands Ang1 and Ang4 as novel GAG binding partners. We localize the binding site of GAGs to the *N*-terminal region of Tie1, which may provide structural insights into the importance of this interaction regarding the formation of Tie1-Tie2 heterodimerization. Furthermore, we use our mutagenesis studies to guide the generation of a mouse model that specifically ablates GAG-Tie1 binding *in vivo* for further characterization of the functional outcomes of GAG-Tie1 binding. We also show that GAGs can form a trimeric complex with Ang1/4 and Tie2 using our microarray technology. Finally, we use our HaloTag glycan engineering platform to modify the cell surface of endothelial cells and demonstrate that HS GAGs can potentiate Tie2 signaling in a sulfation-specific manner, providing the first evidence of the involvement of HS GAGs in Tie/Ang signaling and delineating further the integral role of HS GAGs in angiogenesis.

### 3.1 Tie/Ang Signaling in Developing, Mature, and Remodeling Vasculature

Discovered in 1992, the Tie family of proteins is part of the RTK superfamily and consists of two members: Tie1 and Tie2 (Figure 3-1).[1-4] Along with the related angiopoietin (Ang or Angpt) ligands Ang1, Ang2, and Ang4,[5-7] the Tie/Ang signaling pathway is critical for the proper formation and maintenance of the vascular system.[8-10] Initial studies of Tie1 and Tie2 found that constitutive knockout of either receptor led to severe subcutaneous edema and tissue swelling, causing embryonic lethality for Tie2-deficient mice between embryonic days 10.5 and 12.5 (E10.5 - E12.5) and for Tie1-deficient mice starting at E13.5.[11-13] Tie1-deficient embryos that did not die *in utero* died upon birth due to respiratory failure likely caused by fluid accumulation in the lungs. Closer examination of the embryos showed stunted blood vessel networks, providing a possible mechanism for the edemic phenotype. Later studies linked the loss of Tie1 with impaired lymphogenesis via improper valve and collecting vessel development.[14-16] Moreover, Tie1 has been linked to numerous pathologies including tumor angiogenesis and atherosclerosis,[17, 18] and Tie/Ang proteins have been targeted as both anti-angiogenic and anti-sepsis therapeutics,[19, 20] illustrating the widespread importance of Tie/Ang signaling.

The Tie/Ang signaling axis shares some mechanistic similarities with other RTKs. For example, Tie2 is activated by dimerization or multimerization through binding to Ang1 or Ang4, leading to cross-phosphorylation of its intracellular domains and downstream signaling through the PI3K/Akt survival pathway.[9]

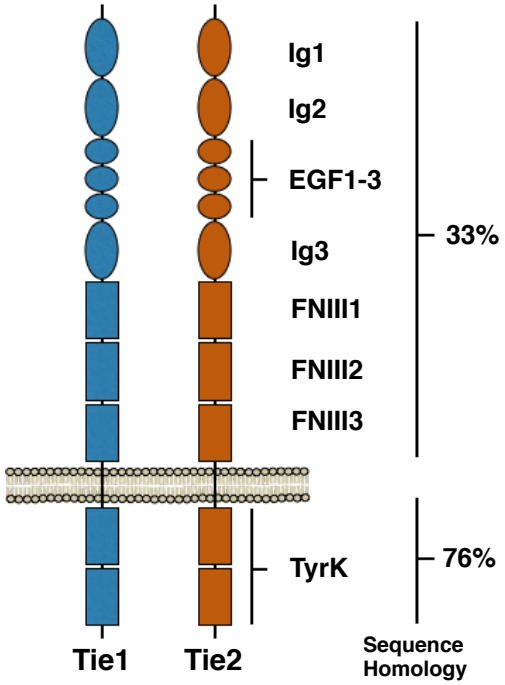



**Figure 3-1. Domain structures of receptor tyrosine kinases Tie1 and Tie2.**

However, Tie2 signaling is different from many other RTKs in that the protein is constantly activated in mature vessels. Ang1 is secreted by pericytes that surround vascular endothelial cells, activating Tie2 to promote cell survival (Figure 3-2). During angiogenesis, pericytes dissociate from the endothelial cells, which then produce Ang2 for autocrine antagonism of Tie2 survival signaling. If new blood vessels form with intact pericyte layers, pro-survival Tie2 signaling through Ang1 resumes. If unsuccessful, endothelial cells undergo apoptosis due to the lack of survival signaling.[9] The effect of Ang4, which appears to be human-specific and functionally distinct from its mouse ortholog Ang3,[7, 21] remains relatively unknown compared to Ang1 and Ang2.

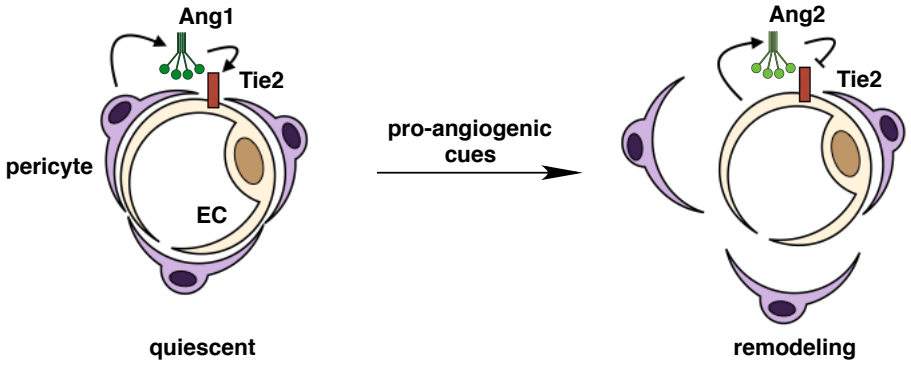



**Figure 3-2. General mechanism of Tie/Ang signaling**. In quiescent vasculature, Ang1 is secreted by pericytes to activate Tie2 on endothelial cells (EC). Upon receiving pro-angiogenic cues like VEGF, pericytes dissociate from ECs, which then secrete Ang2 to antagonize Ang1/Tie2 signaling.

Unlike Tie2, Tie1 is considered an orphan receptor.[9, 22, 23] Without a physiological ligand, its importance in vascular development and remodeling has proven much more difficult to unravel. Tie1 is generally though of as an antagonist of Tie2 signaling through the formation of Tie1/Tie2 heterodimers.[24] Experiments using shRNA-mediated knockdown of Tie1 demonstrated that Tie2 activation by Ang1 is elevated after depletion of Tie1.[25] However, a recent *in vivo* study showed that endothelial deletion of Tie1 in adult mice paradoxically leads to decreased Tie2 activation and downstream signaling,[26] suggesting instead that the Tie1/Tie2 interaction is necessary to promote Ang1-mediated vascular responses. In a separate report, Tie1 was suggested to both sustain and inhibit Tie2 function in different physical locations of newly forming blood vessels through changes in Tie1 expression levels.[22] However, a separate way to account for these differences in activity could be the presence of an undiscovered ligand for Tie1. GAGs are prevalent on the cell surface of most cells including the

endothelium and have previously been shown to bind both to RTKs and their respective ligands,[27-36] making them an ideal candidate for binding to not only Tie1 but also Ang ligands. Moreover, given the prevalence of GAG-RTK interactions in other systems involved in angiogenesis including VEGF/VEGFR,[29, 37] it is not difficult to envision the involvement of GAGs in the Tie/Ang signaling axis.

### 3.2 GAGs Are Physiological Ligands for Orphan Receptor Tie1

#### 3.2.1 ELISA, Carbohydrate Microarray, and Surface Plasmon Resonance

We first examined whether Tie1 and Tie2 were receptors for GAGs using our carbohydrate enzyme-linked immunosorbent assay (ELISA) method.[38] Here, commercially available Tie1- and Tie2-Fc conjugates were immobilized onto Protein A/G coated plates. Biotinylated CS-E and trisulfated hep/HS (HS) were then incubated with the immobilized proteins for 3 h at RT, and bound GAGs were visualized using horseradish peroxidase (HRP)-conjugated streptavidin. Excitingly, we found that both CS-E and HS bound strongly to Tie1, both with nanomolar binding affinity (Figure 3-3; CS-E: 19.9 ± 5.8 nM, HS: 6.16 ± 0.37 nM). These data represent the first evidence of a physiological ligand for the Tie1 receptor since its discovery 25 years ago.[2] Furthermore, we found that Tie2 did not exhibit any measurable binding to either GAG structure, illustrating the specificity of the Tie1-GAG interaction.

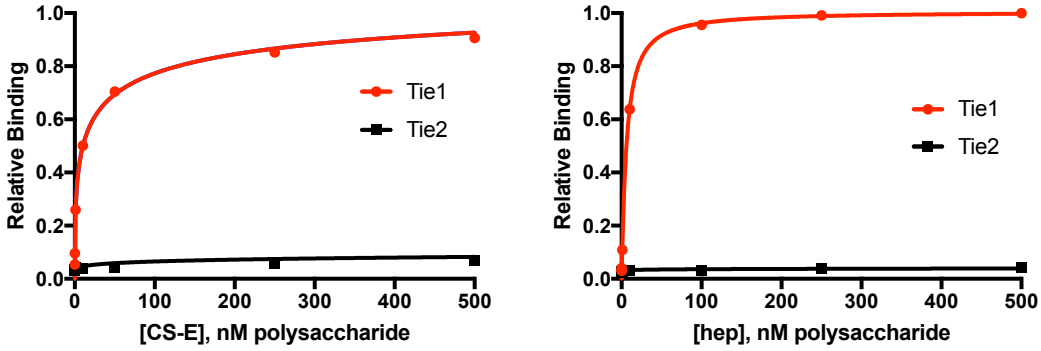



**Figure 3-3. Carbohydrate enzyme-linked immunosorbent assay for Tie1-GAG binding**. Immobilized Tie1-Fc or Tie2-Fc were probed with biotinylated CS-E or biotinylated HS.

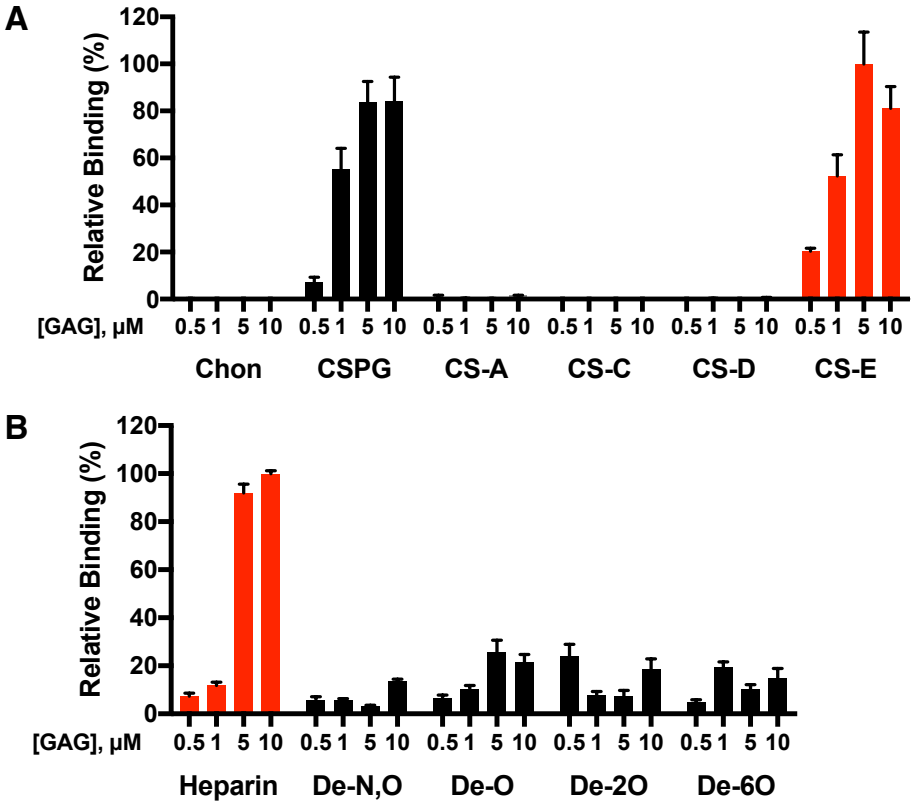
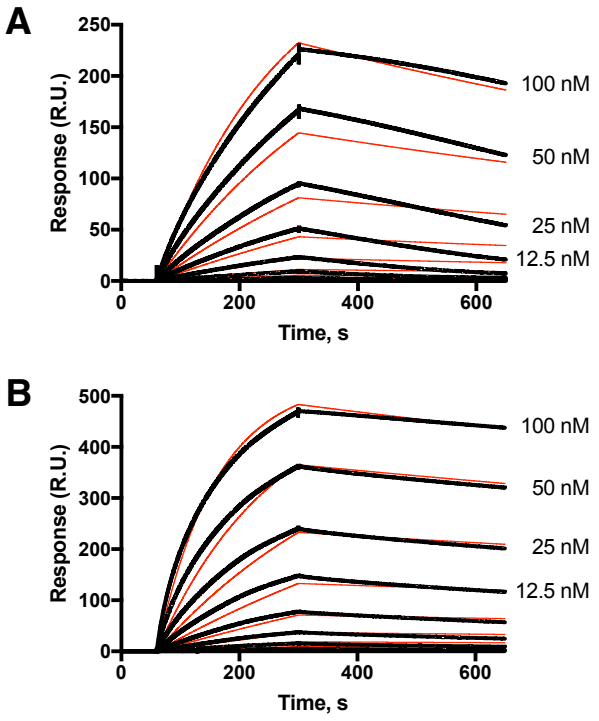



**Figure 3-4. Carbohydrate microarrays to examine sulfation specificity of Tie1-GAG binding**. Tie1-Fc was incubated with printed (A) CS and (B) HS microarray. Bound Tie1-Fc is visualized with an anti-human Fc antibody.




| CS-E | |
|---|---|
| $k_{on}$ | 4.29 x $10^4$ $M^{-1}$ $s^{-1}$ |
| $k_{off}$ | 6.31 x $10^{-4}$ $s^{-1}$ |
| $K_D$ | 14.7 nM |

| HS | |
|---|---|
| $k_{on}$ | 9.60 x $10^4$ $M^{-1}$ $s^{-1}$ |
| $k_{off}$ | 3.01 x $10^{-4}$ $s^{-1}$ |
| $K_D$ | 3.14 nM |

**Figure 3-5. Carbohydrate surface plasmon resonance to examine kinetics of Tie1-GAG binding**. Biotinylated (A) CS-E or (B) HS were immobilized onto a CM5 sensor chip, and Tie1-Fc was flowed over at different concentrations. Kinetic parameters were obtained by fitting to a one-to-one Langmuir binding model.

We next assessed the sulfation specificity of the Tie1-GAG interactions using carbohydrate microarrays.[33] Tie1- and Tie2-Fc conjugates were incubated on glass slides with different CS and HS epitopes electrostatically immobilized on the surface. Bound proteins were detected using a fluorescently tagged anti-human Fc antibody. As expected, Tie1 showed robust binding with the CS-E (Figure 3-4a) and HS (Figure 3-4b) epitopes with minimal binding to all other sulfation patterns.

Finally, we further quantified the Tie1-GAG interactions using surface plasmon resonance (SPR) analysis.[38] In this experiment, SPR CM5 sensor chips

covalently modified with carboxymethylated dextran were functionalized with streptavidin to allow for immobilization of biotinylated CS-E or HS. Solutions of Tie1-Fc at different concentrations were then flowed over the modified chip followed by buffer only, and the response rate was measured over time. We were pleased to see that both CS-E (Figure 3-5a) and HS (Figure 3-5b) gave dissociation constants that were similar to those observed via ELISA (CS-E: 14.7 nM; HS: 3.16 nM). Both carbohydrates also demonstrated similar binding kinetics with $k_{off}$ values around $10^{-4}$ $s^{-1}$, suggesting a theoretical half-life for the complexes of roughly 18 min (CS-E) and 38 min (HS). These off rates are smaller than those reported for known hep/HS binding proteins VEGFR1 and neuropilin 1 (Nrp1),[37] implicating the physiological relevance of these binding events in vascular tissue.

### 3.2.2 Identification of the Tie1-GAG Binding Site

Having shown that Tie1 binds to GAGs through numerous methods, we aimed to decipher the exact location of the Tie1-GAG binding site to help provide structural clues on the importance of these novel interactions. No known consensus primary sequence or secondary structure has been found as a predictive measure for GAG-protein binding. As expected, GAG binding in many cases requires numerous surface-exposed positively charged Arg and Lys residues, which can occur on any number of domain structures. Tie1 contains both immunoglobulin-like (Ig) domains, which interact with hep/HS in FGFR1,[27] and fibronectin type III (FNIII) domains, which bind to CS-E in EphA4 and -B3 receptors.[39, 40]

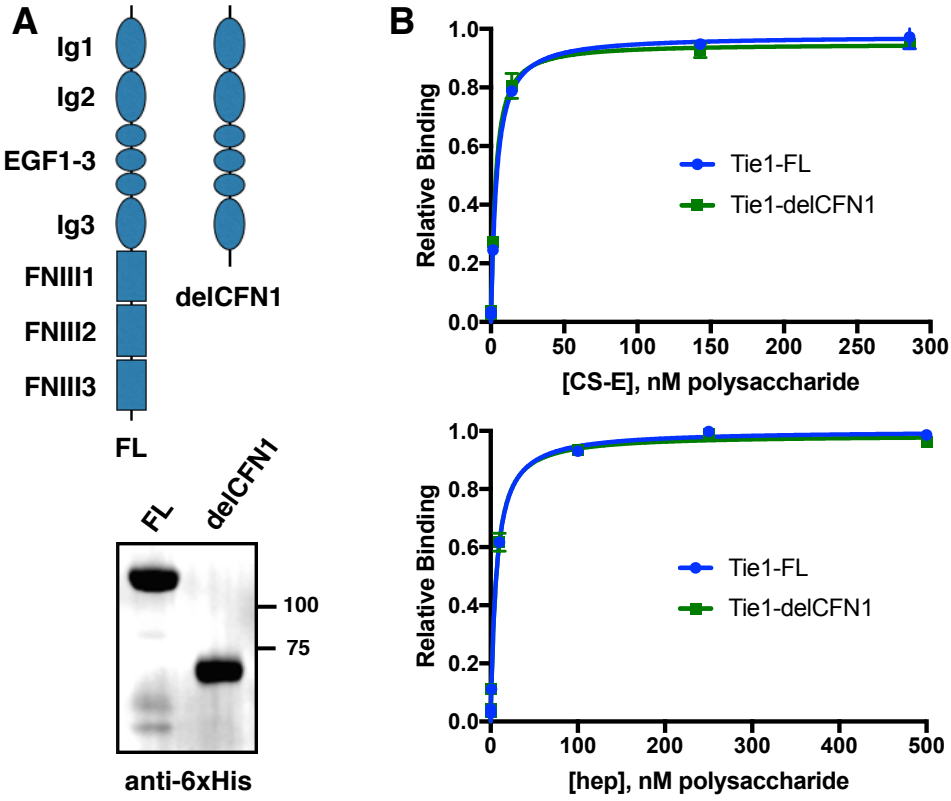



**Figure 3-6. Tie1 truncations to determine Tie1-GAG binding domain**. (A) Full length (FL) and truncated (delCFN1) Tie1-Fc were expressed and purified. (B) Expressed Tie1-Fc constructs were used in carbohydrate ELISA as described earlier. Both constructs maintain binding to GAGs, suggesting that binding occurs in the *N*-terminal domains of Tie1.

As a first step, we produced Tie1-Fc conjugates that contained, the full-length ectodomain (Tie1-FL-Fc) or a *C*-terminal truncation that eliminated the three FNIII domains (Tie1-delCFN-Fc). The two constructs were expressed as secreted proteins and purified by immobilized metal affinity chromatography (IMAC) using a *C*-terminal 6xHis tag (Figure 3-6a). These constructs were then used in the ELISA format described above. We found that both constructs showed binding with CS-E and HS, suggesting that the Tie1-GAG binding site was contained within the *N*-terminal Ig1/2 or EGF1-3 domains (Figure 3-6b). We

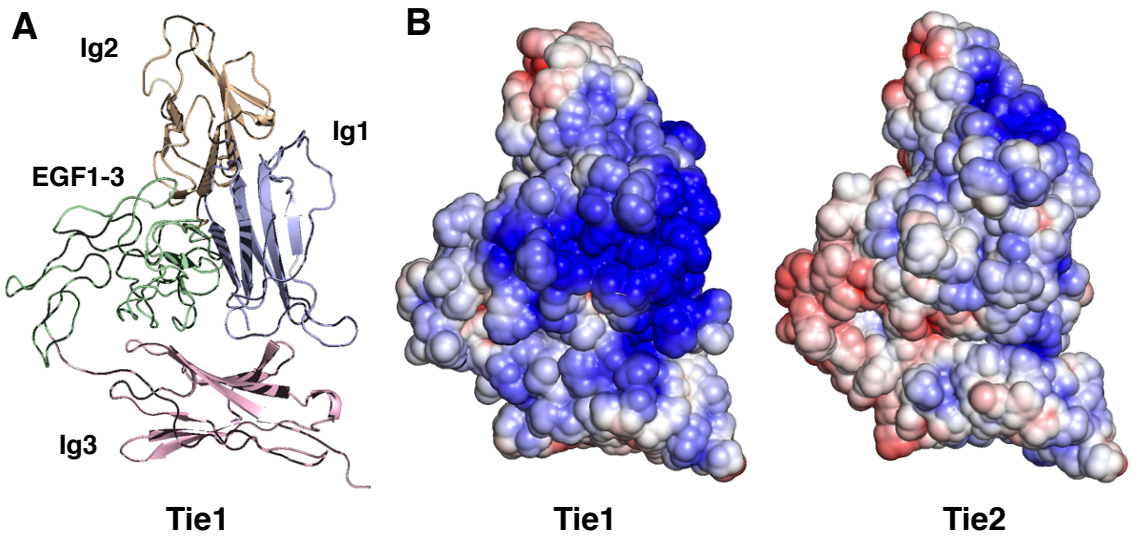



**Figure 3-7. Homology model of Tie1**. (A) Ribbon structure of Tie1 homology model built by SWISS-MODEL using Tie2 crystal structure. (B) Surface electrostatic potential maps of Tie1 and Tie2.

tried to express complementary *N*-terminal truncations of the protein, but none of these constructs expressed at high enough quantities to be used in the ELISA. This result is likely due to the predicted globular structure formed by the *N*-terminal domains in the final folded protein.

Because we could not isolate the Tie1-GAG binding site further by truncation, we turned to computational docking of GAG structures with Tie1 to predict key residues for the interaction.[33] This approach has been successfully used to define likely regions of GAG binding on a number of proteins, including the malarial protein VAR2CSA, tumor necrosis factor α (TNFα), and Ephs A4 and B3.[33, 39, 40] To conduct docking, we first produced a homology model of Tie1 based on the solved Tie2 crystal structure[41] using the SWISS-MODEL program (Figure 3-7a).[42]

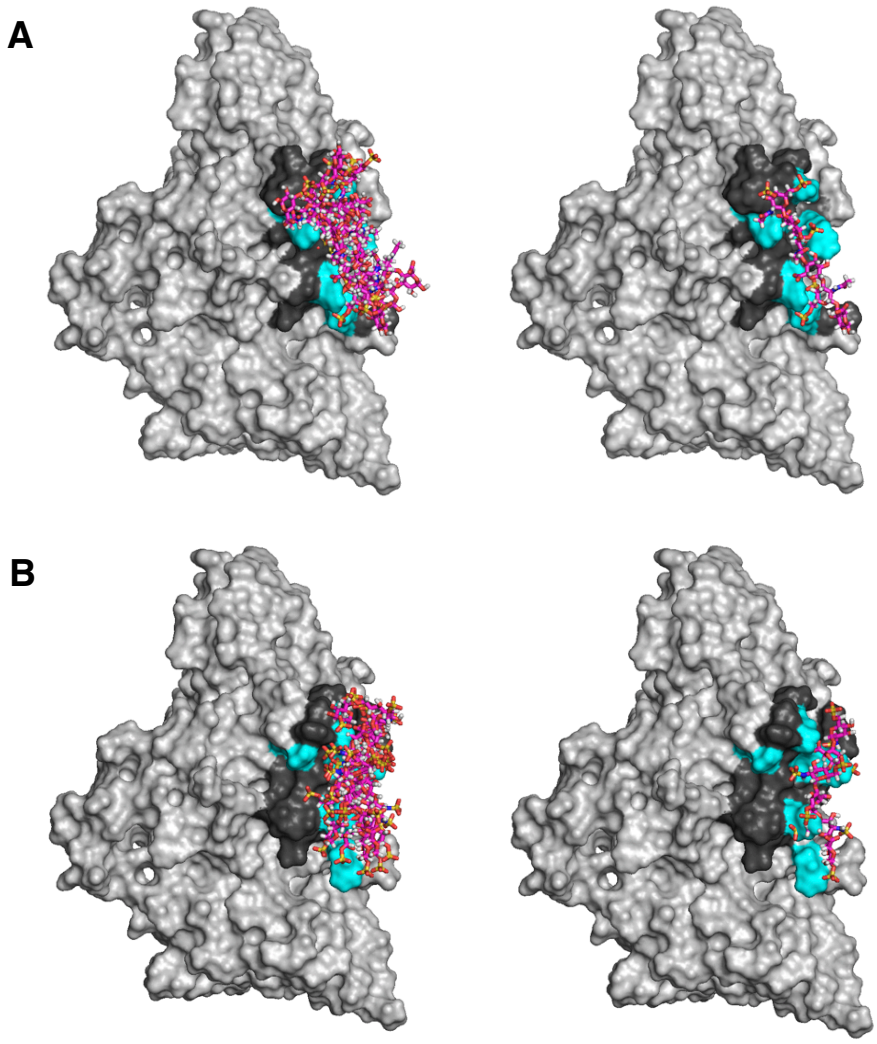



**Figure 3-8. Docking of CS-E and HS hexasaccharides to Tie1**. Top 10 and top 1 binding poses of (A) CS-E and (B) HS onto the Tie1 homology model.

Tie2 shares the same domain structure as Tie1 along with 39% total amino acid identity. Upon inspection of the electrostatic potential maps of the Tie1 and Tie2 structures (Figure 3-7b), the Tie1 model displays a large electropositive region encompassing parts of the Ig1 and three EGF domains. This electropositivity is severely attenuated in the Tie2 structure, providing evidence for the specificity of

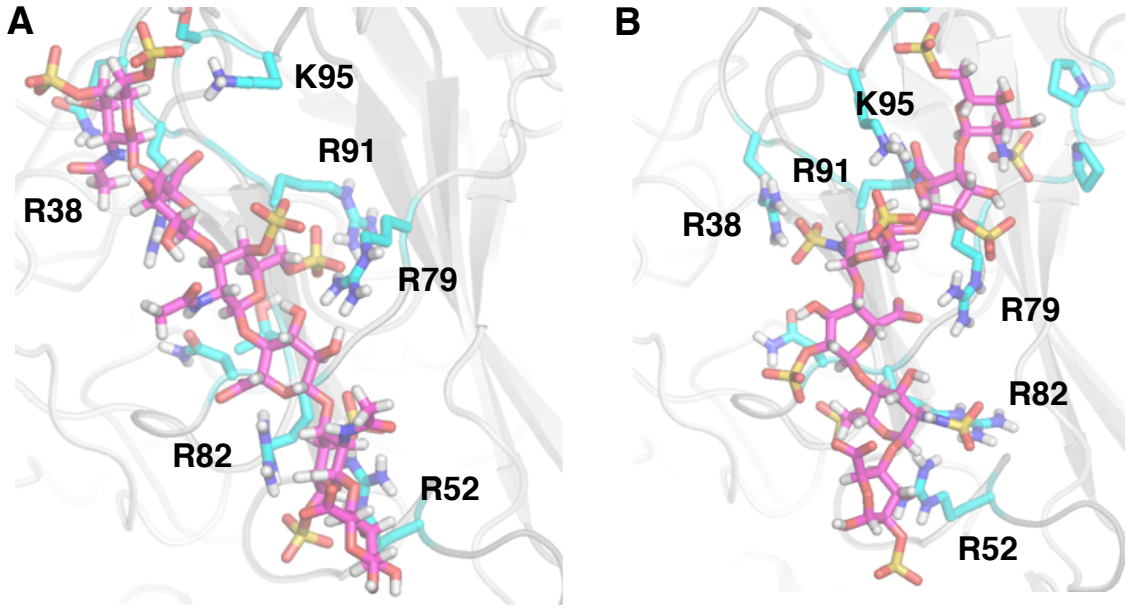



**Figure 3-9. Amino acids in putative GAG binding site of Tie1**. Positively charged amino acids surrounding the top binding pose for (A) CS-E and (B) HS.

GAG binding to Tie1. Interestingly, this region of Tie1 had previously been suggested to interact with a complementary electronegative face of Tie2.[25]

With a model for Tie1 in hand, we then docked hexasaccharide structures of CS-E and HS to the predicted structure and ranked poses based on the calculated binding energies. The top 10 poses of both CS-E and HS clustered in one portion of the large electropositive region centered on the Ig1 domain (Figure 3-8). Upon closer inspection, the top ranked pose of both docked GAG structures showed close association with six positively charged residues: Arg-38, Arg-52, Arg-79, Arg-82, Arg-91, and Arg-95 (Figure 3-9). Using genomic alignment, we found that four of these six residues (Arg-38, Arg-52, Arg-82, Arg-91) were highly conserved


**Figure 3-10. Sequence alignment of *N*-terminal Tie1 gene.** Predicted binding site amino acids are bolded with conserved amino acids highlighted.

throughout all mammals (Figure 3-10), with Lys-95 showing moderate conservation as well.

We turned to the ELISA format to test the accuracy of our computational predictions. Wild-type Tie1-Fc (Tie1-WT) along with a mutant containing all six residues mutated to Ala (Tie1-6A) and a second mutant containing R38A and R82A (Tie1-2A) were cloned and expressed using the secretion system previously described. Importantly, both mutants expressed well, suggesting that the mutations did not affect protein folding as seen for the *N*-terminal truncations. We were pleased to see that binding of both biotinylated CS-E and HS was severely attenuated after mutation (Figure 3-11), confirming that these residues were critical for the Tie1-GAG interaction. We attempted the same experiment using single mutations, but neither showed complete loss of Tie1

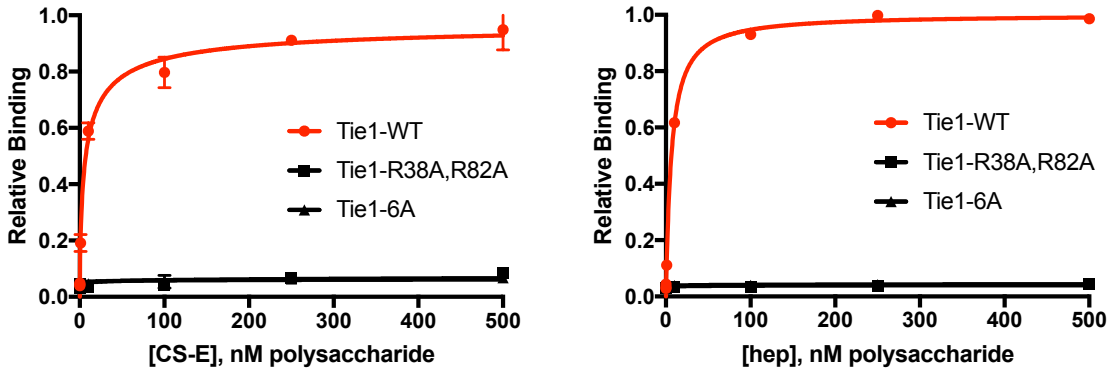



**Figure 3-11. Carbohydrate ELISA of Tie1 mutant constructs**. FL, wild-type (WT), 6A, and 2A mutants were expressed, purified, and used in carbohydrate ELISA binding assays as previously described. Both mutant constructs completely lost binding to GAGs, corroborating the computational results.

binding, suggesting that the double mutation was the minimal perturbation necessary to disrupt the Tie1-GAG interaction.

### 3.2.3 Contextualization of Tie1-GAG Binding

Although up to this point we had tested Tie1 binding to GAGs using both CS-E and HS, we desired to establish which GAG structures would be the most relevant for *in vivo* binding. Tie1 is expressed universally on endothelial cells during late development,[43] after which expression levels are strongly downregulated in the adult except in a small subset of remodeling endothelial cells.[22] Tie1 expression is also stimulated during angiogenic processes such as wound healing and tumor formation and regions of nonlaminar flow like branch points of vessels.[17, 18] Therefore, examining the GAGs produced by endothelial cell lines would provide helpful information to decipher the relative importance of CS-E and HS in Tie1 binding.

Previously, methods to quantify the absolute amounts of different GAG sulfation patterns on the cell surface, termed glycosaminoglycanomics, have found that human umbilical vein endothelial cells (HUVECs) are dominated by HS.[44] In fact, HUVECs, which are used as a model system for systemic endothelial cells, were found to contain the highest proportion of HS and the HS epitope out of all cell lines tested. HS was found at nearly a 10:1 w/w ratio compared to CS. Moreover, CS-E was not detected in HUVECs in this analysis, strongly suggesting that HS was the physiologically relevant GAG for Tie1 binding at least in systemic endothelial cells.

We confirmed these finding using a cell binding assay (Figure 3-12). HUVECs were grown to confluency, treated with or without a mixture of heparinase I and III and fixed. Cells were then incubated with Tie1-WT-Fc or Tie1-2A-Fc, and binding of the constructs was detected using a fluorescently tagged anti-human Fc antibody. Importantly, we saw that Tie1-WT binding was significantly decreased after treatment of the cells with heparinase I/III, with all remaining Tie1-WT staining colocalizing with residual HS signal as detected by an anti-HS antibody. Moreover, binding of Tie1-2A was not observed even with cell surface HS intact. Together, these results suggest that cell surface HS is necessary for Tie1 binding through the predicted Tie1-GAG binding interface and that loss of cell surface HS or the residues necessary for this interaction is sufficient to prevent Tie1 binding.

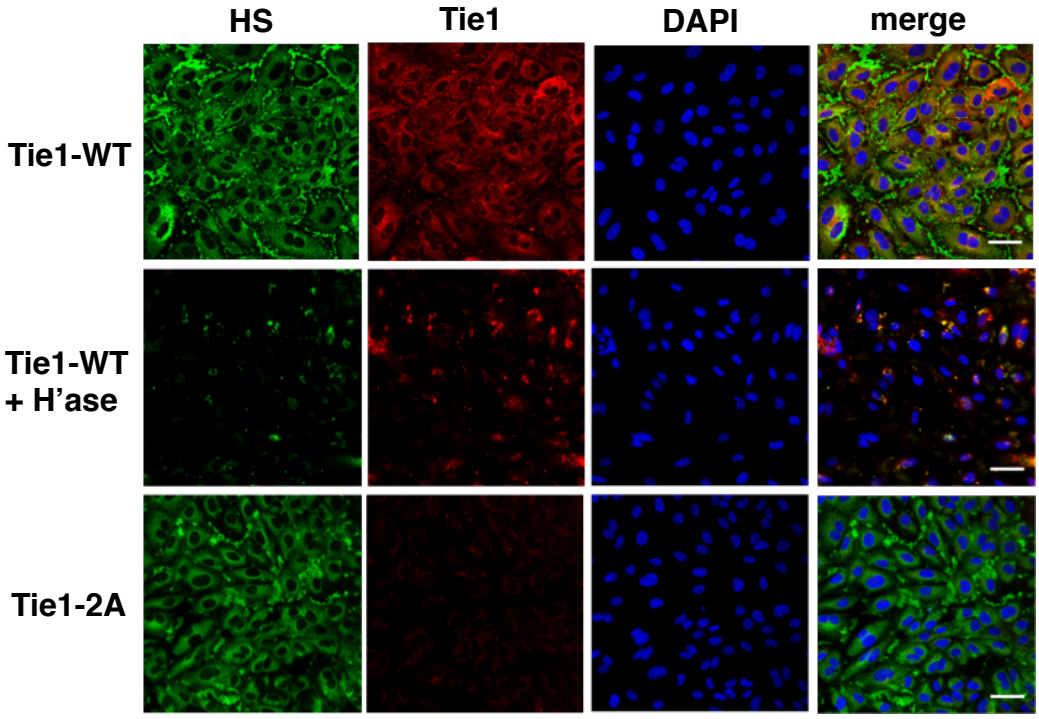



**Figure 3-12. HS-dependent Tie1 binding to endothelial cells**. Untreated or heparinase-treated EA.hy926 cells were probed with Tie1-WT and visualized with an anti-human Fc antibody (red). In parallel, cells were probed using an anti-HS antibody (green) and DAPI (blue). Robust binding was observed for untreated cells, which was lost upon heparinase treatment. Tie1-2A showed no binding to cells, suggesting that binding to HS is necessary to cell surface binding.

### 3.2.4 Towards a Functional Role of Tie-GAG Binding

To further characterize the importance of the Tie1-HS binding interaction, we turned to cellular assays. Unfortunately, Tie1 presents a number of unique challenges to understand its functional role. For example, the most common method to examine the functional output of an RTK is through phosphorylation assays.[45] Although Tie1 contains an intracellular tyrosine kinase domain, evidence of Tie1 homodimeric activation and cross-phosphorylation has not been observed as seen for other RTKs,[46, 47] precluding a commonly used, facile readout for RTK

activity. Early reports using chimeric proteins containing the intracellular Tie1 kinase domain fused to extracellular domain of either tropomyosin-related kinase A (TrkA) or colony stimulating factor 1 (c-fms) receptor provided conflicting results over whether "forced" homodimerization by stimulation with the canonical ligands of TrkA or c-fms could induce cross-phosphorylation,[48, 49] further obfuscating the relevance of Tie1 kinase activity. Conversely, Tie2 activation by angiopoietin ligands can lead to Tie1 phosphorylation through heterodimerization;[46, 47] however, this approach introduces additional HS-binding partners to the experiment (see Section 3.4), preventing the straightforward delineation of the role of HS binding on one protein within the larger system.

To circumvent these concerns, we focused on examining the transient association of Tie1 with other cell surface receptors. The Tie1/Tie2 heterodimer has been documented previously and is the basis of Tie2-induced phosphorylation of Tie1;[46, 47] however, this interaction has been observed using standard co-immunoprecipitation techniques. Instead, cross-linking of surface receptors has been used to preserve the interaction for co-immunoprecipitation.[46, 50, 51] More recently, integrin $\alpha_5\beta_1$ has also been implicated to associate with Tie1 and Tie2 at the cell surface,[51] and its removal prevents Tie2 signaling.[26] As proof of principle, we used the immortalized endothelial cell line EA.hy926 and treated the cells with the cell-impermeable cross-linking agent 3,3'-dithiobis(sulfosuccinimidyl propionate) (DTSSP). Cells were then lysed and immunoprecipitated for Tie1. Cells treated with DTSSP showed robust co-immunoprecipitation of Tie2 as well

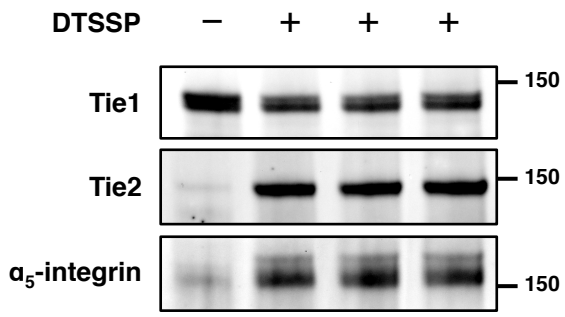



**Figure 3-13. Tie1 association with Tie2 and $\alpha_5$-integrin.** EA.hy926 cells were treated with cross-linker DTSSP, lysed, and immunoprecipitated for Tie1. Samples were analyzed by Western blotting.

as integrin $\alpha_5$ (Figure 3-13). In the near future, we plan to expand this assay out with heparinase treatment and cell surface glycan engineering (see Section 3.4.3).

As an alternative approach to observe Tie1/Tie2 heterodimers, we have also employed proximity ligation assays (PLA).[52] This assay is similar in format to standard immunocytochemistry (ICC) techniques where two proteins are probed for with separate primary antibodies. The method diverges from ICC in that the secondary antibodies are modified with single stranded DNA oligonucleotides rather than fluorescent dyes. Through subsequent annealing and ligation steps, a substrate is formed for rolling circle amplification, at which point fluorescently labeled nucleotides are incorporated into the growing DNA strand. These fluorescent oligonucleotides can then be detected by standard immunofluorescence imaging. Importantly, this procedure will produce punctate staining only at locations where both primary antibodies bound, providing a signal only where the two target proteins are in close proximity with one another. We optimized this assay for the Tie1/Tie2 heterocomplex using two antibodies

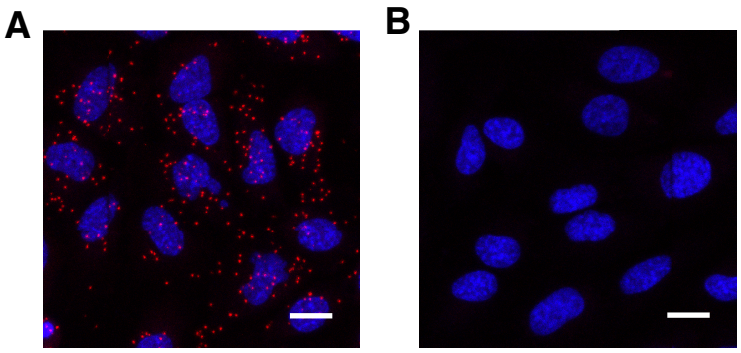



**Figure 3-14. Direct observation of Tie1/Tie2 heterocomplexes by PLA.** EA.hy926 cells were fixed, immunostained with (A) anti-Tie1 and anti-Tie2 antibodies or (B) anti-Tie1 antibody alone, stained with the PLA staining protocol, and visualized by microscopy. Tie1/Tie2 heterocomplexes are revealed as punctate red staining and can be normalized to cell count by DAPI (blue) nuclear staining. Scale bar = 20 μm.

specific for the extracellular domains of each protein (Figure 3-14). Briefly, HUVECs or EA.hy926 cells were fixed, blocked, and probed with the appropriate Tie1 and Tie2 antibodies. Cells were then subjected to PLA procedures according to the manufacturer's specifications. Cells were then co-stained with DAPI and visualized by confocal microscopy. We were pleased to observe the formation of punctate staining at the cell surface when both antibodies were used, whereas a control sample lacking one of the two primary antibodies showed minimal background. As before, we will apply both heparinase and glycan engineering techniques to this assay to further delineate the effect of HS on Tie1/Tie2 heterodimerization.

### 3.2.5 Generation of a Tie1-GAG-Binding Deficient Mouse Model

Further examination of the Tie1-GAG binding event requires the use of functional systems. Here again, the Tie/Ang signaling axis provides a number of

unique challenges. Unlike many other functional outcomes of RTKs such as growth promotion or inhibition, the outcomes of the Tie/Ang pathway such as vessel development, maturation, and remodeling are not easily testable *in vitro*. The closest related *in vitro* assay, measuring apoptosis in response to serum and growth factor starvation, has been used infrequently with angiopoietin ligand stimulation;[53] however, very few other reported *in vitro* assays exist that can directly examine the formation and maintenance of vascular tissue. Instead, many researchers have turned to *in vivo* models to test Tie/Ang function.

The lethality of constitutive Tie1 and Tie2 knockout models prevented further examination of Tie1/2 function in adult animals.[11-13] This led to the development of floxed Tie1 animals crossed with mice expressing cre recombinase under endothelial-specific, tamoxifen-dependent promoters (such as the *Cdh5-creERT2* and *Tal1/SCL-creERT* lines) to provide adult mice with endothelial-specific Tie1 knockout.[17, 26] Although these mice have been invaluable for the study of Tie1 function *in vivo*, they do not provide a straightforward way to ascertain the importance of the Tie1-GAG interaction specifically. Similarly, a number of different HS biosynthesis conditional knockout animals exist, with Ndst1 being commonly used since animals have much lower HS sulfation in the targeted tissue.[54, 55] However, these models also suffer from questions of specificity for our purposes. HS binds to a number of angiogenic ligands and receptors; thus, understanding the ramifications of the Tie1-GAG interaction caused by a global decrease in overall HS sulfation (and thereby overall HS/protein binding) may be

also difficult to parse. Therefore, we turned our attention to a mouse model that could specifically disrupt the Tie1-GAG interaction using insights gained from mutational studies.

To achieve this selective mouse model, we chose to pursue a novel approach using gene editing with the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 platform.[56-58] Here, the Cas9 nuclease protein is directed to specific regions of the genome by a complementary single stranded RNA sequence known as a guide RNA (gRNA). Once localized to the sequence, Cas9 induces a double strand break within the complementary genomic DNA (gDNA) sequence. In the absence of a repair template, endogenous machinery of the cell undergoes non-homologous end joining (NHEJ), thereby producing random insertions and deletions (indels). Alternatively, a repair template is added along with the gRNAs and Cas9. Here, homology-directed repair (HDR) machinery can compete with NHEJ to incorporate the repair template at the DSB site. This approach has been used successfully to incorporate a variety of site mutations, small reporter tags, and fluorescent protein fusions onto endogenously expressed proteins.[59] Importantly, this approach provides a number of benefits compared to other methods like lentiviral delivery or transient expression in that the incorporation event is highly controlled and will not interrupt other genes at random, expression levels will be controlled by the endogenous promoter and subjected to native regulation, and endogenous proteins do not need to be silenced to replace with a mutated sequence. Furthermore, CRISPR/Cas9

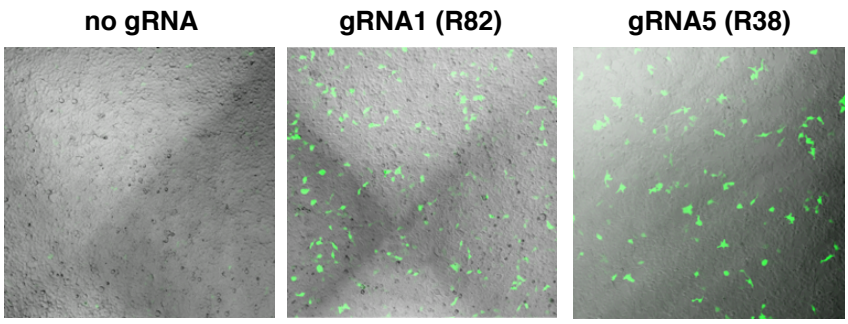



**Figure 3-15. Validation of guide RNAs in HEK-293T cells.** HEK-293T cells were transfected with the pCAG-EG(Tie1)FP plasmid plus empty pX459 (no gRNA) or pX459 containing gRNA1 or gRNA5. gRNA targeting efficiency was measured by the formation of $GFP^+$ cells.

transgenic animal production does not require the extensive screening of stem cell clones like previous transgenic methods.[60] Finally, we believe that this approach will be a powerful new tool for glycobiology, allowing a clear-cut method to understand the importance of a single GAG-receptor binding event without globally affecting other carbohydrate-protein interactions.

We decided to target the R38A and R82A mutations for genomic incorporation. The efficiency of mutation incorporation by HDR decreases rapidly more than 10-20 base pairs (bp) away from the DSB site.[61] Because the codons for these residues are roughly 130 bp away from one another, we decided to use a dual gRNA approach that would cut near each of the targeted mutation sites. gRNAs were designed using the online CHOPCHOP computational tool for the sequences flanking each codon, and four to five gRNAs for each region were selected based on predicted cutting efficiency and minimal off-target homology. These gRNAs were cloned into pX459 gRNA/Cas9 vector to test their cutting efficiency. We chose to use a recently described activity assay for gRNA

screening in which the genomic region of interest is cloned within a broken GFP construct (pCAG-EGxxFP).[62] If cut, the vector undergoes self-mediated HDR using homology regions of GFP flanking the inserted gDNA sequence to produce a functional GFP construct. The Tie1 genomic region of interest was amplified from murine gDNA and inserted into the pCAG-EGxxFP construct to produce pCAG-EG(Tie1)FP. pX459 vectors containing each gRNA and pCAG-EG(Tie1)FP were co-transfected into human embryonic kidney (HEK) 293T cells, which were then visualized after 48 h (Figure 3-15). gRNAs that produced the highest percentage of $GFP^+$ cells were chosen for *in vivo* experiments.

Next, the repair template was designed. Two types of repair templates have been used previously: single stranded oligonucleotides (ssODNs) or double stranded DNA (dsDNA).[59] ssODNs generally are upwards of 200 nucleotides (nt) in length, with roughly 60 nt of homologous sequence on either side of the inserted sequence. Conversely, dsDNA templates are less effective for HDR and require hundreds of bp of homologous sequence for efficient incorporation. Because we aimed to incorporate mutations roughly 130 bp apart, we tried both approaches. Two ssODN constructs were produced containing one of the two mutations flanked by roughly 60 nt of homologous sequence. The dsDNA construct was designed to contain the same central 130-bp sequence flanked by roughly 550 bp of homologous sequence. Within the region surrounding the Arg-38 and Arg-82 codons, a number of mutations were made. First, the two codons were mutated to Ala codons. Second, a silent mutation was incorporated into the

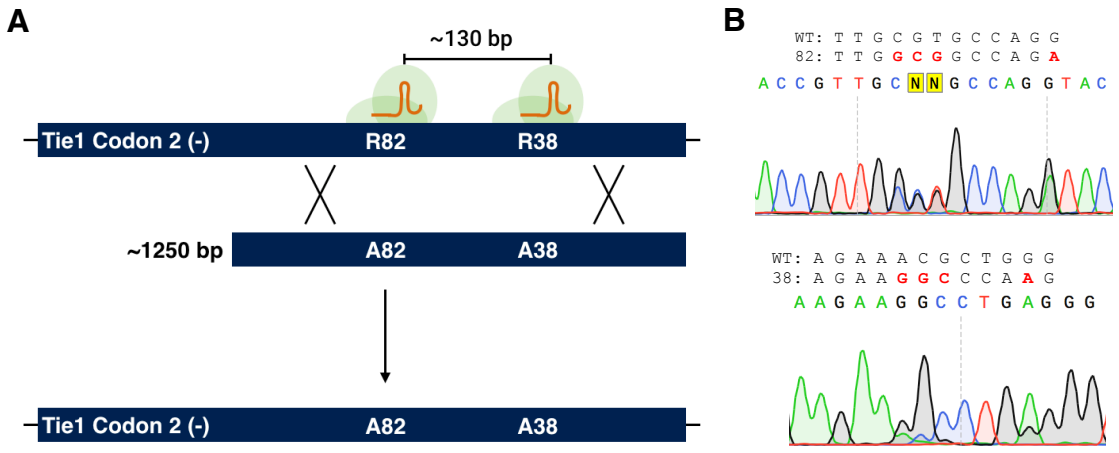



**Figure 3-16. CRISPR/Cas9-mediated *in vivo* Tie1 mutations.** (A) Schematic of CRISPR/Cas9 strategy to introduce R38A and R82A mutations *in vivo*. Two gRNAs specific to regions around each codon of interest along with a dsDNA donor will be used to introduce the mutations. (B) Sequencing results from successful heterozygous mutations in B6SJL-F1/J blastocysts.

protospacer motif (PAM) for each gRNA. Cas9 recognition of gDNA requires an NGG motif known as the PAM immediately at the 3'-end of the gRNA complementary sequence. Loss of the NGG seqeunce prevents Cas9 from cutting the sequence again after the mutation has been incorporated. Finally, silent mutations were added at each site to incorporate restriction enzyme cut sites to allow for facile genotyping. Both the ssODN and ssDNA constructs were commercially purchased along with the gRNA sequence and Cas9 protein.

To test efficiency *in vivo*, the two gRNA/Cas9 complexes were formed *in vitro* and pronuclearly injected into fertilized B6SJL-F1/J murine zygotes at 20 ng/μL each along with either the pair of ssODNs or dsDNA construct at 20 or 10 ng/μL, respectively, in 10 mM Tris pH 8.0, 1 mM EDTA. B6SJL-F1/J mice (the F1 generation crossing of a C57BL/6 female and a SJL/J male) were chosen because

of their relatively large litter sizes and hybrid vigor compared to inbred lines. Zygotes were allowed to develop to the blastocyst phase, at which point they were subjected to genotyping to test for incorporation of the desired mutations. Genotyping was achieved by harvesting gDNA from the blastocysts followed by PCR amplification of the desired sequence and Sanger sequencing (Figure 3-16). In both cases, the majority of the zygotes proceeded to the blastocyst stage (ssODN: 15/18; dsDNA: 15/21), suggesting that the injection solutions were nontoxic. Interestingly, sequencing results showed that the majority of the blastocysts resulting from ssODN injection showed random indels, with no sample showing proper incorporation. This was unexpected as ssODN repair templates had previously been described as highly efficient. We were happy to see that 4/15 blastocysts from dsDNA injection showed proper incorporation of the mutations at both of the sites (Figure 3-16b). We chose this method to proceed to large-scale injection and implantation.

For large-scale injection and implantation, roughly 80 fertilized zygotes were harvested from the oviducts of superovulated, mated B6SJL/J mice, injected with the gRNA/Cas9/dsDNA mixture described above, allowed to mature to the blastocyst stage, and then implanted at ~20 zygotes per surrogate mother. This was repeated three times, which resulted in seven litters containing 19 total pups. The litter sizes were small, but this is expected since loss of Tie1, which could be caused by indels from competing NHEJ processes, is embryonic lethal. Unfortunately, none of the animals possessed the desired genotype. The

experiment was repeated using a higher concentration of dsDNA (20 ng/μL) and incubation of the zygotes with the NHEJ inhibitor SCR7.[63] Here, four litters containing a total of 15 pups were obtained. Excitingly, three pups were heterozygous for the desired mutations, with the other Tie1 allele containing small in-frame deletions. Two of the pups were produced from zygotes incubated with SCR7, whereas one was derived from untreated zygotes. The SCR7-treated pups contained 6- and 9-bp deletions, and the untreated pup contained a larger in-frame deletion of 90 bp, providing anecdotal evidence of the efficacy of SCR7 in inhibiting at least large NHEJ events. The now adult mice are currently set up in breeding pairs with C57BL/6 mice to expand the colony for future homozygous crossing and phenotypic analysis.

### 3.3 HS GAGs Bind to Angiopoietin Ligands and Potentiate Tie2 Signaling

#### 3.3.1 ELISA

In addition to the Tie receptors, we also examined whether the related angiopoietin ligands Ang1, -2, and -4 could also interact with HS. Previously, Ang1 but not Ang2 has been shown to associate with the extracellular matrix (ECM) via its linker peptide region;[64] however, the purified Ang1 constructs used therein did not bind to heparin or soluble Tie2-Fc. In a separate report, Ang3, the functionally distinct murine ortholog to human Ang4, was shown to be anchored to the extracellular matrix through HSPGs like perlecan,[65] raising the possibility of Ang4 association with HS. To test binding, we used a modified version of the

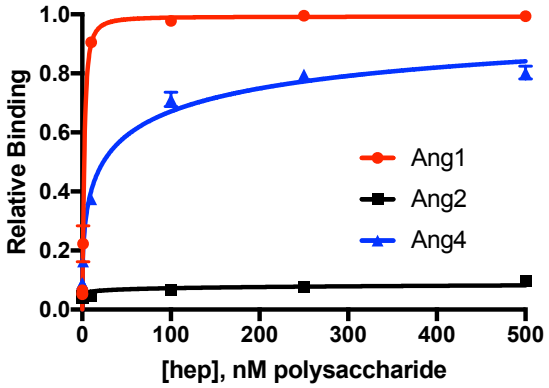



**Figure 3-17. Carbohydrate ELISA using angiopoietin ligands.** Ang-6xHis constructs immobilized on Cu plates were probed with biotinylated HS as previously described. Ang1 and Ang4 show nanomolar affinity to HS, whereas Ang2 does not bind.

carbohydrate ELISA platform. Ang constructs containing a *C*-terminal 6xHis tag were immobilized onto copper functionalized plates and incubated with biotinylated HS. Here, Ang1 and Ang4 showed strong binding to HS, with dissociation constants of 2.23 ± 0.27 nM and 42.0 ± 43.0 nM, respectively, whereas Ang2 did not show binding to HS (Figure 3-17). The strong affinity of Ang1 for HS is in direct conflict with results described from previous experiments.[64] However, the constructs used in the previous paper were over-expressed in murine cancer cell lines LLC and TA3 and the simian kidney cell line Cos7, which may have produced misfolded forms of the protein.

### 3.3.2 Ternary Complex Formation with Tie2

GAGs have been shown previously to both facilitate and impede the formation of a ligand-receptor interaction,[33] thus exerting control over receptor activation and signaling. To rapidly delineate whether HS binding to Ang1 or Ang4 affects their binding to Tie2, we employed carbohydrate microarrays to directly observe

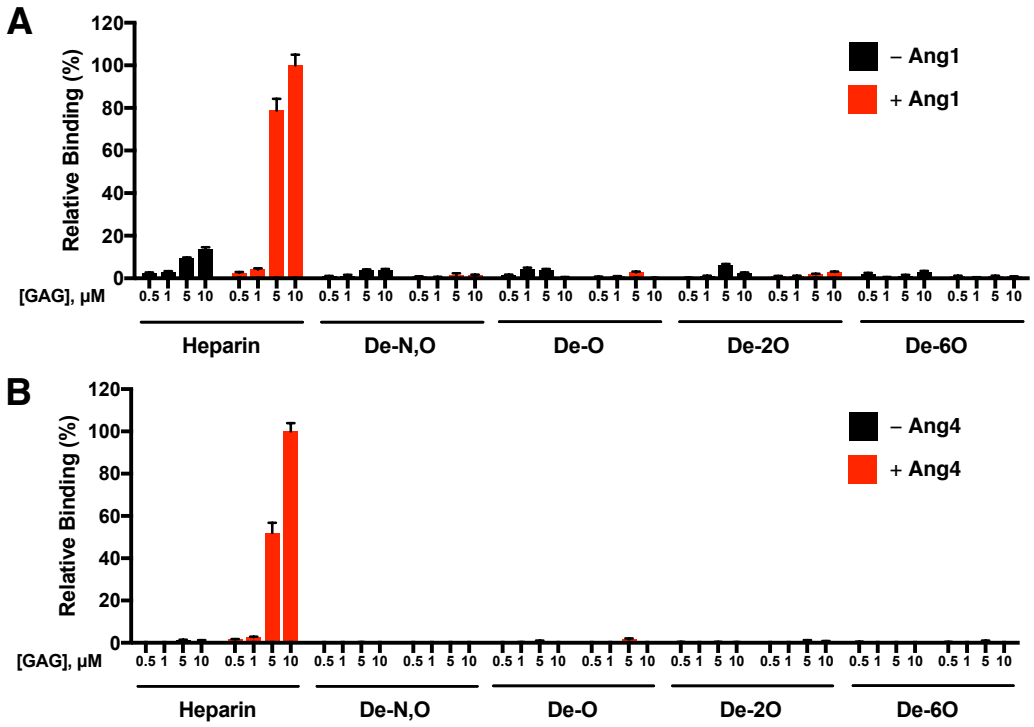



**Figure 3-18. Ternary complex formation of Tie2-Ang-HS.** HS carbohydrate microarrays were incubated with Tie2-Fc with or without (A) Ang1 or (B) Ang4. Immobilized Tie2-Fc was visualized using an anti-human Fc antibody. Binding of Tie2-Fc to the microarray in the presence of Ang1 or Ang4 suggests the formation of carbohydrate-protein-protein ternary complexes.

the formation of ternary, carbohydrate-protein-protein interactions (Figure 3-18). Here, Tie2-Fc, which was previously shown not to bind HS, was incubated with the HS microarrays in the presence or absence of Ang1 or Ang4. Tie2 immobilized onto the microarray through its interaction with HS-bound Ang1/4 was visualized by a fluorescently tagged anti-human Fc antibody. As previously shown, signal from Tie2 alone was not observed on the HS microarray. Interestingly, Tie2 signal dramatically increased in the presence of either Ang1 or Ang4, suggesting that the Ang/HS interaction does not interfere with Ang/Tie2

binding and that the three biomolecules can exist as a single ternary complex. Moreover, the assembly of this ternary complex is sulfation pattern dependent, forming only with the HS motif. This mode of ternary complex formation has previously been demonstrated for both FGF2-FGFR1 and FGF8-FGFR3[66] and suggests that HS at the cell surface can potentiate Ang/Tie2 signaling by bringing Ang ligands into close proximity with the Tie2 receptor.

### 3.3.3 Potentiation of Tie2 Signaling by Glycan Engineering

To examine whether cell surface HS could affect Ang/Tie2 signaling, we decided to utilize our HTP glycan engineering technique.[66] Endothelial cells are generally very refractory to transient transfection;[67] therefore, a lentiviral delivery approach was developed. Here, the HTP transmembrane construct was cloned into a pENTR4 vector and transferred to a pLenti-CMV-Blast destination vector using Gateway cloning.[68] Lentiviruses (HTP-LV) were produced using third generation packaging plasmids in HEK-293T cells and concentrated by PEG precipitation.[69] The endothelial EA.hy926 cell line was treated with different amounts of HTP-LV and subjected to selection with blasticidin after 48 h. Cells from the sample treated with the lowest amount of HTP-LV that survived three days after all untreated cells were killed were used as the EA.hy926-HTP cell line moving forward. To validate the expression of HTP on the cell surface, EA.hy926-HTP and untransduced, parental cells were treated with cell-impermeable AF488-CL and Hoescht 33324 dye for 15 min at 37 °C, washed, and live imaged (Figure

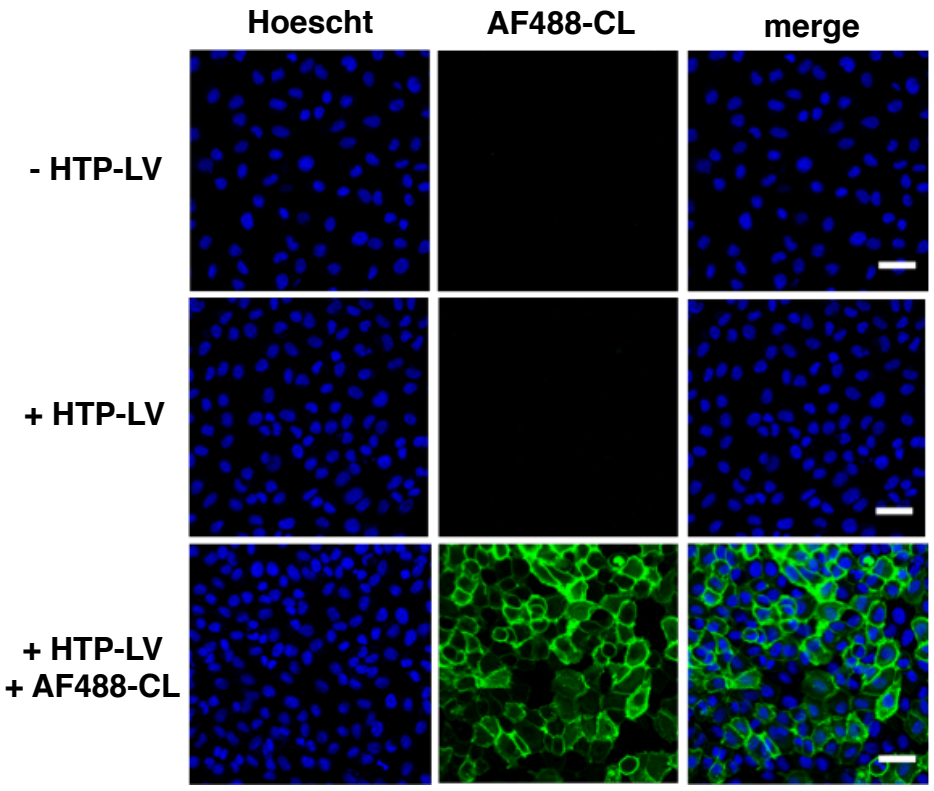



**Figure 3-19. Production of an HTP-expressing endothelial cell line.** EA.hy926 cells were treated with an HTP-expressing lentivirus and selected for incorporation. The resulting cells (+ HTP-LV) and the parental cells (- HTP-LV) were treated with AF488-CL to visualize cell-surface HTP. Scale bar = 50 μm.

3-19). As expected, cells treated with HTP-LV showed robust surface labeling by AF488-CL, whereas the parental cells showed no surface labeling.

With HTP-expressing endothelial cells in hand, we set to test the effect of cell-surface HS on Ang/Tie2 signaling (Figure 3-20). To achieve this, EA.hy926-HTP cells were first treated with heparinase I/III to eliminate endogenous HS and then grown to confluency. Cells then treated with either 5 μg/mL HS-CL or desulfated (deS) HS-CL for 2 h at 37 °C in serum-starved medium (0.5% fetal bovine serum). Cells were washed and then incubated for another 4 h in serum-starved medium

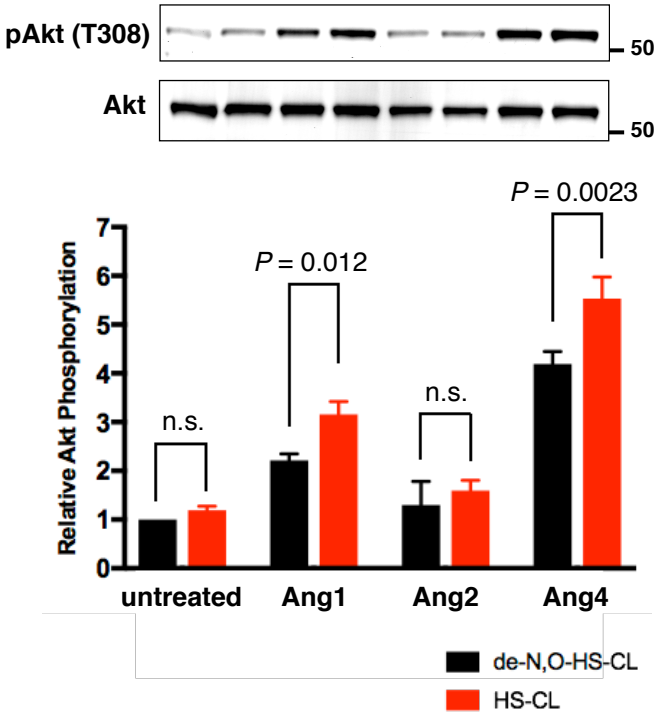



**Figure 3-20. Potentiation of Tie2 signaling by glycan engineering.** EA.hy926 cells expressing HTP were treated with de-HS-CL or HS-CL and stimulated with Ang1, Ang2, or Ang4. Tie2 activation was quantified by measurement of Akt phosphorylation at Thr-308 using Western blotting.

and then treated with 500 ng/mL Ang1, Ang2, or Ang4 for 30 min at 37 °C. Cells were then washed with PBS and immediately lysed. Phosphorylation of Akt at Thr308, which is dependent on 3-phosphoinositide dependent kinase 1 (PDK1), was then measured by Western blotting as a readout for downstream Tie2 signaling. Impressively, we saw that for both Ang1 and Ang4, the presence of HS-CL on the cell surface led to a significant increase in Akt phosphorylation compared to cells displaying with the non-binding de-HS-CL. No significant difference was observed at basal levels or with Ang2, suggesting that this effect is specific to increased Tie2 activation by Ang1 and Ang4 bound to HS. Together with the ternary complexation data, these results suggest that HS, which is highly

upregulated at the cell surface of endothelial cells, can potentiate Tie2 activation by Ang1 and Ang4 through the formation of active ternary signaling complexes.

### 3.4 Conclusions

The results presented herein provide the first evidence for the involvement of HS in the Ang/Tie signaling pathway. The binding of orphan receptor Tie1 to GAGs is the first report of a ligand for this protein since its discovery 25 years ago. Furthermore, the use of biochemical and chemical biology tools to identify amino acid residues necessary for the Tie1-GAG interaction has allowed the production of an animal model that is deficient in this interaction without removing the protein or the carbohydrate. This is the first report of an animal model using this approach, which will provide unparalleled, specific information of the importance of a protein-carbohydrate interaction *in vivo*. Moreover, the interaction between HS and Ang1 and Ang4 parallel the widely observed phenomenon of HS-ligand interactions. Using our glycan engineering method, we were able to provide direct evidence that HS on the cell surface potentiates signaling through HS-Ang1/4-Tie2 ternary complexes as demonstrated by our carbohydrate microarrays. These results as a whole provide crucial information regarding the regulation of the Tie/Ang signaling axis and may help further delineate the molecular mechanisms of a well-known but poorly understood RTK subfamily.

## 3.5 Methods and Materials

### 3.5.1 General Methods

*Molecular biology*. All reagents used for molecular biology were obtained from New England Biolabs unless otherwise noted. All primers were obtained from Integrated DNA Technologies. All miniprep, maxiprep, and gel extraction steps were conducted using Zymogen kits (Genesee Scientific). Cloning was generally achieved using the NEBuilder HiFi Cloning Master Mix (New England Biolabs). Human Tie1 cDNA was obtained from pDONR223-TIE1 plasmid (23946, Addgene) and cloned into a pCDNA3.1 vector (ThermoFisher Scientific) containing an *N*-terminal Ig κ-leader sequence and *C*-terminal human $IgG_1$ Fc and 6xHis tags. Truncations and site mutations were produced using the Q5 site-directed mutagenesis kit (New England Biolabs). The transmembrane HTP construct was produced as previously described.[66] To produce the lentiviral construct, he HTP sequence was first cloned into the pENTR4 vector (17424, Addgene) and then transferred into the pLenti-CMV-Blast vector (17451, Addgene) using the Gateway LR Clonase kit (ThermoFisher Scientific). Vectors used to test gRNA sequences include pCAG-EGxxFP (50716, Addgene) and pX459 (62988, Addgene).

*Cell culture*. All reagents used for cell culture were obtained from ThermoFisher Scientific unless otherwise noted. The EA.hy926 cell line, HEK-293T cell line and HUVECs were obtained from the American Type Culture Collection. EA.hy926 and HEK-293T cells were cultured in DMEM with 10% fetal bovine serum (FBS)

and 1x penicillin/streptomycin (P/S) (referred to now on as complete DMEM). HUVECs were cultured using endothelial cell basal medium (EBM-2, Lonza) with the EGM-2 Bulletkit additives (Lonza) (referred to now as EGM-2 medium) and were used for experiments between 0 and 5 passages. For serum starvation, EA.hy926 cells were cultured in DMEM with 0.5% FBS and 1x P/S. Cells were generally grown on plastic, tissue culture treated plates (Sarstedt) without any coating. For experiments, cells were plated onto plastic plates (for lysate harvesting) or glass-bottom plates (imaging) precoated with 5 $\mu g/cm^2$ bovine fibronectin (Sigma Aldrich) and 10 $\mu g/cm^2$ rat tail collagen type I (Sigma Aldrich). Plates were coated by incubation of plates with the above-mentioned proteins in sterile PBS for 1 h at 37 °C followed by two rinses with sterile PBS and were used immediately.

*Western blotting*. Protein samples were first diluted with one-fourth volume of 4x sodium dodecyl sulfate-polyacrylamine gel electrophoresis (SDS-PAGE) loading buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 40% glycerol, 0.4% bromophenol blue) and heated to 95 °C for 10 min. Samples were then loaded onto NuPAGE 4-12% Bis-Tris protein gels (NP0355BOX, ThermoFisher Scientific) and resolved using 180 V at RT for 1 h with constant voltage. The molecular weight of protein targets was estimated using Precision Plus Protein Dual Color Standards (Bio-Rad). Gels were then transferred onto Immobilion-FL PVDF membrane (IPFL00010, EMD Millipore) using 250 mA at 4 °C for 1-1.5 h with constant amperage. Membranes were then blocked for 1 h at RT with either

5% BSA in TBST (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% Tween) for phosphoprotein detection, 1:1 Li-COR blocking buffer/TBS for goat primary antibodies, or with 5% nonfat milk in TBST, depending on the antibody specifications. In general, primary antibodies were used at 1:1000 (1 μg/mL) in the same buffer as the blocking step with incubation at 4 °C overnight with gentle rocking. Blots were rinsed thrice with TBST prior to incubation with secondary antibodies. In general, highly cross-adsorbed secondary antibodies containing AlexaFluor 680 or AlexaFluor 790 (ThermoFisher Scientific) were used at 1:10000 (0.2 μg/mL) in the same buffer as the blocking step with incubation at RT for 1 h with gentle rocking. Blots were then washed thrice for 5 min with TBST prior to imaging with the Li-COR Odyssey CLx. Quantification was performed using Image Studio Software (Li-COR). Blots were stripped using NewBlot PVDF Stripping Buffer (Li-COR) according to the manufacturer's protocol.

### 3.5.2 Protein Expression and Purification

For each protein construct, five 15-cm plates of 90% confluent HEK-293T cells in 20 mL complete DMEM were transfected with 20-30 μg plasmid using poly(ethyleneimine) (23966-2, Polysciences, Inc.). Briefly, DNA was diluted in 2 mL pre-warmed, serum-free DMEM and PEI (1 mg/mL in sterile dd$H_2O$) was added dropwise with vigorous shaking at a 3:1 w/w PEI/DNA ratio. DNA/PEI complexes were allowed to form for 20 min at RT and then added dropwise to each plate. After 24 h, the medium in each plate was replaced with DMEM containing 2% FBS and 1x P/S. Sometimes, proteins were found to be truncated

once secreted into the medium. To prevent this, the medium was supplemented with 0.5x cOmplete protease inhibitor cocktail without ETDA (PIC-, 11697498001, Sigma Aldrich). At 3 d post-transfection, the medium was harvested and centrifuged to remove cell debris. A new 20-mL aliquot of DMEM, 2% FBS, 1x P/S was added, which was harvested 2 d later, centrifuged, and combined with the first aliquot. The solution pH was adjusted to 8.0 using 2 M Tris, and medium was rotated end-over-end for 2 h at 4 °C with 500 μL Ni-NTA resin (Qiagen) prewashed twice with PBS. The resin was separated from the medium by centrifugation and loaded onto a disposable column (Bio-Rad). The resin was washed with 10 mL phosphate wash buffer (100 mM sodium phosphate pH 7.4, 300 mM NaCl, 10 mM imidazole), and protein was eluted with two aliquots of 1 mL elution buffer (50 mM sodium phosphate pH 7.4, 150 mM NaCl, 250 mM imidazole). Elution samples were combined and buffer exchanged by centrifugation with PBS. Samples were stored at 4 °C for up to one week or flash-frozen and stored at -20 °C.

**3.5.3 Protein-GAG Binding Assays.**

*Carbohydrate enzyme-linked immunosorbent assay.* Biotinylated CS-E and HS were prepared as previously described.[38] Initial ELISAs were performed with commercially available Tie1- and Tie2-Fc conjugates (R&D Systems). Wells of Protein A/G 8-well strips (ThermoFisher Scientific) were rinsed twice with 200 μL PBST (PBS + 0.02% Tween-20) and then incubated with 100 μL 10 μg/mL Tie1-Fc or Tie2-Fc in 1% bovine serum albumin (BSA, Fisher Scientific) in PBS for

1 h at RT with gentle rocking. Wells were rinsed thrice with PBST and next incubated with 200 μL 5% BSA in PBS for 1 h at RT. During this incubation step, a dilution series of biotinylated CS-E or HS was produced in PBS. Wells were again rinsed thrice with PBST, and 50 μL of each GAG-containing solution was added to the wells. Wells were incubated with GAG-containing solutions for 3 h at RT with gentle rocking. Wells were rinsed four times with PBST and then incubated with 100 μL 1:15000 HRP-conjugated streptavidin (ThermoFisher Scientific) for 1 h at RT with gentle rocking. Wells were again rinsed four times with PBST and then incubated with 100 μL development solution (R&D Systems) for 15 min at RT in the dark. Reactions were quenched by the addition of 100 μL 2 N $H_2SO_4$, and absorbance was read at 450 nm. Assays were performed in duplicate for each concentration, and values were reported as the average ± SEM. Curves were fitted and $K_D$ values were calculated with Prism 7 (GraphPad) using the nonlinear curve fitting of standard binding with Hill slope function. For experiments using over-expressed, purified proteins, protein concentrations were first normalized using the bicinchoninic acid (BCA) assay (ThermoFisher Scientific) and used the same as the purchased proteins mentioned above.

*GAG microarrays.* GAG microarrays were generated and conducted as previously described.[33] Briefly, slides were blocked with 10% FBS in PBS for 1 h at RT. Stock solutions of Tie1-Fc, Tie2-Fc, Ang1, and Ang4 (2 μM) were prepared in sterile PBS containing 1% BSA. Blocked slides were washed once with PBS and slowly rocked with 150 μL of 1% BSA/PBS containing 1 μM Tie1-Fc for 2 h at RT. For

ternary complex formation, 150 μL of 1% BSA/PBS containing 1 μM Tie2-Fc with or without 1 μM ligand (*i.e.*, 1 μM Tie2-Fc ± 1 μM Ang1; 1 μM Tie2-Fc ± 1 μM Ang4) were used immediately without prior incubation. Slides were washed three times with PBS and incubated with an AF647-conjugated goat anti-human Fc antibody (1:5000) in 1% BSA/PBS for 1 h at RT in the dark with gentle rocking. Slides were then washed three times with PBS and twice with dd$H_2O$ and blown dry under a stream of filtered air. Arrays were scanned using a G2565BA DNA Microarray Scanner (Agilent), and fluorescence was quantified using GenePix 5.0 software (Molecular Devices) with normalization against local background. The data represent the average of 10 spots per concentration of polysaccharide.

*Surface plasmon resonance*. All experiments were conducted using a Biacore T200 instrument (GE Healthcare). A CM5 sensor chip (GE Healthcare) was first functionalized with streptavidin using the manufacturer's amine coupling protocol. Briefly, all flow cells of a CM5 sensor chip (GE Healthcare) were incubated with a 1:1 molar ratio of *N*-hydroxysuccinimide and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide for 3 min at a flow rate of 10 μL/min followed by 1 μM streptavidin in 0.01 NaOAc pH 5.0 until the response rate leveled off (saturating the flow cell surface). Remaining reactive groups were quenched by ethanolamine. Solutions of biotinylated CS-E or HS were then each flowed over one of the four flow cells (CS-E: flow cell 2; HS: flow cell 4) to provide a response value of roughly 25 RU. As a control, an antibody specific to CS-E was flowed over the CS-E containing flow cell and corresponding control

cell to ensure successful loading of biotinylated GAG. To test binding of Tie1-Fc to the biotinylated GAGs, a two-fold dilution series of Tie1-Fc (starting at 100 nM) was produced and flowed over the flow cells for 300 s at a rate of 20 μL/min. Dissociation of Tie1-Fc was then monitored for 350 s using the same flow rate. Between each run, the surface was regenerated using 2.5 M $MgCl_2$ at a flow rate of 30 μL/min. Response values are reported after substraction of the control (-GAG) cell. The resulting curves were fit to the one-to-one Langmuir binding model equation using the Biacore SPR evaluation software.

### 3.5.4 Computational Methods

*Structure modeling*. The Tie1 homology model was constructed using the program SWISS-MODEL[42] with human Tie2 ligand-binding domain (PDB ID 2GY5)[41] as a template and the sequence of human Tie1 (UniProt ID P35590). The Tie1 homology model was briefly minimized in vacuum using the DREIDING force field[70] to reduce steric clashes built into the model. To produce the GAG ligand for docking, a dodecasaccharide structure was first generated for CS-E and HS, and the ligand side chains were relaxed by a brief minimization step. Ligands were then placed in a water box, and a molecular dynamics (MD) simulation was run for 5 ns. The structure closest to the average structure over the 5 ns MD run was selected as the conformation for docking. The dodecasaccharide was then truncated to a hexasaccharide by removing three monosaccharides from the reducing and non-reducing ends.

*Docking and binding site determination*. First, a coarse round of docking is performed. Here the ligand is docked to the entire protein surface. Prior to the docking run, the protein structure is "alanized" by replacing bulky, nonpolar residues (Val, Leu, Ile, Met, Phe, Try, and Trp) with Ala to allow for more complete sampling of the binding site. The protein is split into regions, and spheres are generated for docking. For each region, docked ligand poses are generated, scored, and clustered into families based on energy and root-mean-square deviation. For each cluster, the lowest energy pose known as the family head is output, generally producing 120 poses for each docked region. These poses are then de-alanized and undergo side chain optimization using SCREAM[71] to produce a unique set of protein side chain conformations for each pose. The top docking regions are identified by ranking the average energy of the top five poses in each region. Top regions are then subjected to an additional round of docking. Here a minimization step is performed to improve protein-ligand interactions following side chain placement and optimization with SCREAM. These poses are clustered as before, and family heads are output. The scoring energy known as the snap binding energy used to rank poses is equal to the complex energy minus protein energy and ligand energy.

**3.5.5 Biological Assays**

*Cell-surface crosslinking*. EA.hy926 cells were plated onto precoated six-well plates and grown to confluency with complete DMEM. Cells were rinsed three times with ice-cold PBS and then incubated with 100 μL 1 mM DTSSP in PBS (freshly

prepared) for 30 min at RT with gentle rocking. Unreacted DTSSP was quenched by addition of 5 μL 1 M Tris pH 8.0 with incubation for 15 min at RT. The crosslinking solution was then removed, and cells were lysed by scraping with ice-cold 200 μL Triton lysis buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 1% Triton X-100, 5% glycerol, 1x PIC-). Lysates were rotated end-over-end for 10 min at 4 °C and then centrifuged for 10 min at 21,000 x *g* to remove all insoluble material. The lysates were transferred to new tubes and rotated end-over-end with 1 μg Gt anti-Tie1 (AF619, R&D Systems) overnight at 4 °C. The next day, 20 μL magnetic Protein A/G beads (ThermoFisher Scientific) prewashed with HBS (50 mM HEPES pH 7.9, 150 mM NaCl) was added to each sample and rotated end-over-end for 1 h at 4 °C. Beads were then washed with 500 μL HBS, 0.1% Triton X-100 thrice for 5 min at 4 °C. After the final wash, all wash buffer was removed, and the beads were resuspended in 30 μL 1x SDS-PAGE loading buffer + 20 mM DTT (freshly prepared) and heated to 95 °C for 10 min. Samples were processed for Western blotted as described above. Blots were probed with 1 μg/mL Gt anti-Tie1 (AF619, R&D Systems), 1 μg/mL Gt anti-Tie2 (AF313, R&D Systems), and 1 μg/mL Gt anti-$\alpha_5$ integrin (AF1864, R&D Systems).

*Proximity ligation assays.* EA.hy926 cells were plated onto precoated 96-well glass bottom plates and grown to confluency. Cells were then washed twice with PBS and then fixed with 4% paraformaldehyde in PBS for 20 min at RT. Cells were rinsed thrice with PBS and blocked with the DuoLink blocking buffer (Sigma Aldrich) for 1 h at RT. Cells were rinsed twice with PBS and incubated overnight

at 4 °C with 1:100 Ms anti-Tie1 (MAB619, R&D Systems) and 1:100 Rb anti-Tie2 (H176, Santa Cruz Biotechnology) in 1% BSA/PBS. Cells were then carried through the Duolink labeling protocol scaled to 30 μL reactions per well according to the manufacturer's protocol (DUO92101, Sigma Aldrich). Cells were visualized using an LSM 710 confocal microscope (Zeiss) with a 40x oil immersion objective and preset filter settings for Texas Red and DAPI. A *Z*-stack of images was collected for each field of view, and a single image was produced using maximal pixel intensity.

*Lentivirus production.* Lentiviruses were produced by co-transfecting two to four 10-cm plates of 90% confluent HEK-293T cells in complete DMEM with the HTP lentiviral plasmid and pLP1, pLP2, and VSV-G helper plasmids (ThermoFisher Scientific) at 4.3 μg, 6.2 μg, 3.1 μg, and 3.7 μg, respectively, using Lipofectamine 3000 (ThermoFisher Scientific) according to the manufacturer's protocol. Medium was harvested and replaced with complete DMEM at 24, 48, and 72 h, centrifuged to remove cell debris, and combined. Lentivirus particles were either directly flash frozen and stored at -80 °C or were concentrated by PEG-6000 precipitation as previously described.[69] Precipitated lentiviruses were resuspended in complete DMEM, flash frozen, and stored at -80 °C.

*Generation and validation of EA.hy926-HTP cell line.* EA.hy926 cells were grown to 80% confluency in a six-well plate with complete DMEM and then treated with 0-20 μL of the concentrated HTP-LV stock along with 8 μg/mL polybrene (Sigma Aldrich). After 48 h, cells were split 1:3 into new six-well plates in complete

DMEM containing 5 μg/mL blasticidin (InvivoGen). Medium was changed every 2 d. After roughly 1-1.5 weeks, all cells in the untreated well had died. Cells were treated for an additional 3 d with 5 μg/mL blasticidin, and cells from the well treated with the lowest amount of lentivirus that survived selection were expanded and cryo-preserved. To validate expression of HTP, treated and untreated EA.hy926 cells were plated onto precoated 96-well glass bottom plates and allowed to grow to confluency. Cells were then treated with 1:1000 cell impermeable AlexaFluor 488-conjugated HTL (G1001, Promega) and 1:1000 Hoescht 33342 (H3750, ThermoFisher Scientific) in complete DMEM at 37 °C for 15 min. Cells were rinsed thrice with complete DMEM and live imaged on an LSM 710 confocal microscope (Zeiss) with a 10x objective.

*Glycan engineering and stimulation assay.* EA.hy926-HTP cells were treated in suspension with 2 U/mL heparinase I/III in complete DMEM for 2 h at 37 °C with gentle mixing every 15 min. Cells were then plated on precoated six-well plates at 100,000 cells per well and grown overnight at 37 °C with complete DMEM. The following day, medium was replaced with serum-starved DMEM (0.5% FBS) containing 5 μg/mL de-HS-CL or HS-CL for 2 h. Medium was removed, and cells were treated with serum-starved DMEM for an additional 4 h. Cells were then stimulated with serum-starved DMEM containing 500 ng/mL Ang1, Ang2, or Ang4 or no ligand for 30 min at 37 °C. Cells were rinsed with ice-cold PBS and then lysed in 200 μL SDS lysis buffer (50 mM Tris pH 7.6, 150 mM NaCl, 1% SDS, 1x PIC-, 1x Phos-STOP (Sigma Aldrich)) at RT. Lysates were sonicated for 10 s at

40% intensity to sheer DNA and then centrifuged for 5 min at 21,000 x *g* to remove any insoluble debris. Lysates were subjected to Western blotting as described above. Blots were probed using 1:1000 Rb anti-phosphoAkt(T308) and 1:1000 Ms anti-Akt (9275 and 9272, Cell Signaling Technology). Assay was performed in triplicate. Phospho-Akt values were normalized to total Akt levels and are reported as the average value as normalized to the untreated de-HS-CL control. Statistical analysis was performed by Prism 7 (GraphPad) using a two-way ANOVA with Sidak's multiple comparisons test.

### 3.5.6 Transgenic Tie1 Mouse Model

*Generating and testing gRNA sequences.* Murine Tie1 gDNA was amplified from a C57BL/6J mouse gDNA sample (Jackson Laboratories) purified by DNEasy Blood and Tissue Kit (Qiagen) using the Q5 2x hot-start PCR master mix (New England Biolabs) using the following primers:

```
Tie1-EGxxFP-F:
CGTGCTGCTGCCCGACAACCACTGAGGATCCGTGCTGTGGATGAGAGGTAAGTT
Tie1-EGxxFP-R:
GGGTCAGCTTGCCGATATCGAATTCGTCGACGAAGAGGCTTGCAGGTAAAGAGC
```

The amplified region was gel purified and cloned into the pCAG-EGxxFP vector between BamHI and SalI sites. Correct incorporation of the sequence was validated using Sanger sequencing with the following primer:

```
EGxxFP-Seq: CTACAACAGCCACAACGTCTAT
```

Guide RNA sequences were designed using the online CHOPCHOP tool (http://chopchop.cbu.uib.no/) against the Dec. 2011 (GRCm38/mm10) murine

genome assembly against regions surrounding the Tie1 Arg-38 and Arg-82 codons. Sequences were selected based on predicted activity and minimal off-target homology. Primers to clone these gRNA sequences into the pX459 vector are shown below:

**Tie1 Arg-82:**

Tie1-gRNA-1F: CACCGGTGAGAACCGTTGCGTGCC
Tie1-gRNA-1R: TTTGGGCACGCAACGGTTCTCACC
Tie1-gRNA-2F: CACCGTTCTCACCAGGTCACGCTG
Tie1-gRNA-2R: TTTGCAGCGTGACCTGGTGAGAAC
Tie1-gRNA-3F: CACCGGCACGCAACGGTTCTCACC
Tie1-gRNA-3R: TTTGGGTGAGAACCGTTGCGTGCC
Tie1-gRNA-4F: CACC**G**ATCGTGCGCACCTTCCCGCC
Tie1-gRNA-4R: TTTGGGCGGGAAGGTGCGCACGAT**C**

**Tie1 Arg-38:**

Tie1-gRNA-5F: CACC**G**CGCAGGTCAGGAAGAAACGC
Tie1-gRNA-5R: AAACGCGTTTCTTCCTGACCTGCG**C**
Tie1-gRNA-6F: CACCGTCCGAGCTCCTCCCTGCTC
Tie1-gRNA-6R: AAACGAGCAGGGAGGAGCTCGGAC
Tie1-gRNA-7F: CACC**G**TTCTTCCTGACCTGCGTGTC
Tie1-gRNA-7R: AAACGACACGCAGGTCAGGAAGAA**C**
Tie1-gRNA-8F: CACC**G**CCTGACCTGCGTGTCTGGTG
Tie1-gRNA-8R: AAACCACCAGACACGCAGGTCAGG**C**
Tie1-gRNA-9F: CACC**G**CCTCACCAGACACGCAGGTC
Tie1-gRNA-9R: AAACGACCTGCGTGTCTGGTGAGG**C**

For gRNA sequences that did not begin with G, a single G:C bp was added (shown in bold) to the primers to allow for efficient transcription by RNA polymerase III. gRNA oligonucleotides were dissolved at 100 μM in $ddH_2O$, and annealing/phosphorylation reactions were set up as follows: 1 μL forward primer, 1 μL reverse primer, 1 μL 10x T4 ligation buffer, 6.5 μL $ddH_2O$, 0.5 μL T4 PNK. Reactions were incubated using a thermocycler at 37 °C for 30 min, followed by

95 °C for 5 min and ramping down to 25 °C at 5 °C/min. The reaction mixture was diluted 1:250 in $ddH_2O$ and used in digestion/ligation reactions as follows: 100 ng pX459 vector, up to 12.5 μL with $ddH_2O$, 2 μL 1:250 diluted reaction, 1 μL 10 mM DTT, 1 μL 10 mM ATP, 2 μL 10x FastDigest buffer (ThermoFisher Scientific), 1 μL FastDigest BpiI, 0.5 μL T7 DNA ligase. Reactions were incubated for six cycles of 5 min at 37 °C then 5 min at 23 °C and could then be used for transformation. Correct incorporation of the gRNA sequence was validated with Sanger sequencing using the following primer:

U6-F-Seq: GACTATCATATGCTTACCGT

To test gRNAs, confluent six-well plates of HEK-293T cells were transfected with 0.5 μg pCAG-EG(Tie1)FP with 0.5 μg of each of the gRNA-containing pX459 vectors. After 48 h, cells were live imaged using an LSM 710 confocal microscope (Zeiss) with a 5x objective using GFP filter sets. gRNA efficiency was estimated by the percentage of $GFP^+$ cells. gRNA1 and gRNA5 were the best performing gRNA sequences for Arg-82 and Arg-38 regions, respectively.

*Designing repair template*. The repair template to be used for injection was designed either as a pair of single stranded oligonucleotides (ssODNs) or a larger double stranded DNA (dsDNA) fragment. In both cases, mutations for Arg-to-Ala and silent mutations disrupting the NGG PAM sequence were added. The R38A mutation produced a StuI restriction site, and a silent mutation was introduced at R82A to produce an EaeI restriction site to be used for rapid genotyping. ssODNs and dsDNA were purchased as DNA Ultramers and a gBlock gene fragment,

respectively (Integrated DNA Technologies). The sequences for the repair templates are shown below, with mutations bolded, the Arg-82 codon in green, and the Arg-38 codon in orange.

**ssODN (Arg-38):**

caggggcgggtccgagctcctccctgctccggcctcaccagacacgcaggtcaggaa
gaaGGCctg**A**gggtctgtgatgcgcaggttggccagcaatgttaagtcaacagacgc
acctaggaccaaa

**ssODN (Arg-82):**

gacgcctaccaaatccgagggcttggagaagcccctcagcgtgacctggtgagaacc
gtt**GGCG**gccag**A**tacaggggctgcccaggcgggaaggtgcgcacgatgcgatcatc
cttctccagcagcagg

**dsDNA:**

cgcaggctcctggtccttttagtacagtaggagggtggctcagagctacagatgaga
gggaagcccctagaggttggtcacaaggatggcaggtgggtcagggtggactaagta
tatggtgggccatggatggaagacctaaggtgtcagcatcatgggtctccctgtccc
accaactgcccctcaccgttgttcttccagatcacgtcagtctgcttttccttgtgc
acatgggcagaaagcacagcagtgtcacctttgttcaccgtgtgtgtgaccttgtct
ggaaacaggtgtgctgtggatgagaggtaagttcagcgatctacaacacagccctcc
ctctggtggtcgatctcctggtaggccagactcacccctgggctgttgtgcacata
gagaactcgagtgcgccgcgcgccagcaccacccacacaggagaagacgcctaccaa
atccgagggcttggagaagcccctcagcgtgacctggtgagaaccgtt**GGCG**gccag
**A**tacaggggctgcccaggcgggaaggtgcgcacgatgcgatcatccttctccagcag
caggggcgggtccgagctcctccctgctccggcctcaccagacacgcaggtcaggaa
gaaGGCctg**A**gggtctgtgatgcgcaggttggccagcaatgttaagtcaacagacgc
acctaggaccaaacaacagtgactccatggtggatcctgccagctcagcagccccag
gagcccgggactccaacagtaacttgtgttaatatttcactgtgtccttagaagact
tcacttttcaaagatttccgcccattcccttcccttacacatgtgctcgctcttta
cctgcaagcctcttcgtcctaaaacataaacaaaagcacacatccttctctggtcca
gacagggaatgagctacgcttctgtctccttcccttactctgtccgtgtcaggtcat
tgccctgactcctgtgtgaaccattcccaccaaggctatccctgagccacctgtccc
tgccccatcatggaggtgcaggttagaagagtcgctgagactggagtgtcatgagga
cgcacaccctctttcttgcttaactaattgaggactatttgactggctgggattgaa
gatcggttagggccagagagacacagacacaggtgggtctgggccctaccttgggaa
gctggaggaccagacacaggcatccctcccagatggaggtgttggctcagagggagg
aacatctaaccagcaggggggcgggggaaggaaaggggacaggaaaagatgcagata
acaaacttcattagaaggtgcttgggtgaggaggaacgaaaggtcaggcgtagggag
aggccaccctacgctcaagtatgagaaatgttgtagtgtggggaggggagctgcccc

gggctagtgcaactggagcgtaaaatctgtaccctgaagtgggcaagaccgaggaga
agaatgtgatcagcctgct

*Preparation of injection mixture*. gRNAs were purchased as the crRNA/tracrRNA system from Integrated DNA Technologies. These constructs contain chemical modifications to increase RNA stability. RNA molecules were dissolved in sterile injection buffer (1 mM Tris pH 7.5, 0.1 mM EDTA) at 1 μg/μL. Mature crRNA:tracrRNA complexes were produced by combining 5 μL 1 μg/μL crRNA and 10 μL 1 μg/μL tracrRNA. Samples were incubated in a thermocycler at 95 °C for 5 min and then ramped down to 25 °C at 5 °C/min. Samples were diluted two-fold in injection buffer, and 4 μL of each complex was combined along with 64 μL injection buffer. Cas9 protein (Integrated DNA Technologies) was diluted to 500 ng/μL in injection buffer, and 8 μL was added to the RNA mixture. Samples were incubated at RT for 15 min. The Cas9/gRNA mixture was aliquoted at 20 μL, and either 1 μL 500 ng/μL ssODN (Arg-82 and Arg-38) or 2.5 or 5 μL 100 ng/μL dsDNA repair template was added with injection buffer to achieve 25 μL total. Samples were centrifuged for 10 min at 21,000 x *g*, and 20 μL was transferred to a new tube. The final concentrations of the injection solution were 20 ng/μL each gRNA/Cas9 complex and 10 or 20 ng/μL repair template. Roughly 20 zygotes per condition from superovulated, mated B6SJL-F1/J (Jackson Laboratories) were harvested and cultured. The injection mixture was pronuclear injected into each zygote, and zygotes were allowed to develop for 3 d.

*Blastocyst genotyping*. Blastocysts were aliquoted into a V-bottom 96-well plate in 10 μL culture medium and flash frozen. To each well, 10 μL 2x blastocyst lysis buffer (200 mM Tris pH 8.3, 200 mM KCl, 0.02% gelatin, 0.45% Tween-20, 60 μg/mL yeast tRNA, 125 μg/mL proteinase K) was added and allowed to incubate for 5 min. Samples were mixed by pipetting, transferred to PCR tubes, and incubated for 10 min at 56 °C and 10 min at 95 °C. gDNA was amplified in two rounds of nested PCR with Q5 2x Hot-Start Master Mix (New England Biolabs). First, a 25 μL reaction was set up using 4 μL lysate and amplified for 35 cycles with annealing temperature of 65 °C. Next, 4 μL of the first reaction was using for a second PCR reaction using a second set of primers with the same annealing temperature. Primers used are shown below:

```
Tie1-gDNA-Geno-F1: AGAGCTACAGATGAGAGGGAAG
Tie1-gDNA-Geno-F2: CTAGAGGTTGGTCACAAGGATG
Tie1-gDNA-Geno-R1: CACATTCTTCTCCTCGGTCTTG
Tie1-gDNA-Geno-R2: TCACCCAAGCACCTTCTAATG
```

Amplicons were gel purified and submitted for Sanger sequencing using the second forward primer.

For production of litters, roughly 80 zygotes were injected once per day for four days. Once zygotes had proceeded to the blastocyst stage, they were implanted into surrogate B6SJL-F1/J mice. Successful litters were tailed for genotyping. gDNA from tail samples was harvested using the DNEasy Kit and amplified in a single step using F1/R1 genotyping primers listed above. Amplicons were gel purified and submitted for Sanger sequencing using the F2 primer.

### 3.6 References


1. J. Partanen *et al.* A novel endothelial cell surface receptor tyrosine kinase with extracellular epidermal growth factor homology domains. *Mol Cell Biol.* **1992**, *12*: 1698-1707.
2. D. J. Dumont, T. P. Yamaguchi, R. A. Conlon, J. Rossant, M. L. Breitman. Tek, a novel tyrosine kinase gene located on mouse chromosome 4, is expressed in endothelial cells and their presumptive precursors. *Oncogene.* **1992**, *7*: 1471-1480.
3. T. N. Sato, Y. Qin, C. A. Kozak, K. L. Audus. Tie-1 and Tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. *Proc Natl Acad Sci USA*. **1993**, *90*: 9355-9358.
4. A. Iwama *et al.* Molecular cloning and characterization of mouse TIE and TEK receptor tyrosine kinase genes and their expression in hematopoietic stem cells. *Biochem Biophys Res Commun.* **1993**, *195*: 301-309.
5. S. Davis *et al.* Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell.* **1996**, *87*: 1161-1169.
6. P. C. Maisonpierre *et al.* Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis. *Science.* **1997**, *277*: 55-60.
7. D. M. Valenzuela *et al.* Angiopoietins 3 and 4: diverging gene counterparts in mice and humans. *Proc Natl Acad Sci USA*. **1999**, *96*: 1904-1909.
8. N. Jones, K. Iljin, D. J. Dumont, K. Alitalo. Tie receptors: new modulators of angiogenic and lymphangiogenic responses. *Nat Rev Mol Cell Biol.* **2001**, *2*: 257-267.

9. H. G. Augustin, G. Y. Koh, G. Thurston, K. Alitalo. Control of vascular morphogenesis and homeostasis through the angiopoietin-Tie system. *Nat Rev Mol Cell Biol.* **2009**, *10*: 165-177.

10. P. Saharinen, M. Jeltsch, M. M. Santoyo, V.-M. Leppänen, K. Alitalo. *The TIE receptor family*. In: D. L. Wheeler, Y. Yarden, Eds., *Receptor Tyrosine Kinases: Family and Subfamilies*. Ed.: Humana Press; 2015.

11. D. J. Dumont *et al.* Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo. *Genes Dev.* **1994**, *8*: 1897-1909.

12. T. N. Sato *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature.* **1995**, *376*: 70-74.

13. M. C. Puri, J. Rossant, K. Alitalo, A. Bernstein, J. Partanen. The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells. *EMBO J.* **1995**, *14*: 5884-5891.

14. G. D'Amico *et al.* Loss of endothelial Tie1 receptor impairs lymphatic vessel development-brief report. *Arterioscler Thromb Vasc Biol.* **2010**, *30*: 207-209.

15. X. Qu, K. Tompkins, L. E. Batts, M. Puri, H. S. Baldwin. Abnormal embryonic lymphatic vessel development in Tie1 hypomorphic mice. *Development.* **2010**, *137*: 1285-1295.

16. X. Qu, B. Zhou, H. S. Baldwin. Tie1 is required for lymphatic valve and collecting vessel development. *Dev Biol.* **2015**, *399*: 117-128.

17. K. V. Woo *et al.* Tie1 attenuation reduces murine atherosclerosis in a dose-dependent and shear stress-specific manner. *J Clin Invest.* **2011**, *121*: 1624-1635.

18. G. D'Amico *et al.* Tie1 deletion inhibits tumor growth and improves angiopoietin antagonist therapy. *J Clin Invest.* **2014**, *124*: 824-834.

19. H. Huang, A. Bhat, G. Woodnutt, R. Lappe. Targeting the ANGPT-TIE2 pathway in malignancy. *Nat Rev Cancer*. **2010**, *10*: 575-585.

20. S. Han *et al.* Amelioration of sepsis by TIE2 activation-induced vascular protection. *Sci Transl Med*. **2016**, *8*: 335ra355.

21. H. J. Lee *et al.* Biological characterization of angiopoietin-3 and angiopoietin-4. *FASEB J*. **2004**, *18*: 1200-1208.

22. S. Savant *et al.* The Orphan Receptor Tie1 Controls Angiogenesis and Vascular Remodeling by Differentially Regulating Tie2 in Tip and Stalk Cells. *Cell Rep*. **2015**, *12*: 1761-1773.

23. S. B. Mueller, C. D. Kontos. Tie1: an orphan receptor provides context for angiopoietin-2/Tie2 signaling. *J Clin Invest*. **2016**, *126*: 3188-3191.

24. H. Singh, T. A. Tahir, D. O. Alawo, E. Issa, N. P. Brindle. Molecular control of angiopoietin signalling. *Biochem Soc Trans*. **2011**, *39*: 1592-1596.

25. T. C. Seegar *et al.* Tie1-Tie2 interactions mediate functional differences between angiopoietin ligands. *Mol Cell*. **2010**, *37*: 643-655.

26. E. A. Korhonen *et al.* Tie1 controls angiopoietin function in vascular remodeling and inflammation. *J Clin Invest*. **2016**, *126*: 3495-3510.

27. J. Schlessinger *et al.* Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol Cell*. **2000**, *6*: 743-750.

28. C. Rolny, D. Spillmann, U. Lindahl, L. Claesson-Welsh. Heparin amplifies platelet-derived growth factor (PDGF)-BB-induced PDGF alpha -receptor but not PDGF beta -receptor tyrosine phosphorylation in heparan sulfate-deficient cells. Effects on signal transduction and biological responses. *J Biol Chem*. **2002**, *277*: 19315-19321.

29. D. Xu, M. M. Fuster, R. Lawrence, J. D. Esko. Heparan sulfate regulates $VEGF_{165}$- and $VEGF_{121}$-mediated vascular hyperpermeability. *J Biol Chem.* **2011**, *286*: 737-745.

30. L. E. Kemp, B. Mulloy, E. Gherardi. Signalling by HGF/SF and Met: the role of heparan sulphate co-receptors. *Biochem Soc Trans.* **2006**, *34*: 414-417.

31. V. Koprivica *et al.* EGFR activation mediates inhibition of axon regeneration by myelin and chondroitin sulfate proteoglycans. *Science.* **2005**, *310*: 106-110.

32. M. W. Barnett, C. E. Fisher, G. Perona-Wright, J. A. Davies. Signalling by glial cell line-derived neurotrophic factor (GDNF) requires heparan sulphate glycosaminoglycan. *J Cell Sci.* **2002**, *115*: 4495-4503.

33. C. J. Rogers *et al.* Elucidating glycosaminoglycan-protein-protein interactions using carbohydrate microarray and computational approaches. *Proc Natl Acad Sci USA*. **2011**, *108*: 9747-9752.

34. P. B. Murray. Heparin is an activating ligand of the orphan receptor tyrosine kinase ALK. *Sci Signal*. **2015**, *8*: ra6.

35. G. Bezakova, M. A. Ruegg. New insights into the roles of agrin. *Nat Rev Mol Cell Biol.* **2003**, *4*: 295-308.

36. T. Arai, A. Parker, W. Busby, Jr., D. R. Clemmons. Heparin, heparan sulfate, and dermatan sulfate regulate formation of the insulin-like growth factor-I and insulin-like growth factor-binding protein complexes. *J Biol Chem.* **1994**, *269*: 20388-20393.

37. M. Teran, M. A. Nugent. Synergistic binding of vascular endothelial growth factor-A and its receptors to heparin selectively modulates complex affinity. *J Biol Chem.* **2015**, *290*: 16451-16462.

38. J. M. Brown *et al.* A sulfated carbohydrate epitope inhibits axon regeneration after injury. *Proc Natl Acad Sci USA*. **2012**, *109*: 4768-4773, S4768/4761-S4768/4710.

39. C. J. Rogers *et al.* Chondroitin sulfate E influences retinotopic mapping via EphB3. *in preparation*. **2017**.

40. G. J. Miller *et al.* Chondroitin sulfate E mediates axonal inhibition via EphA4 activation. *in preparation*. **2017**.

41. W. A. Barton *et al.* Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2-Tie2 complex. *Nat Struct Mol Biol.* **2006**, *13*: 524-532.

42. L. Bordoli *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc*. **2009**, *4*: 1-13.

43. K. Iljin *et al.* A fluorescent Tie1 reporter allows monitoring of vascular development and endothelial cell isolation from transgenic mouse embryos. *FASEB J.* **2002**, *16*: 1764-1774.

44. G. Li *et al.* Glycosaminoglycanomics of cultured cells using a rapid and sensitive LC-MS/MS approach. *ACS Chem Biol.* **2015**, *10*: 1303-1310.

45. S. R. Hubbard, W. T. Miller. Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr Opin Cell Biol.* **2007**, *19*: 117-123.

46. P. Saharinen *et al.* Multiple angiopoietin recombinant proteins activate the Tie1 receptor tyrosine kinase and promote its interaction with Tie2. *J Cell Biol.* **2005**, *169*: 239-243.

47. H. T. Yuan *et al.* Activation of the orphan endothelial receptor Tie1 modifies Tie2-mediated intracellular signaling and cell survival. *FASEB J.* **2007**, *21*: 3171-3183.

48. M. B. Marron, D. P. Hughes, M. D. Edge, C. L. Forder, N. P. Brindle. Evidence for heterotypic interaction between the receptor tyrosine kinases TIE-1 and TIE-2. *J Biol Chem.* **2000**, *275*: 39741-39746.

49. C. D. Kontos, E. H. Cha, J. D. York, K. G. Peters. The endothelial receptor tyrosine kinase Tie1 activates phosphatidylinositol 3-kinase and Akt to inhibit apoptosis. *Mol Cell Biol.* **2002**, *22*: 1704-1713.

50. V. M. Leppanen, P. Saharinen, K. Alitalo. Structural basis of Tie2 activation and Tie2/Tie1 heterodimerization. *Proc Natl Acad Sci USA*. **2017**, *114*: 4376-4381.

51. A. C. Dalton, T. Shlamkovitch, N. Papo, W. A. Barton. Constitutive Association of Tie1 and Tie2 with Endothelial Integrins is Functionally Modulated by Angiopoietin-1 and Fibronectin. *PLoS One.* **2016**, *11*: e0163732.

52. O. Söderberg *et al.* Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods.* **2006**, *3*: 995-1000.

53. N. Oh *et al.* A designed angiopoietin-1 variant, dimeric CMP-Ang1 activates Tie2 and stimulates angiogenesis and vascular stabilization in *N*-glycan dependent manner. *Sci Rep.* **2015**, *5*: 15291.

54. E. Forsberg, L. Kjellen. Heparan sulfate: lessons from knockout mice. *J Clin Invest*. **2001**, *108*: 175-180.

55. J. R. Bishop, M. Schuksz, J. D. Esko. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature.* **2007**, *446*: 1030-1037.

56. P. D. Hsu, E. S. Lander, F. Zhang. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* **2014**, *157*: 1262-1278.

57. H. Wang, M. La Russa, L. S. Qi. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem.* **2016**, *85*: 227-264.

58. A. C. Komor, A. H. Badran, D. R. Liu. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell.* **2017**, *168*: 20-36.

59. H. Yang *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell.* **2013**, *154*: 1370-1379.

60. C. Tong, G. Huang, C. Ashton, P. Li, Q. L. Ying. Generating gene knockout rats by homologous recombination in embryonic stem cells. *Nat Protoc.* **2011**, *6*: 827-844.

61. D. Paquet *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature.* **2016**, *533*: 125-129.

62. D. Mashiko *et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci Rep.* **2013**, *3*: 3355.

63. M. Srivastava *et al.*, An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell (Cambridge, Massachusetts).* **2012**, *151*: 1474-1487.

64. Y. Xu, Q. Yu, Angiopoietin-1, Unlike Angiopoietin-2, Is Incorporated into the Extracellular Matrix via Its Linker Peptide Region. *J Biol Chem.* **2001**, *276*: 34990-34998.

65. Y. Xu, Y. J. Liu, Q. Yu, Angiopoietin-3 is tethered on the cell surface via heparan sulfate proteoglycans. *J Biol Chem.* **2004**, *279*: 41179-41188.

66. A. Pulsipher, M. E. Griffin, S. E. Stone, L. C. Hsieh-Wilson, Long-lived engineering of glycans to direct stem cell fate. *Angew Chem Int Ed Engl.* **2015**, *54*: 1466-1470.

67. M. A. Hunt, M. J. Currie, B. A. Robinson, G. U. Dachs, Optimizing transfection of primary human umbilical vein endothelial cells using commercially available chemical transfection reagents. *J Biomol Tech.* **2010**, *21*: 66-72.

68. E. Campeau *et al.*, A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS One.* **2009**, *4*: e6529.

69. R. H. Kutner, X. Y. Zhang, J. Reiser, Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc.* **2009**, *4*: 495-505.

70. S. L. Mayo, B. D. Olafson, W. A. Goddard III, DREIDING: a generic force field for molecular simulations. *J Phys Chem.* **1990**, *94*: 8897-8909.

71. V. W. Tak Kam, W. A. Goddard III, Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theory Comput*. **2008**, *4*: 2160-2169.

*Chapter 4*

# Deciphering the Importance of *O*-GlcNAc Glycosylation Using Chemical Tools

## 4.1 *O*-GlcNAc Glycosylation of Proteins

Although the human genome only contains between 19,000 and 20,000 genes,[1] cells have developed many mechanisms to increase the chemical diversity of protein structures. Methods of protein diversification can occur both before protein translation through the splicing of messenger RNA molecules and afterwards through chemical changes of amino acid residues, known as post-translational modifications (PTMs). Collectively, these diversifying pathways exponentially increase the number of individual protein chemical structures or "proteoforms."[2] In turn, structural diversity provides a mechanism to increase functional diversity, allowing cells to adapt to a variety of internal and external stimuli.

One type of PTM known as *O*-GlcNAc glycosylation or *O*-GlcNAcylation is remarkably suited as an environmental response element.[3-7] This modification utilizes uridine 5'-diphospho-*N*-acetylglucosamine (UDP-GlcNAc) to attach a single β-linked GlcNAc (GlcNAc) to Ser or Thr residues of many different proteins (Figure 4-1). Because the biosynthesis of UDP-GlcNAc incorporates precursors from nucleotide, carbohydrate, amino acid, and lipid biosynthesis,[5] it has been hypothesized that *O*-GlcNAcylation functions as an integrating signal for global metabolism,[8-10] allowing it to directly connect environmental conditions like nutrient availability with protein function.

*O*-GlcNAcylation is uncommon in both its localization, as one of the few known intracellular glycosylation events, and its enzymatic cycling.[11] PTMs

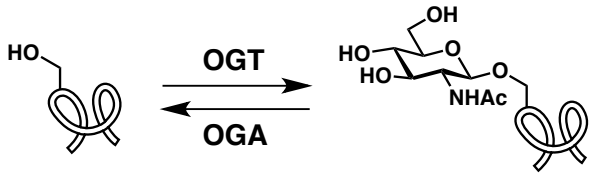



**Figure 4-1. *O*-GlcNAc cycling on proteins**. *O*-GlcNAc is added to Ser or Thr residues on intracellular proteins by OGT using UDP-GlcNAc and removed by OGA.

generally rely on a highly regulated network of writing and erasing enzymes. For example, protein phosphorylation occurs via a network of hundreds of kinases and phosphatases, each with their own substrate specificities. However, *O*-GlcNAcylation does not occur in such a partitioned fashion. Instead, only two enzymes have been discovered to cycle the *O*-GlcNAc PTM: *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA). These proteins work in an extremely widespread fashion, with a current tally of over 4,000 putatively *O*-GlcNAcylated proteins.[12] Unsurprisingly, no ubiquitous consensus sequence has been found for OGT recognition,[13] highlighting the promiscuity of OGT and the difficulty in predicting *O*-GlcNAc sites.[14] Interestingly, recent reports have further expanded the substrate scope of OGT to Cys residues as so-called *S*-GlcNAcylation events.[15, 16]

Given the breadth of *O*-GlcNAcylation substrates, this PTM accordingly has far-reaching effects on a variety of biological processes, including transcription, translation, signal transduction, autophagy, protein homeostasis, circadian rhythm, and metabolism.[8, 17-22] The molecular mechanisms by which *O*-GlcNAcylation exerts these different activities are also diverse. For example, *O*-

GlcNAcylation can compete with phosphorylation either directly by modification of the same residue or indirectly by modification of adjacent sites. The former mechanism can be observed on endothelial nitric oxide synthase (eNOS), in which the native Akt phosphorylation site Ser-1177 is blocked by *O*-GlcNAcylation during hyperglycemia to impair its activity.[23] The latter can be found on the tumor suppressor protein p53, on which *O*-GlcNAcylation at Ser-149 prevents phosphorylation at Ser-155 to prevent proteolytic targeting and degradation.[24] *O*-GlcNAcylation can also attenuate or countermand the activity of distal phosphorylation events. This phenomenon was detected on the transcription factor cAMP response element binding (CREB), where *O*-GlcNAcylation at Ser-40 is induced specifically on activated, Ser-133 phosphorylated CREB to lessen neuronal activity-dependent gene expression.[17] *O*-GlcNAcylation can also block protein-small molecule or protein-protein interactions. For example, phosphofructokinase 1 (PFK1) is *O*-GlcNAcylated at Ser-529 in the binding pocket of its allosteric activator fructose-2,6-bisphosphate (FBP), hindering the formation of the active tetrameric form of the enzyme.[8] In stem cells, transcription factor sex determining region Y-box 2 (Sox2) is *O*-GlcNAcylated at Ser-248, disrupting its interaction with poly-ADP ribose polymerase 1 (PARP1) to control self-renewal and differentiation.[25] Conversely, *O*-GlcNAcylation can promote protein-protein interactions; in the case of signal transducer and activator of transcription 5 (Stat5), *O*-GlcNAcylation at Thr-92 facilitates its interaction with CREB-binding protein (CBP) to enable gene

transcription.[26] In all of these cases, the delineation of the exact site of *O*-GlcNAcylation was instrumental in discovering the functional importance of the PTM. Therefore, it is crucial that robust methods be developed to enable *O*-GlcNAc site identification to better understand the roles of this modification throughout diverse biological processes.

## 4.2 Methods to Identify Sites of *O*-GlcNAcylation

### 4.2.1 Complications in the Analysis of *O*-GlcNAcylation

Identifying sites of *O*-GlcNAcylation is complicated by a number of biological and technical factors. Like other PTMs, protein *O*-GlcNAcylation is substoichiometric, making site identification on proteins that may be already poorly abundant even more difficult.[12] During MS analysis, signal intensity from *O*-GlcNAcylated peptides suffers from ion suppression in the presence of unmodified peptides.[27] Furthermore, the *O*-GlcNAc moiety is readily labile for common MS/MS ionization techniques used for PTM analysis like collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD).[28] Efforts on two fronts have helped to overcome these obstacles. First, the newer ionization technique known as electon transfer dissociation (ETD) has been developed as a milder MS/MS method that allows for the sequencing of peptides without loss of the *O*-GlcNAc group.[29] Second, a number of different enrichment techniques have been produced to both concentrate *O*-GlcNAcylated substrates and separate them from unmodified proteins and peptides.[30-32] These techniques

can be broadly separated into three categories: affinity enrichment, group replacement, and chemical functionalization.

### 4.2.2 Affinity Enrichment of *O*-GlcNAcylated Substrates

One chromatographic method to enrich *O*-GlcNAcylated substrates is known as lectin weak affinity chromatography (LWAC).[30] This method capitalizes on the affinity of natural plant proteins known as lectins for specific carbohydrate structures. In this process, a mixture of modified and unmodified substrates are isocratically passed over a chromatographic column that has wheat germ agglutinin (WGA) covalently attached to the stationary phase. This protein can weakly bind to terminal GlcNAc and sialic acid residues, causing these peptides to be retarded on the column compared to unmodified peptides. The lagging edge of the eluent is kept as the enriched *O*-GlcNAc fraction. As an alternative, succinylated WGA (sWGA) has been developed to no longer bind to sialic acid residues, providing further specificity for *O*-GlcNAcylated proteins.[33] However, this method also enriches for *N*-linked glycans that contain terminal GlcNAc moieties, complicating MS analysis. This can be ameliorated by pre-treatment of samples with PNGase F, which removes all *N*-linked glycans.[34] The natural binding mechanism for WGA requires multiple glycan structures, causing the structure to have a much weaker affinity for single *O*-GlcNAc groups ($K_D$ = ~10 mM for the free monosaccharide).[12] Therefore, longer columns (upwards of 3 m)[31] and/or multiple chromatographic enrichments[32] are often necessary to sufficiently enrich *O*-GlcNAcylated substrates. Nevertheless, this method has

been employed successfully since its development in 2006 in numerous cases to identify sites of *O*-GlcNAcylation.[30-32] The most successful application of this method in combination with downstream high pH reverse phase peptide fractionation identified 1,750 sites from murine synaptosomes, proving its usefulness as a method to identify sites of the *O*-GlcNAc modification.[32] However, the method has only been reported by a single laboratory, so it remains to be seen if LWAC will be more widely embraced by the community.

A second method of affinity enrichment using *O*-GlcNAc-specific antibodies has also been employed to identify *O*-GlcNAcylation sites.[28] A number of antibodies that recognize *O*-GlcNAc have been generated in the past, with the RL-2[35] and CTD110.6[36] antibodies being the most widely used. However, they generally suffer from either specificity for certain *O*-GlcNAcylated structures like RL-2 for nuclear pore complex proteins or recognition of other terminal GlcNAc moieties as is the case for CTD110.6, precluding their use for enriching all *O*-GlcNAcylated substrates over other GlcNAc-modified structures. Recently, three new general *O*-GlcNAc antibodies have been developed and used originally to enrich *O*-GlcNAcylated proteins.[37] Interestingly, the immunoprecipitated proteins differed for each antibody, highlighting the need to use a combination of these antibodies to enrich *O*-GlcNAcylated proteins in an unbiased fashion. In a subsequent report, enriched proteins from immunoprecipitates of HEK-293T cell lysate using all three antibodies were subjected to MS analysis, where 172 *O*-GlcNAcylated peptides were observed.[28]

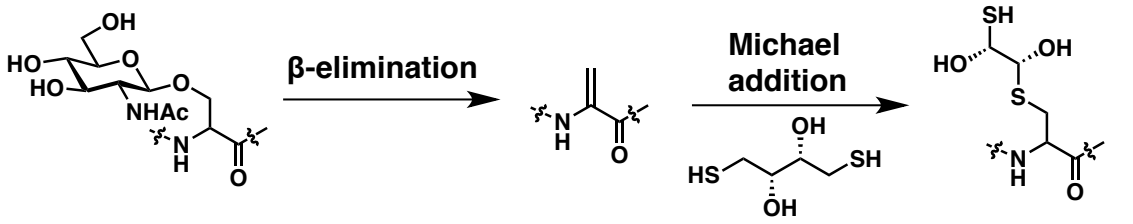



**Figure 4-2. Beta-elimination and Michael addition (BEMA).** BEMA occurs first by removal of *O*-GlcNAc using basic conditions to produce an α,β-unsaturated carbonyl group, which can then react with nucleophiles like DTT by Michael addition.

### 4.2.3 *O*-GlcNAc Enrichment by Replacement

The relative lability of the β-glycosidic linkage of *O*-GlcNAc has been exploited in an enrichment protocol termed β-elimination Michael addition (BEMA, Figure 4-2).[38-40] Originally designed to map phosphorylation sites, this method uses mild basic conditions to cause the elimination of *O*-GlcNAc on Ser and Thr residues and form an α,β-unsaturated carbonyl group, which can then be modified using a sulfhydryl nucleophile like DTT or biotinylated cystamine.[40, 41] Also known as BEMAD when using DTT, this method allows for enrichment of previously *O*-GlcNAcylated proteins and peptides via thiol-capture resin. Moreover, DTT isotopically labeled with $^{2}H$ can be used to differentiate samples for quantification.[42] DTT is stable to CID MS/MS fragmentation, eliminating the need for specialized MS techniques during peptide sequencing. BEMAD can suffer from cross-reactivity, as natural phosphorylated residues as well as alkylated Cys residues produced by reaction with iodoacetamide prior to MS analysis can also undergo β-elimination under stringent conditions.[40] However, treatment of samples with phosphatase prior to analysis and optimization of BEMAD

conditions have improved selectivity.[12] BEMAD has been applied in a variety of biological circumstances to identify *O*-GlcNAcylation sites, including in skeletal muscle, isolated proteasomes, and purified mitochondria.[41-43]

### 4.2.4 *O*-GlcNAc Enrichment by Chemical Functionalization

The third major method to enrich for *O*-GlcNAcylated substrates is through the chemical functionalization of the GlcNAc moiety. The structure of the GlcNAc residue is readily susceptible to direct modification; therefore, two main avenues have been developed to install reactive groups at the *O*-GlcNAc site: metabolic oligosaccharide engineering (MOE) and chemoenzymatic labeling.

MOE, also known as metabolic labeling, uses the endogenous biosynthetic machinery of the cell to incorporate non-natural sugar analogs that contain a chemical handle for further derivatization.[44-46] A number of different functional handles have been used to modify *O*-GlcNAc sites, including azide, alkyne, cyclopropene, and isonitrile groups.[47-50] This method does suffer from some drawbacks, including competition between natural and non-natural substrates leading to substoichiometric incorporation of the functional tag as well as crosstalk between biosynthetic salvage pathways causing the functional tag to be incorporated into a number of different glycan structures. This is epitomized by the discovery that UDP-galactose 4'-epimerase (GALE) interconverts UDP-GlcNAc and UDP-*N*-acetylgalactosamine (GalNAc) as well as the corresponding non-natural sugars.[51] In fact, treatment of cells with peracetylated *N*-azidoacetylgalactosamine ($Ac_4$GalNAz, Figure 4-3) facilitated labeling of *O*-

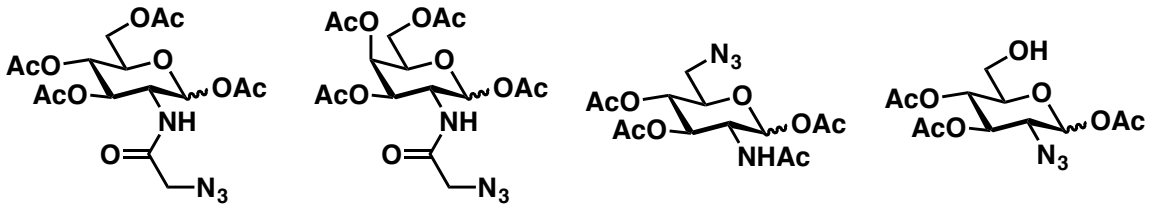



**Figure 4-3. Common carbohydrates used in MOE.**

GlcNAcylated structures better than treatment with the GlcNAc analog. Moreover, the non-natural sugar can incorporate into any glycan that utilizes GlcNAc, including *N*-linked glycans. Recent advances have been made to provide chemical probes that demonstrate selectivity for *O*-GlcNAc over other structures. For example, installation of the azide group at the C6 position of GlcNAc to produce peracetylated 6-azido-6-deoxy-*N*-acetylglucosamine ($Ac_4$6AzGlcNAc)[52] or at the C2 position of glucose (Glc) to produce acetylated 2-azido-2-deoxy-glucose ($Ac_3$2AzGlc)[53] provided labeling of *O*-GlcNAcylated substrates with minimal surface labeling, suggesting that these probes do not incorporate into *N*-linked glycans like previous probes. However, neither of these specific probes has been used to map sites of *O*-GlcNAcylation. In any case, the installed functional handle can then be elaborated using bioorthogonal chemistry to install an enrichment marker, which is discussed further below.

An alternative way to install a chemical handle for functionalization is through the use of a permissive enzyme via chemoenzymatic labeling (Figure 4-4).[12, 54, 55] In this technique, a mutated form of bovine galactosyltransferase (Y289L GalT)[56]

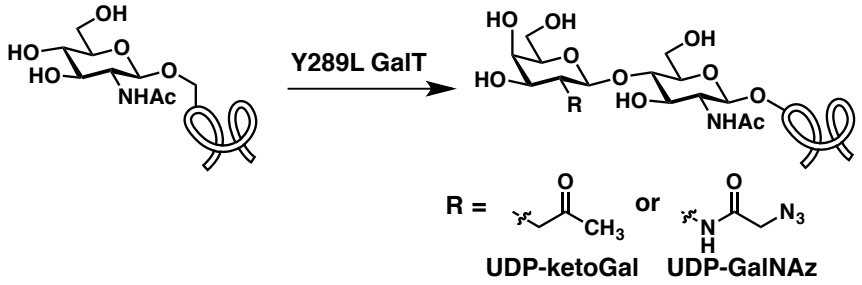



**Figure 4-4. Chemoenzymatic labeling**. *O*-GlcNAcylated substrates are modified by Y289L GalT with a non-natural UDP-galactose sugar. This approach has been used to append a ketone (UDP-ketoGal) or an azide (UDP-GalNAz) functional handle.

is used to modify terminal GlcNAc moieties with a non-natural UDP-galactose (Gal) or UDP-GalNAc sugar. This method was first used to install a ketone handle with UDP-2-acetonyl-2-deoxygalactose (UDP-ketoGal)[55, 57, 58] and more recently has been employed with the azide-containing UDP-GalNAz.[59] This method has benefits over MOE including the stoichiometric addition of the functional handle, the lack of perturbation of the system by addition of high concentrations of exogenous carbohydrates, and the specificity for GlcNAc structures. As with WGA, Y289L GalT will recognize any terminal GlcNAc moiety,[56] but treatment of samples with PNGase F to remove *N*-linked glycans or subcellular fractionation eliminate this cross-reactivity. As with MOE, this method installs a functional handle that allows for further derivatization and enrichment.

Currently, the majority of enrichment tags for *O*-GlcNAc site identification are installed onto an azide- or alkyne-tagged version of GlcNAc using copper-catalyzed azide-alkyne cycloaddition (CuAAC). However, a variety of enrichment tags have been used. For example, addition of 4-pentynyl phosphate has been

used in combination with chemoenzymatic labeling to install a phosphorylation mimicking tag that can be purified by titanium dioxide affinity chromatography, revealing 42 sites of *O*-GlcNAcylation from mouse brain lysate.[60] In a second approach using chemoenzymatic labeling, ethynylbenzaldehyde was added via CuAAC and used with hydrazide resin for peptide enrichment; however, this method was validated only with α-crystallin.[61] Finally, a third method utilizing metabolic labeling reacted modified proteins directly to alkyne-containing agarose resin.[62] After on-bead digestion and release of the modified proteins by β-elimination revealed 185 modification sites from HEK-293T lysate.

The most common enrichment method used with MOE and chemoenzymatic labeling is the installation of a biotin group combined with streptavidin affinity purification. One of the first instances of this method to identify *O*-GlcNAc sites was employed in our own lab in combination with chemoenzymatic labeling using UDP-ketoGal.[58] Here, 20 *O*-GlcNAcylated peptides were discovered from rat brain lysate, and incorporation of a peptide isotopic labeling technique known as dimethyl labeling allowed for quantification of modification dynamics on a subset of these sites. Although successful here, the relatively large size of the biotin tag likely hinders peptide ionization and sequencing by MS. Therefore, newer techniques have explored the incorporation of cleavable biotin moieties to reduce the residual size of the appended tag.

The first cleavable biotin tag used to map *O*-GlcNAc sites incorporated a 2-nitrobenzyl group, which can be selectively cleaved by UV irradiation with 365

nm light.[27] This tag leaves only the aminomethyltriazolyl group appended to GalNAz, significantly decreasing the added size from biotinylated compared to non-cleavable variants. Moreover, the amine residue imbues the peptide with an additional positive charge, which can facilitate MS sequencing by ETD. However, cleavage by UV irradiation was shown to be incomplete.[27] Nevertheless, this approach was successfully applied to both purified mitotic spindles and cerebrocortical brain tissue to identify 141 and 458 *O*-GlcNAc sites, respectively.[63, 64]

A second cleavable tag uses the diphenylsilyl diether group, which can be easily removed by mild acid.[65, 66] After reaction with this tag also containing a dibrominated core to provide an isotopic signature in MS1, metabolically labeled proteins were immobilized on streptavidin resin and subjected to on-bead digestion. The remaining modified peptides were cleaved by mild acid treatment and sequenced, providing 357 unique glycopeptides modified by either GalNAz or *N*-azidoacetylmannosamine (ManNAz).[66] This method was recently expanded to use a mixture of $^{1}$H- and $^{2}$H-labeled probes that also present an isotopic signature in MS.[67] Here, 1765 mono- and di-*O*-GlcNAcylated mass signatures were identified over 15 samples corresponding to 379 uniquely *O*-GlcNAcylation sites. However, 36% of the modified peptides were only identified by the isotopic signature and were unable to be sequenced, underscoring the difficulty in sequencing *O*-GlcNAcylated peptides. Nevertheless, these results collectively demonstrate that the use of a cleavable

biotin tag is a useful approach to enrich for *O*-GlcNAcylated substrates and map modification sites. Newer methods should take advantage of benefits from multiple approaches like minimizing the residual tag left after cleavage while simultaneously improving on shortcomings of current approaches like incomplete cleavage and poor ETD fragmentation.

### 4.3 References


1. I. Ezkurdia *et al.* Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Gen.* **2014**, *23*: 5866-5878.
2. L. M. Smith, N. L. Kelleher, The Consortium for Top Down Proteomics, Proteoform: a single term describing protein complexity. *Nat Methods.* **2013**, *10*: 186-187.
3. N. E. Zachara, G. W. Hart. The emerging significance of *O*-GlcNAc in cellular regulation. *Chem Rev*. **2002**, *102*: 431-438.
4. C. Slawson, G. W. Hart. *O*-GlcNAc signalling: implications for cancer cell biology. *Nat Rev Cancer*. **2011**, *11*: 678-684.
5. J. A. Hanover, M. W. Krause, D. C. Love. Bittersweet memories: linking metabolism to epigenetics through *O*-GlcNAcylation. *Nat Rev Mol Cell Biol.* **2012**, *13*: 312-321.
6. M. R. Bond, J. A. Hanover. A little sugar goes a long way: the cell biology of *O*-GlcNAc. *J Cell Biol.* **2015**, *208*: 869-880.
7. X. Yang, K. Qian. Protein *O*-GlcNAcylation: emerging mechanisms and functions. *Nat Rev Mol Cell Biol.* **2017**.

8. W. Yi. *et al.* Phosphofructokinase 1 Glycosylation Regulates Cell Growth and Metabolism. *Science*. **2012**, *337*: 975-980.

9. J. W. Bullen *et al.* Cross-talk between two essential nutrient-sensitive enzymes: *O*-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *J Biol Chem*. **2014**, *289*: 10592-10606.

10. X. Rao *et al. O*-GlcNAcylation of G6PD promotes the pentose phosphate pathway and tumor growth. *Nat Commun*. **2015**, *6*: 8468.

11. G. W. Hart, Y. Akimoto. *The O-GlcNAc Modification*. In: A. Varki *et al.*, Eds., *Essentials of Glycobiology*. 2nd Ed. Cold Spring Harbor (NY): 2009.

12. J. Ma, G. W. Hart. *O*-GlcNAc profiling: from proteins to proteomes. *Clin. Proteomics*. **2014**, *11*: 8.

13. S. A. Whelan, G. W. Hart. Identification of *O*-GlcNAc sites on proteins. *Methods Enzymol*. **2006**, *415*: 113-133.

14. R. Jochmann, P. Holz, H. Sticht, M. Sturzl, Validation of the reliability of computational *O*-GlcNAc prediction. *Biochim Biophys Acta*. **2014**, *1844*: 416-421.

15. J. C. Maynard, A. L. Burlingame, K. F. Medzihradszky. Cysteine *S*-linked *N*-acetylglucosamine (*S*-GlcNAcylation), A new post-translational nodification in nammals. *Mol Cell Proteomics*. **2016**, *15*: 3405-3411.

16. H. Xiao, R. Wu. Global and site-specific analysis revealing unexpected and extensive protein *S*-GlcNAcylation in human cells. *Anal Chem*. **2017**, *89*: 3656-3663.

17. J. E. Rexach *et al.* Dynamic *O*-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nat Chem Biol*. **2012**, *8*: 253-261.

18. Y. Zhu *et al. O*-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nat Chem Biol*. **2015**, *11*: 319-325.

19. P. Ramakrishnan *et al.* Activation of the transcriptional function of the NF-κB protein c-Rel by *O*-GlcNAc glycosylation. *Sci Signal.* **2013**, *6*: ra75.

20. B. Guo *et al. O*-GlcNAc-modification of SNAP-29 regulates autophagosome maturation. *Nat Cell Biol.* **2014**, *16*: 1215-1226.

21. F. Zhang *et al. O*-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell.* **2003**, *115*: 715-725.

22. M. D. Li *et al. O*-GlcNAc signaling entrains the circadian clock by inhibiting BMAL1/CLOCK ubiquitination. *Cell Metab.* **2013**, *17*: 303-310.

23. B. Musicki, M. F. Kramer, R. E. Becker, A. L. Burnett. Inactivation of phosphorylated endothelial nitric oxide synthase (Ser-1177) by *O*-GlcNAc in diabetes-associated erectile dysfunction. *Proc Natl Acad Sci USA*. **2005**, *102*: 11870-11875.

24. W. H. Yang *et al.* Modification of p53 with *O*-linked N-acetylglucosamine regulates p53 activity and stability. *Nat Cell Biol.* **2006**, *8*: 1074-1083.

25. S. A. Myers *et al.* SOX2 *O*-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *Elife.* **2016**, *5*: e10647.

26. C. Gewinner *et al.* The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J Biol Chem.* **2004**, *279*: 3563-3572.

27. Z. Wang *et al.* Enrichment and site mapping of *O*-linked *N*-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol Cell Proteomics.* **2010**, *9*: 153-160.

28. P. Zhao *et al.* Combining high-energy C-trap dissociation and electron transfer dissociation for protein *O*-GlcNAc modification site assignment. *J Proteome Res.* **2011**, *10*: 4088-4104.

29. S. A. Myers, S. Daou, E. B. Affar, A. L. Burlingame. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments—a study of the *O*-GlcNAcylated protein host cell factor C1. *Proteomics.* **2013**, *13*: 982-991.

30. K. Vosseller *et al. O*-linked *N*-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics.* **2006**, *5*: 923-934.

31. R. J. Chalkley, A. Thalhammer, R. Schoepfer, A. L. Burlingame, Identification of protein *O*-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci USA*. **2009**, *106*: 8894-8899.

32. J. C. Trinidad *et al.* Global identification and characterization of both *O*-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics.* **2012**, *11*: 215-229.

33. M. Monsigny *et al.* Properties of succinylated wheat-germ agglutinin. *Eur J Biochem.* **1979**, *98*: 39-45.

34. X. Wei, L. Li. Comparative glycoproteomics: approaches and applications. *Brief Funct Genomic Proteomic.* **2009**, *8*: 104-113.

35. C. M. Snow, A. Senior, L. Gerace. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. *J Cell Biol.* **1987**, *104*: 1143-1156.

36. F. I. Comer, K. Vosseller, L. Wells, M. A. Accavitti, G. W. Hart. Characterization of a mouse monoclonal antibody specific for *O*-linked *N*-acetylglucosamine. *Anal Biochem.* **2001**, *293*: 169-177.

37. C. F. Teo *et al.* Glycopeptide-specific monoclonal antibodies suggest new roles for *O*-GlcNAc. *Nat Chem Biol.* **2010**, *6*: 338-343.

38. H. E. Meyer, E. Hoffmann-Posorske, H. Korte, L. M. Heilmeyer, Jr. Sequence analysis of phosphoserine-containing peptides. Modification for picomolar sensitivity. *FEBS Lett*. **1986**, *204*: 61-66.

39. Y. Oda, T. Nagasu, B. T. Chait. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol.* **2001**, *19*: 379-382.

40. L. Wells *et al.* Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics.* **2002**, *1*: 791-804.

41. T. Overath *et al.* Mapping of *O*-GlcNAc sites of 20S proteasome subunits and Hsp90 by a novel biotin-cystamine tag. *Mol Cell Proteomics.* **2012**, *11*: 467-477.

42. J. Ma *et al.* Comparative proteomics reveals dysregulated mitochondrial *O*-GlcNAcylation in diabetic hearts. *J Proteome Res.* **2016**, *15*: 2254-2264.

43. J. Hedou, B. Bastide, A. Page, J. C. Michalski, W. Morelle. Mapping of *O*-linked β-*N*-acetylglucosamine modification sites in key contractile proteins of rat skeletal muscle. *Proteomics.* **2009**, *9*: 2139-2148.

44. K. N. Chuh, A. R. Batt, M. R. Pratt. Chemical methods for encoding and decoding of posttranslational modifications. *Cell Chem Biol.* **2016**, *23*: 86-107.

45. K. K. Palaniappan, C. R. Bertozzi. Chemical glycoproteomics. *Chem Rev.* **2016**, *116*: 14277-14306.

46. M. E. Griffin, L. C. Hsieh-Wilson, Glycan engineering for cell and developmental biology. *Cell Chem Biol.* **2016**, *23*: 108-121.

47. S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi. *In vivo* imaging of membrane-associated glycans in developing zebrafish. *Science*. **2008**, *320*: 664-667.

48. B. W. Zaro, Y.-Y. Yang, H. C. Hang, M. R. Pratt. Chemical reporters for fluorescent detection and identification of *O*-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc Natl Acad Sci USA*. **2011**, *108*: 8146-8151.

49. D. M. Patterson, K. A. Jones, J. A. Prescher. Improved cyclopropene reporters for probing protein glycosylation. *Mol Biosyst*. **2014**, *10*: 1693-1697.

50. S. Stairs *et al.* Metabolic glycan imaging by isonitrile-tetrazine click chemistry. *ChemBioChem*. **2013**, *14*: 1063-1067.

51. M. Boyce *et al.* Metabolic cross-talk allows labeling of *O*-linked β-*N*-acetylglucosamine-modified proteins via the *N*-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci USA*. **2011**, *108*: 3141-3146.

52. K. N. Chuh *et al.* Changes in metabolic chemical reporter structure yield a selective probe of *O*-GlcNAc modification. *J Am Chem Soc*. **2014**, *136*: 12283-12295.

53. B. W. Zaro, A. R. Batt, K. N. Chuh, M. X. Navarro, M. R. Pratt. The small molecule 2-azido-2-deoxy-glucose is a metabolic chemical reporter of *O*-GlcNAc modifications in mammalian cells, revealing an unexpected promiscuity of *O*-GlcNAc transferase. *ACS Chem Biol*. **2017**, *12*: 787-794.

54. M. E. Griffin *et al.* Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag. *Mol Biosyst*. **2016**, *12*: 1756-1759.

55. N. Khidekel *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc*. **2003**, *125*: 16162-16163.

56. B. Ramakrishnan, P. K. Qasba. Structure-based design of beta 1,4-galactosyltransferase I (β4Gal-T1) with equally efficient *N*-acetylgalactosaminyltransferase activity: point mutation broadens β4Gal-T1 donor specificity. *J Biol Chem.* **2002**, *277*: 20833-20839.

57. N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA*. **2004**, *101*: 13132-13137.

58. N. Khidekel *et al.* Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol.* **2007**, *3*: 339-348.

59. P. M. Clark *et al.* Direct in-gel fluorescence detection and cellular imaging of *O*-GlcNAc-modified proteins. *J Am Chem Soc*. **2008**, *130*: 11576-11577.

60. B. L. Parker, P. Gupta, S. J. Cordwell, M. R. Larsen, G. Palmisano. Purification and identification of *O*-GlcNAc-modified peptides using phosphate-based alkyne CLICK chemistry in combination with titanium dioxide chromatography and mass spectrometry. *J Proteome Res.* **2011**, *10*: 1449-1458.

61. T. Nishikaze, S. Kawabata, S. Iwamoto, K. Tanaka. Reversible hydrazide chemistry-based enrichment for *O*-GlcNAc-modified peptides and glycopeptides having non-reducing GlcNAc residues. *Analyst*. **2013**, *138*: 7224-7232.

62. H. Hahne *et al.* Proteome wide purification and identification of *O*-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J Proteome Res.* **2013**, *12*: 927-936.

63. Z. Wang *et al.* Extensive crosstalk between *O*-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal.* **2010**, *3*: ra2.

64. J. F. Alfaro *et al.* Tandem mass spectrometry identifies many mouse brain *O*-GlcNAcylated proteins including EGF domain-specific *O*-GlcNAc transferase targets. *Proc Natl Acad Sci USA*. **2012**, *109*: 7280-7285.

65. J. Szychowski *et al.* Cleavable biotin probes for labeling of biomolecules via azide–alkyne cycloaddition. *J Am Chem Soc*. **2010**, *132*: 18351-18360.

66. C. M. Woo, A. T. Iavarone, D. R. Spiciarich, K. K. Palaniappan, C. R. Bertozzi. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact *N*- and *O*-glycopeptide discovery and analysis. *Nat Methods*. **2015**, *12*: 561-567.

67. C. M. Woo *et al.*, Development of IsoTaG, a chemical glycoproteomics technique for profiling intact *N*- and *O*-glycopeptides from whole cell proteomes. *J Proteome Res*. **2017**, *16*: 1706-1718.

*Chapter 5*

# Chemically Cleavable Tagging Method for the Enrichment and Detection of *O*-GlcNAc Glycosylated Proteins and Peptides

Portions of this chapter are published as:

**Griffin ME**, Jensen EH, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. "Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag." *Mol. Biosys.* **2016**, *12*: 1756-1759.

doi: 10.1016/j.chembiol.2015.12.007. Research article.

**5.1 Abstract**

Post-translational modification by *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) plays many important roles in controlling protein function. However, relatively few *O*-GlcNAc modification sites have been mapped due to the difficulty of enriching and detecting *O*-GlcNAcylated peptides from complex samples. Here we describe an improved approach to quantitatively label and enrich *O*-GlcNAcylated proteins for site identification. Chemoenzymatic labeling followed by copper(I)-catalyzed azide–alkyne cycloaddition (CuAAC) installs a new mass spectrometry (MS)-compatible linker designed for facile purification of *O*-GlcNAcylated proteins from cell lysates. The linker also allows subsequent quantitative release of *O*-GlcNAcylated proteins for downstream MS analysis. We validate the approach by unambiguously identifying several established *O*-GlcNAc sites on the proteins α-crystallin and *O*-GlcNAc transferase (OGT), as well as discovering new, previously unreported sites on OGT. Notably, these novel sites on OGT lie in key functional domains of the protein, underscoring how this site identification method may reveal important biological insights into protein activity and regulation. We then describe the further optimization of this method to identify sites of *O*-GlcNAc modification on native proteins from complex lysate. Together, these experiments lay the groundwork for a straightforward method to label and enrich *O*-GlcNAcylated proteins and peptides for MS identification.

## 5.2 General Approach

Understanding the roles of *O*-GlcNAcylation in specific physiological contexts will require a more comprehensive characterization of the *O*-GlcNAc proteome and the modification sites on proteins. Notably, although thousands of proteins have been putatively shown to be *O*-GlcNAcylated,[1-7] relatively few glycosylation sites have been mapped. Mass spectrometric (MS) identification of *O*-GlcNAcylated peptides from complex mixtures has been challenging due to the substoichiometric nature of *O*-GlcNAcylation and is further exacerbated by suppression of *O*-GlcNAc peptide ionization in the presence of the unmodified peptide.[8] Thus, improved methods to enrich *O*-GlcNAcylated peptides or proteins are much needed, particularly approaches that can be directly used in conjunction with MS/MS sequencing to achieve a more comprehensive understanding of *O*-GlcNAc modification sites.

Robust enrichment of *O*-GlcNAcylated proteins can be accomplished using a two-step chemoenzymatic approach (Figure 5-1).[3, 9] First, the *O*-GlcNAc moiety is tagged with a non-natural azide group by treatment of cell lysates with UDP-GalNAz and a mutant galactosyltransferase (Y289L GalT)[10] that specifically recognizes terminal GlcNAc moieties. Next, a biotin group is attached via copper(I)-catalyzed azide–alkyne cycloaddition (CuAAC),[11] which allows for affinity purification. Although a limited set of alkyne-biotin linkers are commercially available, many existing linkers are not ideal for mapping *O*-

GlcNAc modification sites. In particular, harsh conditions are usually required to

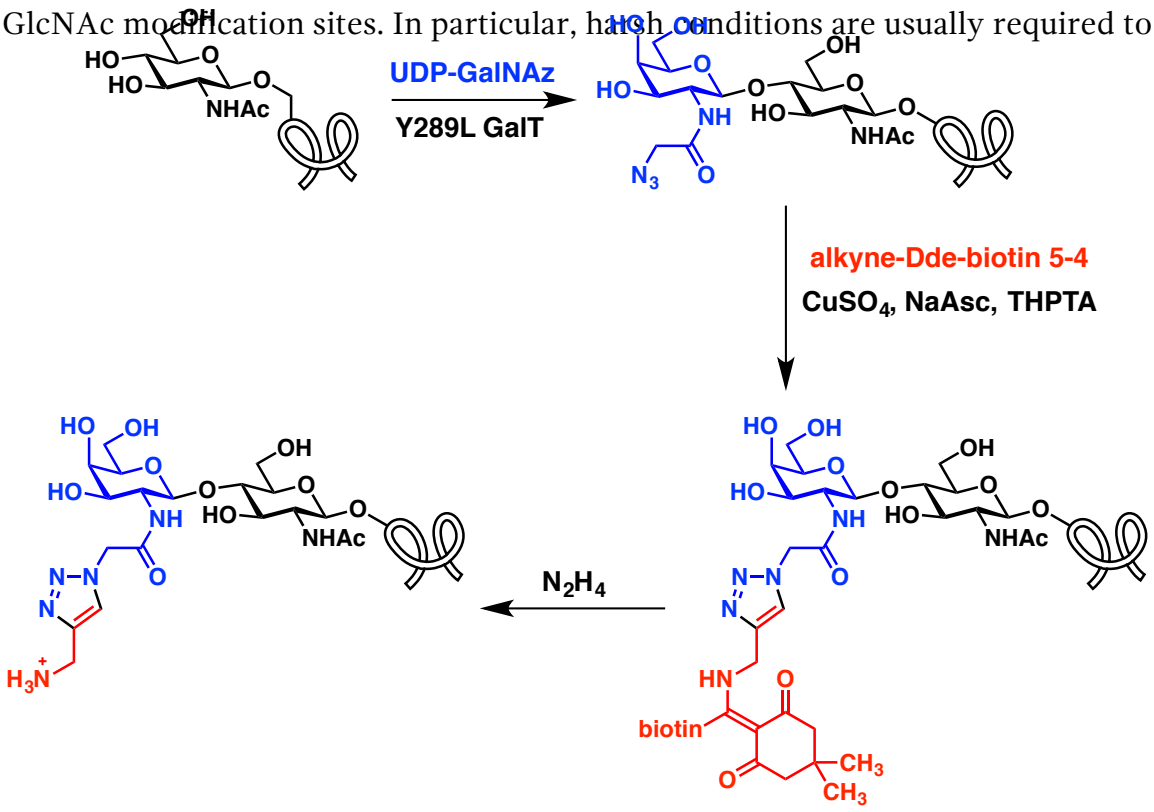



**Figure 5-1. Two-step chemoenzymatic approach.** *O*-GlcNAcylated substrates are modified by Y289L GalT with UDP-GalNAz to install an azide handle, which can be functionalized with a cleavable biotin linker using CuAAC. After enrichment, labeled *O*-GlcNAcylated substrates can be released by facile chemical cleavage with hydrazine.

disrupt the femtomolar biotin–streptavidin interaction,[12] which may hydrolyze the labile *O*-GlcNAc moiety. Additionally, many linkers contain a large spacer between the biotin group and the alkyne functionality,[12] which appends a relatively large mass to the glycopeptide and can preclude its identification. Therefore, a facile method to release the labeled peptides and proteins with minimal added mass would greatly facilitate downstream analysis.

Several cleavable linkers have been previously developed for enrichment of

O-GlcNAcylated proteins.[6-8, 13] However, each suffers from significant drawbacks for site identification. For example, a photocleavable linker (alkyne-PC-biotin **5-1**, Figure 5-2) was employed in conjunction with UDP-GalNAz and Y289L GalT to sequence modified peptides from mouse brain lysate.[7, 8] Importantly, the moiety retained after cleavage provided a positively-charged amine group, which increased the overall peptide charge and facilitated ionization by electron-transfer dissociation (ETD), the most successful MS/MS method for *O*-GlcNAc peptide sequencing.[4, 14] Unfortunately, cleavage of the linker was found to be incomplete.[8] In a recent report, a dibromine-containing, acid-cleavable linker (**5-2**, Figure 5-2) was employed to identify various glycan modifications including *O*-GlcNAc.[13] However, cleavage of the linker revealed only a neutral hydroxyl group, and the halogenated glycopeptides demonstrated poor fragmentation efficiency using ETD. Recently, this linker was updated to use a mixture of $^{1}$H- and $^{2}$H to produce a similar isotopic signature (**5-3**, Figure 5-2); however, 36% of the modified peptides were only identified by the isotopic pattern and were unassignable to a specific sequence.[15] The analysis was further complicated by GALE-mediated epimerization of GalNAz into GlcNAz,[16] which allows simultaneous labeling of *O*-GalNAcylated and *O*-GlcNAcylated peptides that cannot be distinguished by MS and require manual annotation based on subcellular localization. Therefore, we aimed to develop a tagging approach that would be quantitatively appended and released, specifically modify only *O*-GlcNAcylation sites, and incorporate a positive charge upon cleavage to

facilitate ETD-MS detection.

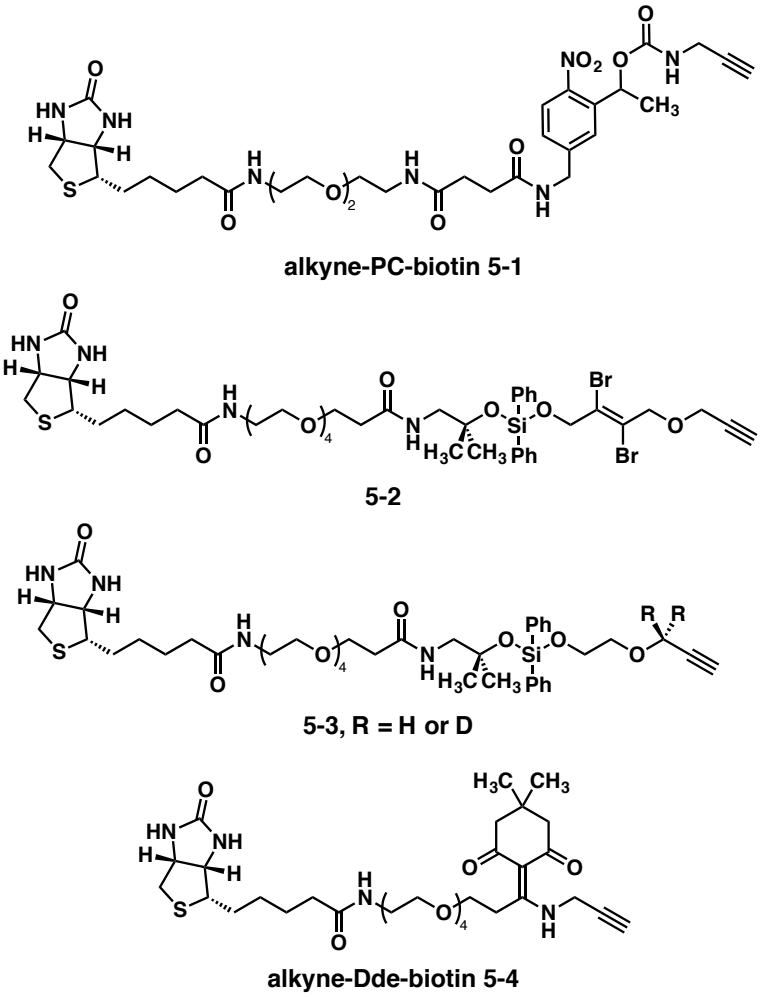



**Figure 5-2. Cleavable biotin tags used for *O*-GlcNAc site identification.**

## 5.3 Method Development

### 5.3.1 Validation of Chemically Cleavable Linker

To achieve these goals, we chose to examine the 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl (Dde) functional group. The Dde moiety has been used extensively as a protecting group for lysine in peptide synthesis,[17]

demonstrating its compatibility with biomolecules. The group is stable to both acid and base and can be quantitatively removed by hydrazine.[18] However, it

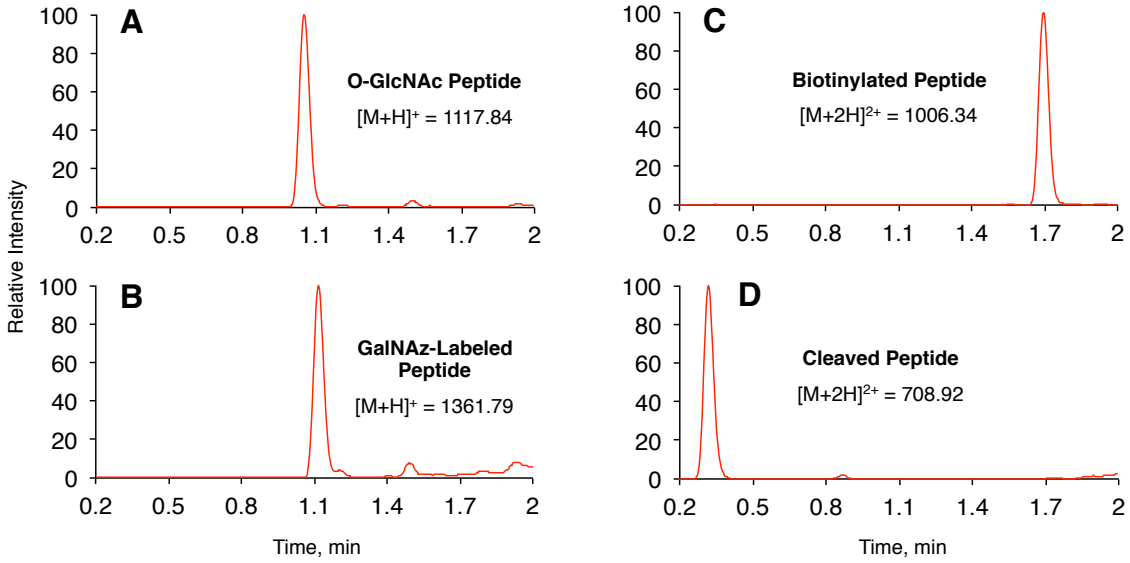
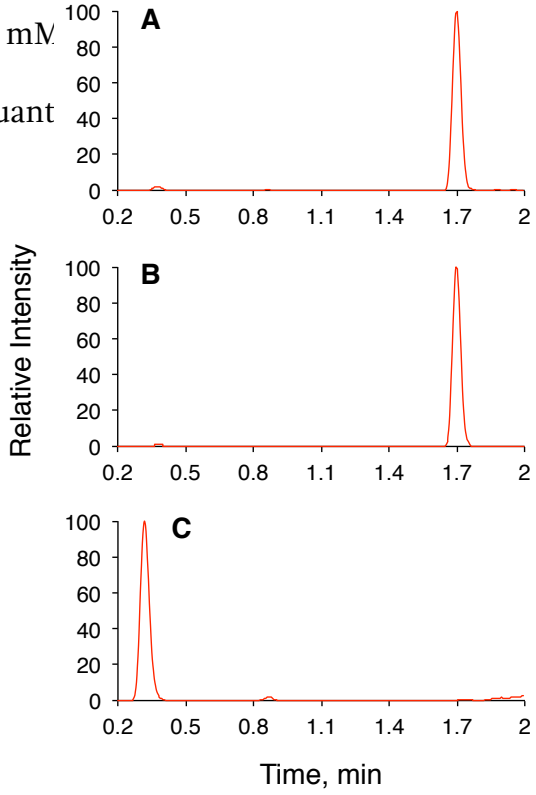



**Figure 5-3. Validation of Dde linker 5-4.** Reverse phase LC-MS analysis of O-GlcNAc peptide labelling reactions at (A) time 0, (B) 16 h after addition of 1 and Y289L GalT, (C) 1 h after CuAAC with 2, and (D) 1 h after cleavage with 2% aqueous hydrazine. (A) and (B) show base peak chromatograms. (C) and (D) show extracted ion chromatograms of the starting material and product within 1 m/z of calculated values.

was reported that the Dde group is incompatible with sodium dodecyl sulphate (SDS) and amine-containing buffers, common additives to protein labeling protocols.[19]

We first investigated the labeling of a model O-GlcNAcylated peptide with commercially available alkyne-Dde-biotin (**5-4**, Figure 5-2) followed by cleavage of the linker using liquid chromatography (LC)-MS (Figure 5-3). Commercially available peptide TAPT(gS)TIAPG (Figure 5-3a), where gS is the O-GlcNAcylated residue, was incubated with 100 ng mL Y289L GalT and 1 mM

UDP-GalNAz in 10 mM




ernight at 4 °C. LC-MS analysis revealed quant alNAz- labeled

**Figure 5-4. Stability of Dde linker 5-4.** Reverse-phase LC-MS analysis of alkyne-Dde-biotin-labelled peptide after 1 h incubation with (A) 6 M urea, (B) 1% RapiGest, or (C) 2% hydrazine. All graphs show extracted ion chromatograms of the starting material and possible product within ± 1 *m/z*.

product (Figure 5-3b). Next, the azide-containing peptide was reacted with 100 mM of **5-4** in 10 mM sodium phosphate pH 7.6 containing 2 mM sodium ascorbate (NaAsc), 100 mM THPTA, and 1 mM $CuSO_4$. After 1 h, stoichiometric biotinylation of the peptide was observed (Figure 5-3c). Treatment with 2% aqueous hydrazine for 1 h at RT resulted in quantitative cleavage of the linker to afford a minimal, positively charged aminomethyltriazolyl group (Figure 5-3d).

To test whether the linker would be stable under stringent wash conditions, we incubated the labeled peptide with 1% RapiGest, a MS-compatible analogue of SDS, or 6 M urea for 1 h at RT (Figure 5-4). In both cases, the linker remained intact, highlighting the compatibility of the linker with rigorous washing steps.

### 5.3.2 Comparing Dde and PC Linkers

We next tested the performance of our linker in comparison to the previously described, widely utilized photocleavable linker alkyne-PC-biotin **5-1**.[8] Briefly, HEK-293T cell lysate was subjected to chemoenzymatic labeling with UDP-GalNAz using Y289L GalT as described above. The azide-labeled protein was then split into two equal fractions and reacted with either **5-1** or **5-4** by CuAAC. An aliquot of each sample was reserved for analysis, and the remainder of each sample was subjected to cleavage using 2% hydrazine monohydrate or by UV irradiation at 365 nm. The samples were resolved by SDS-PAGE and probed for biotin using streptavidin conjugated to AlexaFluor 680 dye (Figure 5-5). Notably, a stronger biotin signal was observed for lysate labeled with **5-4** compared to **5-1**, suggesting higher labeling efficiency with the Dde linker. Furthermore, although both linkers cleaved well, the **5-1** showed slightly higher residual signal compared to **5-4**, suggesting that the Dde moiety was also released more efficiently than the photocleavable group. These results demonstrate that our new approach provides an improvement in both labeling efficiency and recovery of O-GlcNAcylated proteins compared to the most

widely used method.

### 5.3.3 Validation Using Known *O*-GlcNAcylated Proteins

We then evaluated the potential of the approach to pull down known O-

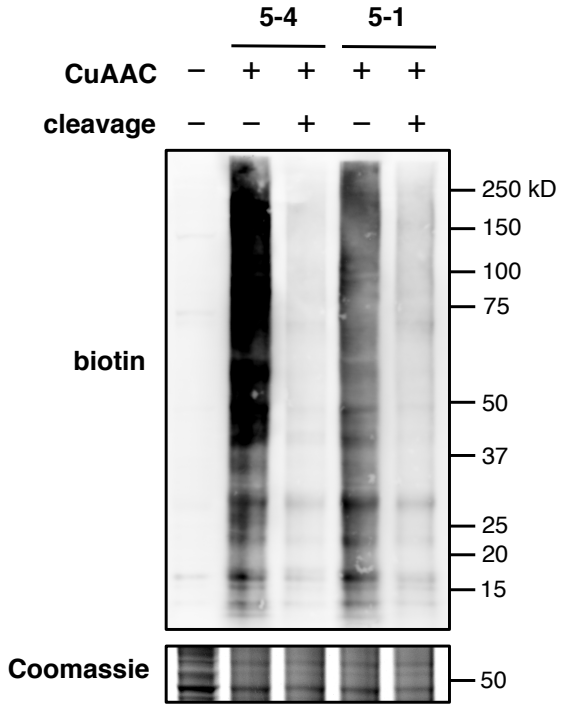



**Figure 5-5. Comparison of Dde linker 5-4 and PC linker 5-1.** Protein lysate was labeled using the chemoenzymatic approach using **5-4** or **5-1** and then resolved by SDS-PAGE. Protein lysates labelled with **5-4** show higher biotin signal after labelling (lane 2 vs. lane 4) and lower residual signal after cleavage (lane 3 vs. lane 5) compared to the PC linker.

GlcNAcylated proteins and identify sites of modification. The well-characterized O-GlcNAcylated protein α-crystallin was selected to assess the sensitivity of the method because it has a relatively low glycosylation stoichiometry (<10%).[20] Short-form OGT (sOGT)[21, 22] from Sf9 cells has multiple sites of O-GlcNAcylation[4] and was thus used to determine whether comprehensive site mapping could be achieved. To test the robustness of our method in a complex

mixture, each protein was added to 200 mg of adult mouse cortical lysate and 10 subjected to chemoenzymatic labeling and CuAAC using **5-2**. The labeled proteins were applied to high-capacity Neutravidin resin and washed with 0.5 mL of 1% SDS, 6 M urea, and phosphate buffered saline (PBS). The resin was then incubated for 1 h with 2% aqueous hydrazine to cleave the O-GlcNAcylated proteins from the resin. Eluted samples were precipitated, re-dissolved in denaturing buffer and subjected to reduction, alkylation, and proteolytic digestion. Digested peptides were separated by nanoLC-MS and analyzed on an LTQ-Velos by a combination of collision-induced dissociation (CID) and ETD-MS.

Impressively, a large number of O-GlcNAcylation sites were identified on α-crystallin and sOGT (Table 5-1). The known O-GlcNAc site on α-crystallin A (Ser-162)[23] was readily recognized despite the low abundance of the O-GlcNAc modification at this site. Importantly, we observed both known and novel sites on sOGT.[4, 24] For example, we identified the previously reported Thr-662 site, which is found in the catalytic domain of sOGT. The new linker design also revealed a number of new O-GlcNAcylation sites within the N-terminal tetratricopeptide repeat-containing (TPR) domains (Ser-10/Thr-12, Ser-20, Ser-52, and Ser-56) of sOGT, and we observed a doubly modified peptide at both Ser-10/Thr-12 and Ser-20, highlighting the sensitivity of the approach to identify novel glycosylation sites and multiply occupied states. As the TPR domains of OGT are thought to mediate protein-protein interactions,[25-27] such

modifications could play an integral role in OGT regulation and may provide a mechanism to selectively modulate its activity toward specific substrates.

| Protein | Peptide sequence | Site(s) | Mascot ion score | Mascot delta ion score | MS |
|---|---|---|---|---|---|
| αCryA | AIPV*S*REEKPSSAPSS | S162 | 24.9 | 23.5 | ETD |
| sOGT | I*SPT*FADAYSNMGNTLK | S10*/T12* | 46.5 | – | ETD |
| sOGT | ISPTFADAY*S*NMGNTLK | S20* | 21.6 | 13.6 | ETD |
| sOGT | I*SPT*FADAY*S*NMGNTLK | S10*/T12*, S20* | 38.4 | – | ETD |
| sOGT | EMQDVQGALQCY*T*R | T38 | 41.8 | 35.0 | CID |
| sOGT | AIQINPAFADAH*S*NLASIHKDSGNIPEALASYR | S52* | 53.5 | 7.9 | ETD |
| sOGT | AIQINPAFADAHSNLA*S*IHKDSGNIPEALASYR | S56* | 56.8 | 15.7 | ETD |
| sOGT | LYLQMWEHYAAGNKPDHMIKPVEV*T*ESA | T662 | 33.1 | 8.0 | ETD |

**Table 5-1. Validation of 5-4 to identify *O*-GlcNAc sites.** *O*-GlcNAc sites identified following labeling with **5-4**, Neutravidin affinity purification, and hydrazine-mediated elution. Sites and regions of modification are denoted in red or maroon italics, respectively. Novel site identifications are marked by an asterisk.

## 5.4 Extension to Native *O*-GlcNAcylation Sites

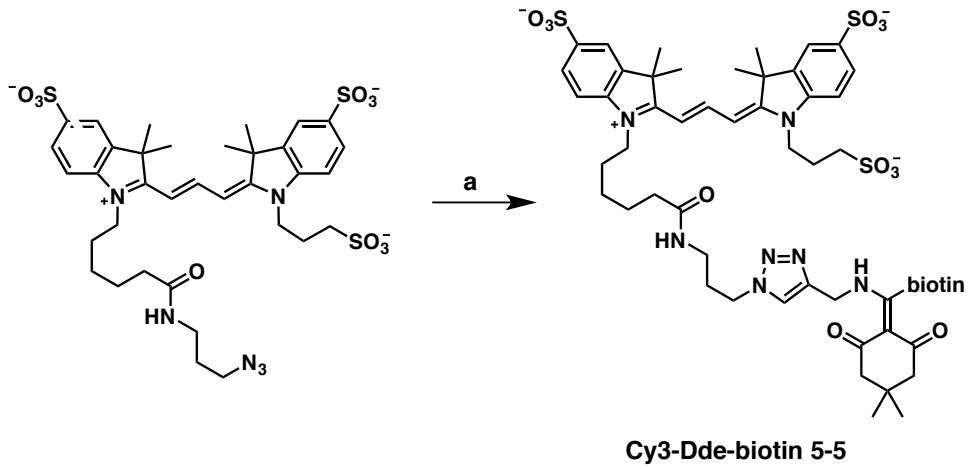
### 5.4.1 Choosing a Workflow

We next aimed to expand this enrichment and identification method to detect *O*-GlcNAcylation events on native proteins. Our original method called for the enrichment of intact proteins and tryptic digestion after elution.[28] Although this method was successful for exogenously added proteins and would be amenable to proteins purified by immunoprecipitation, the preponderance of unmodified peptides that would still be produced by the digestion of enriched

*O*-GlcNAcylated proteins may impede the detection of modified peptides by ion suppression. Therefore, two separate approaches were considered. First, we considered an approach for on-bead digestion. Here, labeled proteins would be immobilized onto Neutravidin resin, and the mixture would be subjected to tryptic digestion. Unmodified peptides would no longer be attached to the resin support and could be removed prior to chemical elution of the remaining, labeled peptides. This approach has been successfully used in a number of cases including the identification of *O*-GlcNAcylated sites in conjunction with isotopic labeling using the dibrominated or deuterated linker described earlier.[15, 29] Alternatively, we envisioned a digestion-first protocol, which would digest complex lysate directly and then proceed with labeling and enrichment steps using the tryptic peptides. This tactic had also been used successfully to enrich *O*-GlcNAcylated peptides using the photocleavable linker.[7, 8]

Selection of the proper workflow depended heavily on the extended stability of the linker to the conditions of on-bead tryptic digestion. Although the linker had shown stability to stringent wash conditions for 1 h, it was unclear if the linker could withstand elevated temperatures for 16 h. To test this, we took a practical approach by conjugating Cy3-azide dye to alkyne-Dde-biotin **5-4** using CuAAC conditions. The resulting compound, Cy3-Dde-biotin **5-5** (Figure 5-6), was used crude by incubating with Neutravidin beads, producing pink beads visible to the naked eye. Samples were then incubated for 24 h in a variety of buffer conditions (1x PBS, pH 7.4; 2 M urea, PBS; 1% SDS, PBS; 20 mM

HEPES pH 7.9, 1% SDS) at room temperature or at 37 °C with gentle rotation. The beads were then pelleted and visually inspected to see if the dye color was lost from the beads. Interestingly, all conditions incubated at 37 °C showed a




**Cy3-Dde-biotin 5-5**

**Figure 5-6. Synthesis of Cy3-Dde-biotin 5-5**. (a) **5-4** (1.2 eq), NaAsc (24 eq), THPTA (1.2 eq), $CuSO_4$ (12 eq), PBS, RT, 2 h.

decrease in dye color localized on the beads with a pink hue in the buffer (Figure 5-7), suggesting that incubation at 37 °C caused hydrolysis of the linker. To confirm that the hydrolysis was not catalyzed by some unknown additive found in the beads, solutions of **5-5** were incubated in 1x PBS, pH 7.4 overnight at room temperature or 37 °C with or without the addition of the Neutravidin beads. Again, both samples incubated at 37 °C showed less dye color on the beads regardless of whether the beads were present (Figure 5-8), suggesting that the elevated temperature alone is enough to facilitate hydrolysis.

Previous reports using Dde as a protecting group during peptide synthesis have described some instances of migration between Lys residues under

extended incubation times.[18, 30] These events only occurred when excess base was present. A previous report indicated that the Dde group would also degrade under similar b s pH 8.5.[19] Interestingly, the previously

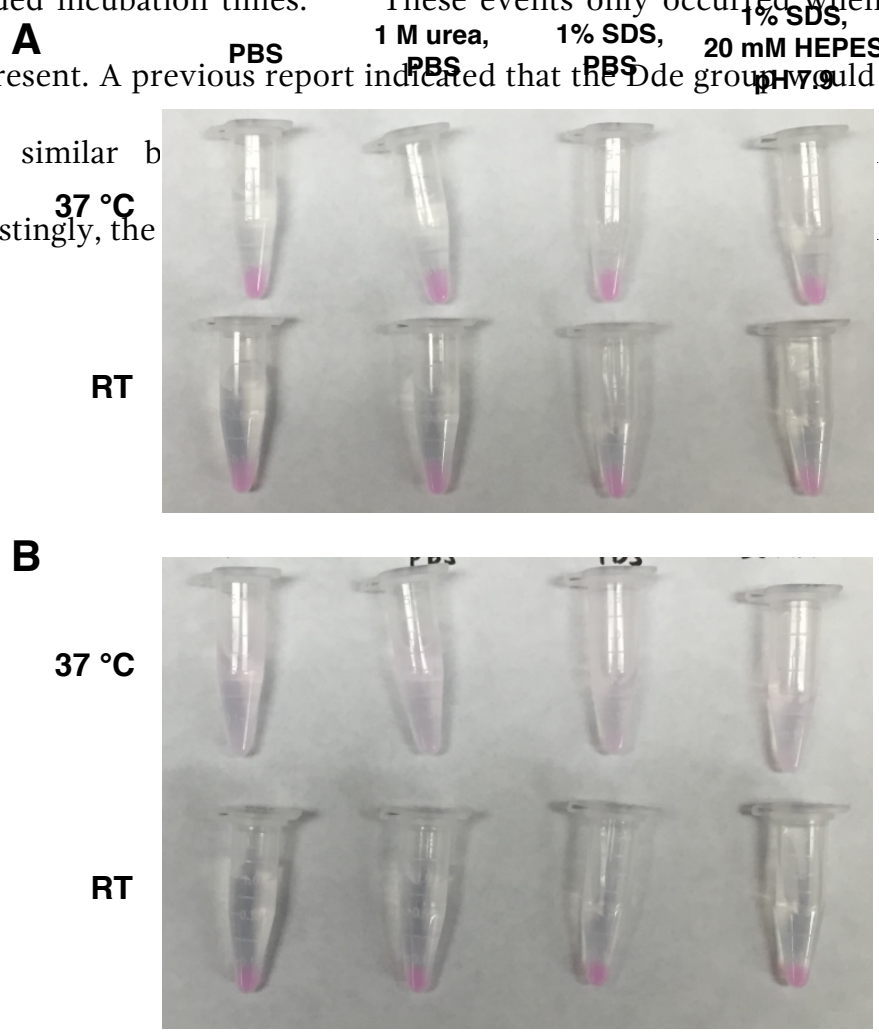



**Figure 5-7. Long-term stability of Dde linker**. Cy3-Dde-biotin **5-5** was (A) immobilized on Neutravidin beads and (B) incubated overnight in different buffers at RT and 37 °C. Beads incubated at 37 °C showed loss of dye to the solution, suggesting that the linker is sensitive to higher temperatures.

reported. To rectify the lability of the Dde group in peptide synthesis, a

sterically hindered alternative known as Ddiv or ivDde was developed by replacing the ethyl group of D sovaleryl group.[18] The biotin group provides a steric e e ivDde group; yet, the linker may be made more resista asing the steric bulk

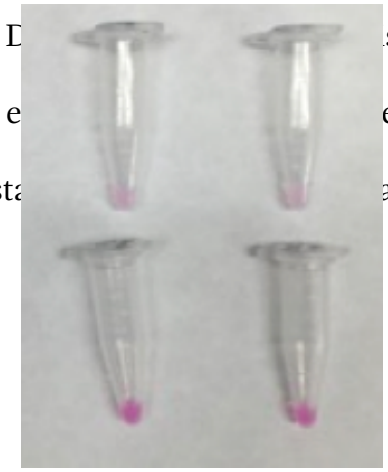



**Figure 5-7. Long-term stability of Dde linker in solution**. Cy3-Dde-biotin **5-5** was diluted in PBS and incubated overnight with or without beads at RT or 37 °C. Regardless of whether beads were present, less dye was immobilized onto beads when incubated at 37 °C.

adjacent to the cyclohexylidene group. Nevertheless, we decided to move forward with the alternative digestion-first approach.

### 5.4.2 Digestion-First Method

In this method, complex lysate would first be subjected to trypsin digestion. After numerous attempts using in solution digestion failed, we decided to turn our attention to filter-assisted sample preparation (FASP).[31] Here, filters are used to buffer exchange solubilized proteins after reduction and alkylation into a solution amenable to trypsin activity. There have also been recent reports of

improved FASP methods,[32, 33] and these approaches were attempted as well. Procedure details are given in the experimental methods section. Briefly, cell lysates from two 15-cm plates of HEK-293T cells at 75% confluency were lysed using SDS lysis buffer, quantified by the BCA assay, and adjusted to 2 mg/mL protein with lysis buffer. Samples (20 mg) were reduced by incubation for 5 min at 95 °C with 20 mM dithiothreitol (DTT) and then concentrated on Amicon centrifugal filter units (10 mg per filter unit). Samples were diluted with 8 M urea, 50 mM Tris pH 7.9 containing 20 mM IAA and allowed to react in the dark for 30 min at RT. The reduced and alkylated lysate was then concentrated and washed twice with urea buffer and twice with Tris buffer. Sequencing-grade trypsin was added to each sample at 1:50 w/w and incubated overnight at 37 °C. Peptides were separated from undigested material by centrifugation, and the filter was washed twice with Tris buffer. The filtrate was combined, acidified to pH 3 using TFA, and desalted with HLB columns. For the enhanced methods, the buffer in the final wash step was replaced with 0.2% w/v sodium deoxycholate, 50 mM Tris pH 7.9. Alternatively, the sample was first washed with 4% w/v sodium deoxycholate, 8 M urea, 50 mM Tris pH 7.9 before continuing with the urea and Tris washes. Acidification after filtration precipitated all remaining sodium deoxycholate, allowing for it to be easily separated prior to desalting. Trypsin is active in the presence of sodium deoxycholate, and the detergent is thought to remove excess, protein-bound SDS, which could inhibit trypsin activity. However, in all cases, we observed

yields of roughly 5-7 mg peptides from a starting amount of 20 mg protein. Therefore, we proceeded with the unmodified FASP method for all future sample preparation.

The remaining steps of the peptide labeling protocol were left largely unchanged from the original protein labeling protocol with a few exceptions. Generally, HLB desalting columns were used in place of protein precipitation as described in the original protocol. The clean-up step for the CuAAC step in other protocols used strong cation exchange (SCX) to separate the peptides from excess alkyne reagent.[7, 8, 34] However, repeated experiments using SCX columns failed to retain the peptides. Instead, we opted to remove excess Dde-alkyne-biotin by adding in azide agarose resin after the peptide reaction was complete. Remaining, unreacted reagent conjugated to the agarose, which could easily be filtered off. Finally, we decided to switch from 2% w/v hydrazine to a buffered solution containing 2% w/v hydroxylamine. The β-*O*-GlcNAc linkage is cleaved under basic conditions,[1] so we aimed to produce an elution buffer with a neutral pH. The $pK_a$ of the hydrazinium ion is about 8, so a buffered solution of pH 7 will only contain about 10% of the reactive, deprotonated nucleophile. Conversely, hydroxylammonium has a $pK_a$ of about 6, which would yield a solution containing roughly 90% of the deprotonated compound at pH 7.

### 5.4.3 Comparing Dde and PC Linkers with Endogenous *O*-GlcNAc Sites

With these adjustments, we conducted the protocol using both **5-1** and **5-4** as a direct comparison. Samples were eluted twice and analyzed separately to

ensure that complete elution was achieved and avoid incomplete cleavage that was observed previously with the **5-1**.[8] In total, 445 unique *O*-GlcNAcylated peptides on 256 different proteins were identified across all four samples using higher-energy collisional dissociation (HCD) MS/MS (Table 5-2). Here, peptides are counted as the number of unique peptide sequences; however, possible multiple modifications or different modification sites on the same peptide were not counted as additional unique peptides. Therefore, this count is likely an underestimation of the number of possible sites. Importantly, peptides containing either tag could easily be identified with their MS/MS spectra by the presence of up to three signature ions caused by the fragmentation of glycosidic bonds within the linker: 300.1 (GalNAz-GlcNAc), 318.1 ($H_2O$ adduct), and 503.2 (GlcNAc-Ser/Thr) *m/z*, which are observed from both linkers due to the identical aminomethyltetrazolyl group left after cleavage.[8] **5-4** strongly outperformed **5-1** in peptide coverage. Of the 442 unique peptides identified in all samples, 414 were observed in the Dde samples, whereas only 257 were observed in the PC sample. In both cases, a single elution step was sufficient to release nearly all of the identified peptides, with only 7 unique peptides identified solely in the second Dde elution step and 16 in the second PC elution.

The modified peptides revealed by this analysis covered a variety of known *O*-GlcNAcylated proteins.[5, 7, 8, 13] As expected, sites on many multiply *O*-GlcNAcylated proteins such as host cell factor 1 (Hcfc1), Msx2-interacting protein (Mint), myosin phosphatase target subunit 1 (Mypt1), nucleoporins

(Nup98, Nup153, Nup214), glutamine and serine rich protein 1 (Qser1), and lysine deficient protein kinase 1 (Wnk1) were observed in the dataset. Interestingly, Qser1 and the related protein proline and serine rich protein 1 (Prsr1), which was also observed to be *O*-GlcNAcylated in this dataset, are ubiquitously expressed but have no known function. Thus, these results may help provide evidence for the role of these proteins in the cell and its regulation by *O*-GlcNAcylation.

**5.5 Conclusions and Outlook**

The results presented herein provide the framework for a new approach to identify sites of *O*-GlcNAcylation on proteins using a novel cleavable tag **5-4** based on the Dde functional group. This linker provides numerous benefits over existing linkers due to its complete cleavage by hydrazine and hydroxylamine as well as its ability to leave only the small aminomethyltetrazolyl group after cleavage. Moreover, the remaining linker provides an additional positive charge to the labeled peptide, aiding in ETD fragmentation. There still remain areas of improvement for the linker, including synthetic variants closer to the ivDde structure to avoid cleavage at elevated temperatures. Nevertheless, we envision that this linker can be combined with isotopic labeling methods such as SILAC or dimethyl labeling to allow for the quantitative measure of *O*-GlcNAcylation dynamics based on individual sites. These experiments will provide critical information regarding the regulation of protein *O*-GlcNAcylation to help delineate its physiological functions in diverse biological systems.

**Table 5-2. *O*-GlcNAc sites identified from HEK-293T lysate**. Samples were processed using a digestion-first protocol and labeled using the two-step chemoenzymatic approach with **5-4** or **5-1**. Peptides were identified by HCD-MS/MS and are presented as unique sequences.

| Uniprot | Peptide sequence | Peptide counts | | | |
|---|---|---|---|---|---|
| | | *Dde-1* | *Dde-2* | *PC-1* | *PC-2* |
| 4ET_HUMAN | SVLHPPGSGSHAAAVSVQTTPQNVPSR | 1 | | | |
| AAK1_HUMAN | LTDPIPTTETSIAPR | | | 1 | |
| ABLM1_HUMAN | STSQGSINSPVYSR | 1 | | | |
| ADRM1_HUMAN | SQSAAVTPSSTTSSTR | 1 | 1 | 1 | 2 |
| AF10_HUMAN | NPGTTVSAASPFPQGSFSGTPGSVK | 2 | | | |
| AGFG1_HUMAN | APVGSVVSVPSQSSASSDK | 6 | 3 | 5 | 3 |
| | SSSADFGTFNTSQSHQTASAVSK | 6 | 3 | 7 | 4 |
| | VVASVHASISGSSASSTSSTPEVKPLK | 1 | | | |
| AGFG2_HUMAN | TLLGDPAPSLSVAASTSSQPVSQSHAR | 4 | 1 | 2 | |
| AHNK_HUMAN | GPQVSSALNLDTSK | 4 | 2 | 3 | 2 |
| AHNK2_HUMAN | ESEIPTSEIQTPSYGFSLLK | 1 | | | |
| AINX_HUMAN | SNVASSAACSSASSLGLGLAYR | 1 | | | |
| ALMS1_HUMAN | ISVASEPVDQTTGTPAVTSTSYSQYR | 1 | | 1 | |
| | TETPSVSSSLYSYR | 1 | | | |
| | VSVAPGPVGQTTGAPTITSPSYSQHR | 1 | | 2 | |
| AMRA1_HUMAN | TSASSVSLLSVLR | 1 | | | |
| ANKH1_HUMAN | NAFPLGAPTLVTSQATTLSTFQPANK | 4 | 3 | 2 | 3 |
| | SIHANFSSGVGTTAASSK | 2 | | | |
| | VSTSPVGLPSIDPSGSSPSSSSAPLASFSGIPGTR | 2 | | | |
| ANR17_HUMAN | IGSSAPTTTAANTSLMGIK | 4 | | 2 | 1 |
| | LKVEDEPEVLTEPPSATTTTTIGISATWTTLAGSHGK | 2 | | | |
| | MTTVALSSTSQTATALTVPAISSASTHK | 3 | | 1 | |
| | MTVPPLATSSAPVAVPSTAPVTYPMPQTPMGCPQPTPK | 1 | | | |
| | QHFSPLSLLTPCSSASNDSSAQSVSSGVR | 2 | | 1 | |
| | TSNATTTTVTTTASNNNTAPTNATYPMPTAK | 1 | | | |
| APMAP_HUMAN | AGPNGTLFVADAYK | | | 1 | |
| ARI1A_HUMAN | GGTPGSGAAAAAGSKPPPSSSASASSSSSSFAQQR | 2 | | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| | NPQMPQYSSPQPGSALSPR | 1 | | | |
| ARI1B_HUMAN | VMPTVPTSQVTGPPPQPPPIR | 9 | 2 | 6 | 2 |
| ARI3A_HUMAN | LPVSLAGHPVVAAQAAAVQAAAAQAAVAAQAAALEQLR | 1 | | | |
| ARIP4_HUMAN | VVTTTDIVIPGLNSSTDVQAR | 5 | | 4 | 1 |
| ARNT_HUMAN | HSNPTQGATPTWTPTTR | 2 | | | 3 |
| ASPP2_HUMAN | ENLPVSSDGNLPQQAASAPSR | | | 1 | |
| ATF1_HUMAN | TTPSATSLPQTVVMTSPVTLTSQTTK | 3 | 2 | 5 | |
| ATF7_HUMAN | SAAEAVATSVLTQMASQR | 1 | | | |
| ATX1L_HUMAN | APSATSPSGQLPHHSSTQPLDLAPGR | 2 | | 2 | |
| ATX2L_HUMAN | SAAPAPISASCPEPPIGSAVPTSSASIPVTSSVSDPGVGSISPASPK | 3 | 1 | 2 | 1 |
| BAG3_HUMAN | SQSPAASDCSSSSSSASLPSSGR | | | 2 | |
| BAZ2B_HUMAN | GGLSTGVASLSSTINPCGHLFR | 2 | | | |
| | LPSSAASSTTPTSSSTPSVASVVSK | 2 | | | |
| BCORL_HUMAN | TPPMPVLTPVHTSSK | 1 | | | |
| BPTF_HUMAN | FLFTPLATTATTASTTTTTVSTTAAGTGEQR | 5 | | 2 | 1 |
| | GQPVSTAVSAPNTVSSTPGQK | 4 | 2 | 2 | 2 |
| | LEQQKPTVIATSTTSPTSSTTSTISPAQK | | | 1 | |
| | STVTTTTTTVTK | 1 | | | |
| | TVITEVTTMTSTVATESK | 4 | 2 | 6 | 3 |
| | VMVAPISGSVTTGTK | 3 | 1 | 2 | 1 |
| BRD8_HUMAN | LLEAGPTQFTTPLASFTTVASEPPVK | | | 1 | |
| CACL1_HUMAN | AAPAPTASSTININTSTSK | | | 1 | |
| CARF_HUMAN | SSSSTNTSLLTSK | 1 | | | 1 |
| CATL1_HUMAN | YSVANDTGFVDIPK | 2 | | 1 | |
| CBL_HUMAN | VPVSAPSSSDPWTGR | 3 | 1 | 2 | 2 |
| CDK12_HUMAN | TSAVSSQANSQPPVQVSVK | 5 | 2 | 4 | 2 |
| CDK13_HUMAN | TENQHVPTTSSSLTDPHAGVK | 2 | | | |
| CDK8_HUMAN | VVPPTTTSGGLIMTSDYQR | 3 | | | |
| CE170_HUMAN | EINDVAGEIDSVTSSGTAPSTTVSTAATTPGSAIDTR | 1 | | | |
| CIC_HUMAN | GYGSAPSSSASSPASSSASAATSFSLGSGTFK | 1 | | 1 | |
| CKAP5_HUMAN | ISTSTGISPQMEVTCVPTPTSTVSSIGNTNGEEVGPSVYLER | 2 | | 2 | 1 |
| CLIP1_HUMAN | VQAEDEANGLQTTPASR | 1 | | | |
| CNOT1_HUMAN | APLAGQVSTMVTTSTTTTVAK | 16 | 3 | 18 | 5 |
| CNOT2_HUMAN | SLSQGTQLPSHVTPTTGVPTMSLHTPPSPSR | 1 | | | |
| CNOT4_HUMAN | SNPVIPISSSNHSAR | 1 | | | |
| CRTC2_HUMAN | SLQQPGLPSQSCSVQSSGGQPPGR | 2 | 1 | 1 | |
| CUL4B_HUMAN | MAEESSSSSSSSSPTAATSQQQQLK | 1 | | 2 | |
| CUX1_HUMAN | QAPLSQSDITILTPK | 1 | | | |
| DAPLE_HUMAN | TCSTSATTTAPSNSTPIAR | 1 | | 1 | |
| DIDO1_HUMAN | TYFPGPPGDGHPEPSPLEDLSPCPASCGSGVVTTVTVSGR | | 1 | 1 | |
| | VLSSLKPAAPSPATAATTAAAASTAASSTASSASK | 4 | | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| DLG5_HUMAN | SLTPSTTVSSILR | 1 | | | |
| DSRAD_HUMAN | NAEFLTCNIPTSNASNNMVTTEK | 1 | | 4 | 1 |
| E41L1_HUMAN | DVLTSTYGATAETLSTSTTTHVTK | 2 | | | |
| E41L2_HUMAN | TITYESPQIDGGAGGDSGTLLTAQTITSESVSTTTTTHITK | 2 | | 1 | |
| ELF1_HUMAN | FILQAIPSSQPMTVLK | 2 | | | |
| ELF2_HUMAN | ALTPVSIAHGTPVMR | 1 | | | |
| | VAMQVPVVMTSLGQK | 1 | | | |
| EMD_HUMAN | LSPPSSSAASSYSFSDLNSTR | 3 | | | |
| EMSA1_HUMAN | VKEEQYLGHEGPGGAVSTSQPVELPPPSSLALLNSVVYGPER | | | 1 | |
| EMSY_HUMAN | IISSNIVSGTTTK | 2 | | | |
| | ITFTKPSTQTTNTTTQK | 1 | | | |
| | MSNIMQSIANSLPPHMSPVK | 2 | | | |
| | QTASQVEQPIITQGSSVTK | 1 | | | |
| | TTSGSIITVVPK | 1 | | 2 | |
| | VIIVTTSPSSTFVPNILSK | 3 | 1 | 3 | 2 |
| EP300_HUMAN | SGSSPNLNMGVGGPGQVMASQAQQSSPGLGLINSMVK | 1 | | | |
| EP400_HUMAN | AAAAPFQTSQASASAPR | 1 | | | 1 |
| | AQPAITTGGSAAVLAGTIK | 1 | | | |
| | AVGSPATATPDLVSMATTQGVR | 3 | 2 | 4 | 2 |
| | AVTSVTASAVVTTNLTPVQTPAR | 5 | 1 | 1 | 1 |
| | SLVPQVSQATGVQLPGK | 1 | 1 | 1 | |
| | TAAPTTASAAPQGPLR | 1 | | | |
| | TQFLTTPISQAQK | 3 | 1 | | 1 |
| | TSVTGTSMPTGAVSGNVIVNTIAGVPAATFQSINK | 3 | | | |
| EPC1_HUMAN | LTVPSSVATVNSIAPINAR | 1 | | | |
| EYA4_HUMAN | TEPLNSSETTATTGDGALDTFTGSVITSSGYSPR | 1 | | | |
| F193A_HUMAN | SPPSVSSASSGSGSSSPITIQQHPR | 2 | | | |
| | TATTTPGFVDTR | | | 1 | |
| | VVMATSSATSSVSCTATTVQSSNSQFR | 2 | 1 | 1 | |
| F208B_HUMAN | ETPLPVSLPSDK | 2 | | | |
| | SLSDTLVSTTAPSGIVNVSVK | 1 | | | |
| | VASYSGTVTQATFTR | 1 | | | |
| FLIP1_HUMAN | VTSTITITPVTTSSAR | 2 | 1 | 1 | 2 |
| FNBP4_HUMAN | ATEISTAVVQR | | 1 | 1 | |
| FOXC1_HUMAN | GSPQSAAAELSSGLLASAAASSR | 2 | | | |
| FOXK1_HUMAN | EPAAAVAATATTTPATATTASASASSTGEPEVK | 13 | 4 | 11 | 3 |
| | HAVPTNSLAGNAYALTSPLQLLATQASSSAPVVVTR | 1 | | 2 | |
| | VVTTSANSANGYILTSQGAAGGSHDAAGAAVLDLGSEAR | 5 | 1 | 4 | |
| FOXK2_HUMAN | FAQSAPGSPLSSQPVLITVQR | 1 | | | |
| GABPA_HUMAN | YVLASQEQQMNEIVTIDQPVQIIPASVQSATPTTIK | 2 | | | |
| GANP_HUMAN | SPTSVGAFPSTSAFGQEAGEIVNSGFGK | 3 | 2 | 1 | |

| | | | | | |
|---|---|---|---|---|---|
| GCR_HUMAN | VSASSPSLAVASQSDSK | | 1 | | |
| GEMI5_HUMAN | TVIESSPESPVTITEPYR | 1 | | | |
| GMEB2_HUMAN | VVSTLPSTVLGK | 2 | | | |
| GPKOW_HUMAN | EGVLPLTAASTAPISFGFTR | 1 | | | |
| GSCR1_HUMAN | QVPVSGYLASAAGPSEPVTLASAGVSPQGAGLVIQK | 1 | | | |
| GSE1_HUMAN | RVPMGPIIVPPGGHSVPSTPPVVTIAPTK | 1 | | | |
| HCFC1_HUMAN | APVTVTSLPAGVR | 3 | 2 | 3 | 1 |
| | AVTTVTQSTPVPGPSVPPPEELQVSPGPR | 2 | | | |
| | HSHAVSTAAMTR | 1 | | | |
| | IPPSSAPTVLSVPAGTTIVK | 12 | 5 | 7 | 5 |
| | ISVATGALEAAQGSK | 7 | 3 | 5 | 2 |
| | LVTPVTVSAVKPAVTTLVVK | 6 | 3 | 3 | 1 |
| | QEAAASLVTSTVGQQNGSVVR | 7 | 2 | 5 | 1 |
| | QTSATSTTMTVMATGAPCSAGPLLGPSMAR | 7 | 3 | 8 | 1 |
| | SGTVTVAQQAQVVTTVVGGVTK | 16 | 4 | 13 | 3 |
| | SPISVPGGSALISNLGK | 2 | 1 | 1 | 1 |
| | SPITIITTK | 6 | 3 | 5 | 2 |
| | SSVGAGEPR | 1 | | | |
| | TAAAQVGTSVSSATNTSTRPIITVHK | 20 | | 7 | 2 |
| | TIPMSAIITQAGATGVTSSPGIK | 25 | 13 | 18 | 9 |
| | TMAVTPGTTTLPATVK | 11 | 4 | 12 | 4 |
| | VASSPVMVSNPATR | 3 | | 1 | 1 |
| | VMSVVQTKPVQTSAVTGQASTGPVTQIIQTK | 7 | 2 | 6 | 4 |
| | VMTSGTGAPAK | 1 | | 1 | 2 |
| | VTGPQATTGTPLVTMRPASQAGK | 2 | | 2 | |
| | YDIPATAATATSPTPNPVPSVPANPPK | 2 | | 1 | |
| HGS_HUMAN | AEPMPSASSAPPASSLYSSPVNSSAPLAEDIDPELAR | 11 | 3 | 10 | 3 |
| HSPB1_HUMAN | LATQSNEITIPVTFESR | 3 | 1 | 1 | 1 |
| HTF4_HUMAN | GSTSSSPYVAASHTPPINGSDSILGTR | 1 | | | |
| | LSYPPHSVSPTDINTSLPPMSSFHR | 1 | | | |
| I2BP2_HUMAN | AAASLAAVSGTAAASLGSAQPTDLGAHK | 3 | | 4 | |
| | INGEAQPWLSTSTEGLK | 2 | | 2 | |
| I2BPL_HUMAN | FEYPPPPVSLGSSSHTAR | 8 | 3 | 5 | 2 |
| | GPPTPAPPGAPGGPACLGGTPGVSATSSSASSSTSSSVAEVGVGAGGK | 1 | | 1 | 1 |
| | SRFEYPPPPVSLGSSSHTAR | 5 | | 4 | |
| IF4G3_HUMAN | EQEGQTSETTAIVSIAELPLPPSPTTVSSVAR | 11 | | 3 | 1 |
| | KEQEGQTSETTAIVSIAELPLPPSPTTVSSVAR | 2 | | 1 | |
| | STIAAPTSSALSSQPIFTTAIDDR | 8 | 2 | 2 | 3 |
| ITB1_HUMAN | VCECNPNYTGSACDCSLDTSTCEASNGQICNGR | 2 | | 5 | 2 |
| ITSN2_HUMAN | AQSLIDLGSSSSTSSTASLSGNSPK | 1 | | 1 | |
| JHD2C_HUMAN | HSVPQSLPQSNYFTTLSNSVVNEPPR | 4 | | 2 | |

| | | | | | |
|---|---|---|---|---|---|
| | SPTHLTVSSTNTLR | 1 | | | |
| JUNB_HUMAN | LIVPNSNGVITTTPTPPGQYFYPR | 1 | | | |
| K2026_HUMAN | CLTSALQIPVTVALPTPATTSPK | 1 | | | |
| K2C8_HUMAN | TTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFSR | 3 | | 1 | |
| KANL3_HUMAN | LPTPMQSLGAITTGTSTIVR | 3 | | | |
| | SSSSEGGVSASPVPSVVSSSTAPSALHTLQSR | 2 | | 1 | |
| | VPTTITLTLR | 1 | | | |
| KCMF1_HUMAN | SNMHFTSSSTGGLSSSQSSYSPSNR | | | 1 | |
| KDM3B_HUMAN | NSILASSGFGAPLPSSSQPLTFGSGR | 1 | 1 | 2 | |
| | TLEQVGQGIVASAAVVTTASSTPNTVR | 10 | 1 | 8 | 5 |
| | TLVVQDEPVGGDTPASFTPYSTATGQTPLAPEVGGAENK | 1 | | | |
| | VEHSPFSSFASQASGSSSSATTVTSK | 7 | 1 | 5 | 2 |
| KMT2D_HUMAN | SLPSDPFSR | 1 | | | |
| KRT81_HUMAN | LCEGIGAVNVCVSSSR | | 1 | | |
| LAP2A_HUMAN | SSSSSSQPEHSAMLVSTAASPSLIK | 1 | | | |
| LAR4B_HUMAN | EPSVPASCAVSATYER | 1 | | | |
| | TLSADASVNTLPVVVSR | 1 | | | |
| LIMD1_HUMAN | TPSVSAPLALSCPR | 1 | | | |
| LIN54_HUMAN | LGAQTPVTISANQIILNK | 1 | | | |
| | TITISESGVIGSTLNSTTQTPNK | 1 | | 1 | |
| LMNA_HUMAN | ASASGSGAQVGGPISSGSSASSVTVTR | 9 | 5 | 8 | 3 |
| | SVGGSGGGSFGDNLVTR | 1 | 1 | | 1 |
| LMO7_HUMAN | TSTTGVATTQSPTPR | | | 1 | |
| LPP_HUMAN | STGEPLGHVPAR | 2 | 1 | 1 | 2 |
| LRIF1_HUMAN | ILATATTSTSGMVEASQMPTVIYVSPVNTVK | 1 | | | |
| MAFK_HUMAN | STELSSTSVPFSAAS | 2 | 2 | 2 | 1 |
| | VATTSVITIVK | 2 | 1 | 2 | 1 |
| MAP4_HUMAN | ASPSKPASAPASR | 1 | | | 1 |
| MATR3_HUMAN | DLSAAGIGLLAAATQSLSMPASLGR | 3 | 1 | 2 | 1 |
| MAVS_HUMAN | VPTTLMPVNTVALK | 1 | | | |
| MBD5_HUMAN | DIPNPLIAGISNVLNTPSSAAFPTASAGSSSVK | 1 | | | |
| MCAF1_HUMAN | NPTASAAPLGTTLAVQAVPTAHSIVQATR | 1 | | | |
| | NPVSLPSLPNPTKPNNVPSVPSPSIQR | 1 | | | |
| | TSLPTVGPSGLYSPSTNR | 2 | | | |
| MED15_HUMAN | FPPTTAVSAIPSSSIPLGR | 9 | | 7 | |
| MGAP_HUMAN | IPGVSTPQTLAGTQK | 1 | | | |
| | TTGITTPVASVAFPK | 2 | | | |
| MINT_HUMAN | ADRPSLEKPEPIHLSVSTPVTQGGTVK | 3 | | 2 | |
| | AQSTPSPALPPDTK | 1 | | 1 | |
| | ASDVDTSSSTLR | 1 | | 1 | 1 |
| | QPLFVPTTSGPSTPPGLVLPHTEFQPAPK | 3 | | 2 | |

| | | | | | |
|---|---|---|---|---|---|
| | SLVSTPAGPVNVLK | 1 | | | |
| | VNTSEGVVLLSYSGQK | 2 | 2 | 2 | |
| MKL2_HUMAN | VSESPSPVTTNTPAQFASVSPTVPEFLK | 1 | | | |
| MLXIP_HUMAN | EGMLASTVSQSNVVIAPAAIAR | 2 | | | |
| MUC5A_HUMAN | GCPVTSTPVTAPSTPSGR | | | 1 | |
| MYPT1_HUMAN | DSVPTAVTIPVAPTVVNAAASTTTLTTTTAGTVSSTTEVR | 3 | | 1 | 3 |
| | QDDLISSSVPSTTSTPTVTSAAGLQK | 12 | 2 | 5 | 1 |
| | RQDDLISSSVPSTTSTPTVTSAAGLQK | 6 | | 2 | 3 |
| | TKPLASVTNANTSSTQAAPVAVTTPTVSSGQATPTSPIK | 3 | | 1 | 1 |
| NASP_HUMAN | ATLVESSTSGFTPGGGGSSVSMIASR | 2 | | | |
| NCOA3_HUMAN | AVSLDSPVSVGSSPPVK | 1 | | | |
| NCOA6_HUMAN | SIVTTLVPSELISAVPTTK | 3 | | 1 | 1 |
| NCOR1_HUMAN | HTSVVSSGPSVLR | 1 | | | |
| | IMPLPAGGPSISQGLPASR | 2 | | | |
| | NQVSSQTPQQPPTSTFQNSPSALVSTPVR | 3 | 1 | 5 | 2 |
| | SSHLEVSQASQLLQQQQQQQLR | 1 | | | |
| | YPPHSVQYTFPNTR | 3 | | | |
| NCOR2_HUMAN | AISSASIEGLMGR | | | | 1 |
| | VVTLAQHISEVITQDYTR | 1 | | 1 | |
| | YPPHSLSYPVQIAR | 3 | | | |
| NFIA_HUMAN | ASPHATPSTLHFPTSPIIQQPGPYFSHPAIR | 3 | | | |
| NFIC_HUMAN | LALPPATKPATTSEGGATSPTSPSYSPPDTSPANR | 1 | | | |
| NFRKB_HUMAN | IQTVPASHLQQGTASGSSK | 1 | | 2 | |
| | LMPALGVSVADQK | 1 | | | |
| | QVPVSTTVVSTSQAGK | 2 | | 2 | 1 |
| | TVAVASGAASTPISISTGAPTVR | 6 | 1 | 2 | 2 |
| NOTC1_HUMAN | CNCLLPYTGATCEVVLAPCAPSPCR | | | 1 | |
| NOTC2_HUMAN | YSCVCSPGFTGQR | | 1 | | 1 |
| NOTC3_HUMAN | CQCPAGYTGPLCENPAVPCAPSPCR | 1 | | | |
| NPM_HUMAN | MSVQPTVSLGGFEITPPVVLR | 1 | | | |
| NU153_HUMAN | ALTLTVVSESAETMTASSSSCTVTTGTLGFGDK | 4 | | 2 | |
| | CQPVFSFGNSEQTK | | | 1 | |
| | CVSCMSEKPGSSVPASSSSTVPVSLPSGGSLGLEK | 2 | | 1 | |
| | FGVSSSSSGPSQTLTSTGNFK | 11 | 1 | 5 | 3 |
| | GFDTSSSSSNSAASSSFK | | 2 | 1 | 2 |
| | IGVSSDSGSINPMSEGFK | 2 | 1 | 4 | 2 |
| | ISLPITSSSLPTFNFSSPEITTSSPSPINSSQALTNK | 8 | 4 | 6 | 2 |
| | QQEPVTSTSLVFGK | 6 | 2 | 4 | 2 |
| | SSSAGFSFGTGVINSTPAPANTIVTSENK | 6 | 4 | 3 | 3 |
| | STEANVLPPSSIGFTFSVPVAK | 31 | 6 | 12 | 10 |
| | VQMTSPSSTGSPMFK | 2 | 2 | | 2 |

| | | | | | |
|---|---|---|---|---|---|
| NU214_HUMAN | ASSTSLTSTQPTK | 2 | 2 | | 1 |
| | EPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVPDAR | 6 | 2 | 2 | |
| | GGGFFSGLGGKPSQDAANKNPFSSASGGFGSTATSNTSNLFGNSGAK | 1 | | | |
| | KEPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVPDAR | 8 | 1 | 4 | |
| | LGELLFPSSLAGETLGSFSGLR | 4 | 2 | 2 | 3 |
| | NNPATPSTAMGSSVPYSTAK | 4 | | 4 | 1 |
| | NPFSSASGGFGSTATSNTSNLFGNSGAK | 10 | 4 | 5 | 3 |
| | SSATVTGEPPSYSSGSDSSK | 1 | | | |
| | TFGGFASSSFGEQKPTGTFSSGGGSVASQGFGFSSPNK | 5 | | 2 | |
| | TGGFGAAPVFGSPPTFGGSPGFGGVPAFGSAPAFTSPLGSTGGK | 1 | | 1 | |
| NUFP2_HUMAN | TIQNSSVSPTSSSSSSSSTGETQTQSSSR | 1 | | 1 | |
| NUP53_HUMAN | ASTSDYQVISDR | 1 | | | |
| NUP98_HUMAN | AASLMNIPSTSSWSVPPPLTSVFTMPSPAPEVPLK | 1 | | 3 | |
| | FTSGAFLSPSVSVQECR | 2 | | 1 | 2 |
| | GPQNQVGAGTTTGLFGSSPATSSATGLFSSSTTNSGFAYGQNK | 1 | | 1 | 1 |
| | KGPQNQVGAGTTTGLFGSSPATSSATGLFSSSTTNSGFAYGQNK | 3 | | | 1 |
| P121A_HUMAN | APPTLQAETATKPQATSAPSPAPK | 2 | | | |
| P121B_HUMAN | QSFLFGTQNTSPSSPAAPAASSASPMFKPIFTAPPK | 2 | | | |
| P66A_HUMAN | GTTATSAQANSTPTSVASVVTSAESPASR | 2 | | 4 | 3 |
| | TPLSTGGTLAFVSPSLAVHK | 2 | | | |
| P66B_HUMAN | LQQQAALSPTTAPAVSSVSK | 2 | 1 | | |
| PACS1_HUMAN | LAQATSSSSSTSAAAASSSSSSTSTSTSMAVAVASGSAPPGGPGPGR | 1 | | | |
| PAPOA_HUMAN | EQLDTETSTTQSETIQTAASLLASQK | | | 1 | |
| PCF11_HUMAN | SPEEPSTPGTVVSSPSISTPPIVPDIQK | | | 1 | |
| PDLI5_HUMAN | ANNSQEPSPQLASSVASTR | 2 | | 2 | 1 |
| | EVVKPVPITSPAVSK | 4 | | | |
| PF21A_HUMAN | FTPTTLPTSQNSIHPVR | 3 | 1 | 3 | 1 |
| PHAR4_HUMAN | FIISTSITTAPAATTAATSLAK | 17 | 4 | 7 | 2 |
| PHC1_HUMAN | QPGTAQAQALGLAQLAAAVPTSR | 4 | | 2 | 1 |
| PHC3_HUMAN | STSQTQSLTICHNK | 1 | | 1 | |
| PHF3_HUMAN | GSAVATSHFEVGNTCPSEFPSK | 1 | | | |
| PICAL_HUMAN | KPHTSLTTAASPVSTSAGGIMTAPAIDIFSTPSSSNSTSK | 5 | | 3 | 1 |
| | SSGDVHLSISSDVSTFTTR | | | 1 | |
| PKCB1_HUMAN | SSAQTSAAGATATTSTSSTVTVTAPAPAATGSPVK | 1 | | | |
| | SSAQTSAAGATATTSTSSTVTVTAPAPAATGSPVKK | 1 | | | |
| | TPPSTTVGSHSPPETPVLTR | 1 | | 1 | |
| PKHA5_HUMAN | IVNVSLADLR | 4 | 1 | 2 | 1 |
| PLEC_HUMAN | AGVAAPATQVAQVTLQSVQR | 1 | | | |
| PLIN3_HUMAN | TLTAAAVSGAQPILSK | 3 | | 2 | 1 |
| | VASMPLISSTCDMVSAAYASTK | 4 | 1 | 2 | |
| PLRG1_HUMAN | MPSESAAQSLAVALPLQTK | 1 | | | |

| | | | | | |
|---|---|---|---|---|---|
| POGZ_HUMAN | LAPSFPSPPAVSIASFVTVK | 6 | 3 | 2 | 1 |
| | STPSTSTTPTATQPTSLGQLAVQSPGQSNQTTNPK | 4 | | 1 | |
| PRC2B_HUMAN | DSDFSLPPGSASGPTGSPVVK | 6 | | | |
| | GGLPVSQSQEIFSSLQPFR | 4 | 3 | 2 | 2 |
| PRC2C_HUMAN | AVSEMSTEIGTMISVSSAEYGTNAK | 3 | | 3 | |
| | ESVTDYTTPSSSLPNTVATNNTK | 2 | 2 | 2 | 1 |
| | ETIQQSSSLTSVPPTTFSLTFK | 6 | 1 | 7 | 2 |
| | RETIQQSSSLTSVPPTTFSLTFK | 5 | | 3 | |
| PRSR1_HUMAN | GFLTSNDTNLINSSALSSAVTSGLASLSSLTLQNSDSSASAPNK | 1 | | | |
| | GPHPGTSDLHISSTPAATTLPVMIK | 2 | | | |
| PTN12_HUMAN | TNISTASATVSAATSTESISTR | 2 | 1 | 2 | |
| PUM1_HUMAN | SASSASSLFSPSSTLFSSSR | 2 | 1 | 2 | 2 |
| QSER1_HUMAN | KTEALQVATTSPTANTTGTATTSSTTVGAVK | 2 | | | |
| | QSSLSCSPIGDSTQVSNGGLQQK | 1 | | | |
| | SCSTEQPLTSTK | | | 1 | 1 |
| | TAQAAASGTTLLPQFR | 2 | | 1 | |
| | TEALQVATTSPTANTTGTATTSSTTVGAVK | 1 | 1 | 1 | |
| | TSQGTVPTALAFER | 3 | | 1 | 2 |
| RBM12_HUMAN | VNLPTTVSNFNNPSPSVVTATTSVHESNK | 16 | | 8 | 2 |
| RBM14_HUMAN | AQPSVSLGAAYR | 1 | | | 2 |
| RBM27_HUMAN | AANIVIQTEPPVPVSINSNITR | 5 | 4 | 7 | 1 |
| | MMSKPQTSGAYVLNK | 1 | | | |
| | TQTQRPNLIGLTSGDMDVNPR | 1 | | | |
| RBM4_HUMAN | HLLPTSGAAATAAAAAAAAAAVTAASTSYYGR | 5 | | 1 | |
| RBM7_HUMAN | TMDNMTSSAQIIQR | 1 | | 1 | 1 |
| RBMS2_HUMAN | MLAQSALSPYLSSPVSSYQR | 2 | 1 | | |
| RC3H2_HUMAN | VGVNNTVTTTAGNVISVIGSTETTGK | 1 | | | |
| RFX5_HUMAN | TAEVPVSEASGQAPPAK | 1 | | | 1 |
| RGPD3_HUMAN,RGPD4_HUMAN | EGFSIPVSADGFK | 2 | | | |
| | TDVIQGDDVADAASEVEVSSTSETTTK | 1 | | | |
| RIPK1_HUMAN | MQSLQLDCVAVPSSR | | 1 | | |
| RLA2_HUMAN | LASVPAGGAVAVSAAPGSAAPAAGSAPAAAEEK | | | 1 | |
| RPRD2_HUMAN | SAVSTSVPTKPTENISK | 3 | | 1 | |
| | SFNYSPNSSTSEVSSTSASK | 1 | | 1 | |
| | SLFSPQNTLAAPTGHPPTSGVEK | 1 | 1 | 1 | |
| | TPAPATTTSHNPLANILSK | 2 | | 1 | |
| RPTOR_HUMAN | NYALPSPATTEGGSLTPVR | 1 | | 1 | |
| RRP1B_HUMAN | TPTSSPASSPLVAK | 1 | | | |
| RSBNL_HUMAN | QLQPPAAPSPQSYGSPASWSFAPLSAAPSPSSSR | 1 | | | |
| S30BP_HUMAN | GTTTNATSTTTTTASTAVADAQK | 7 | 2 | 8 | 2 |
| | KGTTTNATSTTTTTASTAVADAQK | 3 | | 3 | |

| | | | | | |
|---|---|---|---|---|---|
| | SKWDSAIPVTTIAQPTILTTTATLPAVVTVTTSASGSK | 3 | | | |
| | WDSAIPVTTIAQPTILTTTATLPAVVTVTTSASGSK | 5 | 1 | 3 | |
| SAP_HUMAN | DNATEEEILVYLEK | 1 | | | |
| | TNSTFVQALVEHVK | 5 | | 2 | |
| SARNP_HUMAN | FGIVTSSAGTGTTEDTEAK | | | 1 | |
| SBNO1_HUMAN | FIQTTASTRPSVSAPTVR | 1 | | | |
| | TPPVTTNR | | | 1 | |
| SC16A_HUMAN | TLENPVNVYNPSHSDSLASQQSVASHPR | 2 | | | |
| SC24A_HUMAN | ASSQPTVSGNTSLTTNHQYVSSGYPSLQNSFIK | 1 | 1 | | |
| SC24B_HUMAN | SSPVVSTVLSGSSGSSSTR | 7 | 2 | 5 | 2 |
| | TPPTANHPVEPVTSVTQPSELLQQK | 9 | 2 | 4 | 2 |
| SC24C_HUMAN | APPSSGAPPASTAQAPCGQAAYGQFGQGDVQNGPSSTVQMQR | 5 | 1 | 2 | |
| SCAF8_HUMAN | ETVQTTQSPTPVEK | 1 | | | 1 |
| | SSEPVKETVQTTQSPTPVEK | 1 | | | |
| SCML2_HUMAN | NPMYIHTSVSQDFSR | 1 | 1 | | |
| SCYL2_HUMAN | SSASSTFTSVPSMGIGMMFSTPTDNTK | 1 | | | |
| SET1A_HUMAN | GSTPYSQDSAYSSSTTSTSFKPR | 1 | | | |
| | QDTPSSFGQFTPQSSQGTPYTSR | 2 | | | |
| SET1B_HUMAN | DFSFTPTFSEPSGPLLLPVCPLPTGR | | | 1 | |
| SFPQ_HUMAN | FPPLGGGGGIGYEANPGVPPATMSGSMMGSDMR | 1 | | | |
| SH3R1_HUMAN | IGVFPGNYVAPVTR | 1 | | | |
| | LQGNGVAGSPSVVPAAVVSAAHIQTSPQAK | 1 | | | |
| SI1L3_HUMAN | RPVSFPETPYTVSPAGADR | 1 | | | |
| SIX4_HUMAN | VLQSSANSATTTSYSPSVPVSFPGLIPSTEVK | 2 | | | |
| SKIL_HUMAN | TCTSVPETLHLNPSLK | 1 | | | |
| SLAI1_HUMAN | MPSTTAISSNISSPVTVR | 2 | | 1 | |
| SMAP2_HUMAN | STAPVMDLLGLDAPVACSIANSK | | | 1 | |
| SMG7_HUMAN | KTPVSEAR | 1 | | | |
| | YPNNSMFNEVYGK | 1 | | | |
| SNRPA_HUMAN | AVQGGGATPVVGAVQGPVPGMPPMTQAPR | 1 | | | |
| SON_HUMAN | ILDSFAAAPVPTTTLVLK | 4 | 2 | 2 | 2 |
| | SMMSSYSAADR | | | | 1 |
| SP2_HUMAN | TPSGEVQTVLVQDSPPATAAATSNTTCSSPASR | 2 | | 1 | |
| SPAT2_HUMAN | CDSLLTCPPASKPSAFPSK | 1 | | | |
| SPTB2_HUMAN | DDEEMNTWIQAISSAISSDKHEVSASTQSTPASSR | 2 | | | |
| | HEVSASTQSTPASSR | 3 | | 2 | 2 |
| SRC8_HUMAN | NASTFEDVTQVSSAYQK | 5 | | 2 | 1 |
| SRCAP_HUMAN | AETQGANHTPVISAHQTR | 1 | | 2 | |
| SRGP1_HUMAN | NSPTPATSTESLSPLHNVALR | 2 | | 1 | |
| SRP14_HUMAN | AAAAAAAAAPAAAATAPTTAATTAATAAQ | 2 | 2 | 4 | |
| SRRM2_HUMAN | IPAASAAAMNLASAR | 4 | 1 | 3 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| | MAPALSGANLTSPR | 1 | | | |
| | TPAALAALSLTGSGTPPTAANYPSSSR | 4 | | | |
| STAT3_HUMAN | FICVTPTTCSNTIDLPMSPR | 2 | 1 | 3 | |
| SUGP1_HUMAN | AQTSTDAPTSAPSAPPSTPTPSAGK | 3 | | 2 | |
| SYNJ1_HUMAN | IDPFEDLSFNLLAVSK | 1 | | | |
| SYNPO_HUMAN | VSTPATTTSTFSR | 1 | 1 | 1 | 1 |
| TAB1_HUMAN | VYPVSVPYSSAQSTSK | 2 | 1 | 2 | 2 |
| TAB2_HUMAN | VVVTQPNTK | 1 | | 1 | |
| TAF4_HUMAN | QVSQAQTTVQPSATLQR | 3 | | 4 | 1 |
| TAF9_HUMAN | ASIPATSAVQNVLINPSLIGSK | 2 | | 2 | |
| TAF9B_HUMAN | LSVGAVSSKPTTPTIATPQTVSVPNK | 4 | | | |
| TANC1_HUMAN | MSSSTSSLTSSSSFSDGFK | 1 | | | |
| TCF20_HUMAN | VGQFGQHYQSSASSSSSSSFPSPQR | 3 | | | |
| TIF1A_HUMAN | QWQISSGQGTPSTTNSTSSTPSSPTITSAAGYDGK | 2 | | | |
| TNR6A_HUMAN | ASNYNVPLSSTAQSTSAR | 3 | | 1 | |
| | LTWSPGSVTNTSLAHELWK | | | 1 | |
| TNR6B_HUMAN | DNTTGSNSSLNTSLPSPGAWPYSASDNSFTNVHSTSAK | 1 | | | |
| TOB1_HUMAN | STQPLTFTTATFAATK | 1 | | | |
| TOX4_HUMAN | GLQLGQTSTATIQPSQQAQIVTR | 5 | 2 | 3 | 2 |
| | QMLPSSITMSQGGMVTVIPATVVTSR | 6 | 1 | 3 | |
| TPD54_HUMAN | SWHDVQVSSAYVK | 1 | | | |
| TPR_HUMAN | GIASTSDPPTANIKPTPVVSTPSK | 2 | | | |
| TRI33_HUMAN | QHSNPGHAGPFPVVSVHNTTINPTSPTTATMANANR | 2 | | | |
| TRIP6_HUMAN | TGSLKPNPASPLPASPYGGPTPASYTTASTPAGPAFPVQVK | | | 1 | |
| TROAP_HUMAN | TSVSQASGLLLETPVQPAFSLPK | 1 | | | |
| UBAP2_HUMAN | GVSVSSSTTGLPDMTGSVYNK | 4 | 3 | 1 | 2 |
| | IPYQSPVSSSESAPGTIMNGHGGGR | 1 | | | |
| | LLQLPSTTIENISVSVHQPQPK | 6 | | 3 | |
| UBN1_HUMAN | TPASSSSALSHPAKPHSVSSAGSSYK | 1 | | | |
| UBP2L_HUMAN | QAFTPSSTMMEVFLQEK | 1 | | | |
| | SPAVATSTAAPPPPSSPLPSK | 7 | 3 | 5 | 3 |
| | TAQALAQLAAQHSQSGSTTTSSWDMGSTTQSPSLVQYDLK | 1 | | | |
| UBQL2_HUMAN | SQNRPQGQSTQPSNAAGTNTTSASTPR | 1 | | 2 | 4 |
| UTRO_HUMAN | VVLVSSASDIPVQSHR | 3 | | 1 | |
| VCIP1_HUMAN | TEPSVFTASSSNSELIR | 3 | 3 | 2 | 3 |
| VEZF1_HUMAN | KTPTTVVPLISTIAGDSSR | 11 | | 3 | |
| | TPTTVVPLISTIAGDSSR | 13 | 4 | 4 | 3 |
| | TSLVSTIAGILSTVTTSSSGTNPSSSASTTAMPVTQSVK | 5 | | 2 | |
| WAC_HUMAN | INEVLTAAVTQASLQSIIHK | 1 | | | |
| WNK1_HUMAN | EGPVLATSSGAGVFK | 3 | 2 | 3 | 1 |
| | FSAPGQLCISMTSNLGGSAPISAASATSLGHFTK | 3 | | 1 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| | QPIPASSMPQQIGIPTSSLTQVVHSAGR | 7 | 1 | 5 | 1 |
| | TLSPEMITVTSAVGPVSMAAPTAITEAGTQPQK | 11 | | 3 | |
| | VFPSEITDTVAASTAQSPGMNLSHSASSLSLQQAFSELR | 3 | | | |
| WNK3_HUMAN | QIMAPVTNSSSYSTTSVR | 1 | 1 | 2 | |
| XRN1_HUMAN | EAQSSQATPVQTSQPDSSNIVK | | | 2 | |
| YETS2_HUMAN | QLTTGSVVQGTLGVSTSSAQGQQTLK | 2 | | | |
| YTHD3_HUMAN | IGGDLTAAVTK | 2 | | 2 | |
| | TVGTALSSSGMTSIATNSVPPVSSAAPKPTSWAAIAR | 3 | | | |
| ZBT20_HUMAN | SVLQQPSVNTSIGQPLPSTQLYLR | 1 | | 1 | |
| ZC3HE_HUMAN | TSQEELLAEVVQGQSR | 1 | | | |
| ZCH14_HUMAN | TAQQPALVVETSTAATGTPSTVLHAARPPIK | 2 | | 2 | |
| ZEP1_HUMAN | LEQVYNIAVTSSVGLTSPSSR | 2 | | 1 | |
| | QVFLLSVPSLDCLPITR | 2 | | 1 | |
| | SNGPSAALVTTSTPSALPTGEK | 1 | 1 | 1 | |
| | SNSMPTTGYSAVPANIIPPPHPLR | 1 | | | |
| | VNIQEQSQQPVTSLSLFNIK | 2 | | | |
| ZEP2_HUMAN | ALYHNPPLSMGQYLQAEPIVLGPPNLR | 1 | 2 | | |
| | APQTLPLGLESSIPLCLPSTSDSVATLGGSK | 1 | | | |
| | DGLQSGSSSFSSLSPSSSQDYPSVSPSSR | 2 | | | |
| | EQTYPCYSGASGLHPK | 2 | | 3 | 1 |
| | SESAEQQISPPNTNAK | 1 | | | |
| | SFDYGNLSHAPVSGAAASTVSPSR | 2 | | 1 | 1 |
| | SNSVPTSSATNLTIPPSLR | 2 | | | |
| | STETPSEQVLQEDFASANAGSLQSLPGTVVPVR | 2 | | 3 | 1 |
| | TLVTNAAMQGIGFNIAQVLGQHAGLEK | 2 | | | 1 |
| ZFHX3_HUMAN | NFQHPLVSTANLIGPGHSFYGK | 1 | | 1 | |
| ZFR_HUMAN | AGYSQGATQYTQAQQTR | 5 | 2 | 3 | 5 |
| | QAAAAAAAAAATAAWTGTTFTK | 4 | | 2 | |
| | QQEAPPPPPPATTQNYQDSYSYVR | 8 | 3 | 7 | 2 |
| | QYYQQPTATAAAVAAAAQPQPSVAETYYQTAPK | 3 | 1 | 3 | |
| | STPVTSAVQIPEVK | 2 | | 1 | |
| ZHX1_HUMAN | EEIVENPSSSASESNTSTSIVNR | 2 | | 1 | |
| | ENEIKPDREEIVENPSSSASESNTSTSIVNR | 1 | | | |
| | IHPSTASTVVTPAAVLPGLAQVITAVSAQQNSNLIPK | 1 | | | |
| ZHX3_HUMAN | EGDHSFINGAVPVSQASASSAK | 2 | | 1 | |
| | VPEVTCIPTTATLATHPSAK | 2 | | 1 | |
| ZN281_HUMAN | TNESQISNNINMQSYSVEMPTVSSSGGIIGTGIDELQK | 1 | | | |
| ZN318_HUMAN | TVVAHTSPWMPVVTTSTQTK | 1 | | | |
| ZN384_HUMAN | GCGLAPPHYPTLLTVPASVSLPSGISMDTESK | 3 | | | |
| | SDQLTPHSQASVTQNITVVPVPSTGLMTAGVSCSQR | 4 | 1 | 2 | |
| ZN507_HUMAN | NETIPDIPVSVDNLQTHTVQTASVAEMGR | 2 | | | |

| | | | | | |
|---|---|---|---|---|---|
| ZN532_HUMAN | QVTIKPVATAFLPVSAVK | 1 | | | |
| ZNT6_HUMAN | GTDDLNPVTSTPAKPSSPPPEFSFNTPGK | 1 | | | |
| ZO1_HUMAN | TPSTEAAHIMLR | 1 | | | |
| ZYX_HUMAN | GPPASSPAPAPK | 1 | | | |
| | VSSGYVPPPVATPFSSK | 11 | 4 | 4 | 2 |
| | **Total Peptides** | *Dde-1* | *Dde-2* | *PC-1* | *PC-2* |
| | 442 | 407 | 118 | 241 | 133 |

## 5.6 Experimental Methods

*Reagents and materials.* All proteins, chemicals, and reagents were of analytical grade, obtained from Sigma Aldrich (St. Louis, MO), and used without further purification unless specified. The *O*-GlcNAcylated peptide TAPT(gS)TIAPG, high-capacity Neutravidin agarose resin, spin columns, and C18 desalting tips were purchased from ThermoFisher Scientific (Waltham, MA). Thiamet G was received from Tocris Biosciences (Avonmouth, Bristol, UK). cOmplete protease inhibitor cocktail without EDTA (PIC-EDTA) was purchased from Roche Diagnostics Corp. (Indianapolis, IN). Baculovirus preparation and protein expression of short-form OGT (sOGT) in Spodoptera frugiperda (*Sf9*) cells was performed as previously described.[24] Cerebral cortices were obtained from adult C57BL/6 mice bred in house. All protein concentrations were measured using a

BCA assay kit (ThermoFisher Scientific). RapiGest and UDP-GalNAz were synthesized as referenced.[35, 36] Y289L GalT was expressed and purified as described previously. Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), alkyne-Dde-biotin **5-4**, the alkyne photocleavable biotin linker (alkyne-PC-biotin, **5-1**), Cy3-azide, and azide agarose beads were purchased from Click Chemistry Tools (Scottsdale, AZ).

*Peptide labeling.* The labeling protocol was adapted from a previously reported method. The peptide TAPT(gS)TIAPG (20 μM final) was dissolved in a 200 μL solution of 10 mM HEPES pH 7.9, 5.5 mM $MnCl_2$, 1 mM UDP-GalNAz, and 100 ng/μL Y289L GalT and rotated end-over-end overnight at 4 °C. Prior to enzyme addition, an aliquot was removed as an initial time point for LC-MS analysis. The reaction was acidified to 0.1% TFA, desalted using a C18 tip, and an aliquot was saved for analysis. The labeled peptide (10 μM final) was diluted into a 400 μL solution of 10 mM sodium phosphate pH 7.6, 100 μM alkyne-Dde-biotin **5-4**, 2 mM sodium ascorbate, and 100 μM THPTA. $CuSO_4$ was added (1 mM final), and the reaction was incubated while rotating end-over-end at RT for 1 h. After removing a sample, the reaction was acidified and desalted again. The peptide (10 μM) was then split into fractions of 50 μL containing 25 mM sodium phosphate pH 7.6 and either 1% RapiGest, 6 M urea, or 2% hydrazine monohydrate and incubated for 1 h at RT. Samples were acidified, desalted, and subjected to LC-MS analysis.

*LC-MS analysis of O-GlcNAc peptide labeling.* Liquid chromatography and mass spectrometry (LC-MS) were performed using an LTQ linear ion trap mass spectrometer combined with an Accela LC and PAL autosampler (Thermo Scientific, Waltham, MA). Approximately 10 pmol peptide from each sample was injected onto a CORTECS UPLC C18+ column (2.1 x 50 mm, Waters Corp., Milford, MA). Flow rate was set at 0.4 mL/min. Solvent A (ddH2O, 1% formic acid) and Solvent B (acetonitrile, 1% formic acid) were used to create a gradient. The gradient consisted of 0-0.2 min, 5% B; 0.2-3.5 min 5-65% B, 3.5-4.0 min 65% B with injection into the MS starting at 0.2 min to avoid salt contamination. All peptide products were found to elute during the linear gradient between 0.2 and 2.0 min. For the biotinylated and cleaved peptide, alkyne reagent 2 was not sufficiently removed by the C18 tips. Therefore, the reaction was monitored using an extracted ion chromatogram by extracting all ions within ± 1 m/z of the calculated masses.

*Linker comparison by protein labeling.* Labeling with Y289L GalT and UDP-GalNAz was conducted as previously described.[3] Briefly, 500 μg of HEK-293T cell lysate in 1% SDS, 1x PBS pH 7.4 (10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, 137 mM NaCl, 2.7 mM KCl), 10 μM Thiamet G, and 1x PIC-EDTA was diluted to a protein concentration of 1 mg/mL using 1% SDS, 1x PBS pH 7.4 and precipitated by adding three volumes of methanol, one volume of chloroform, and two volumes of ddH2O with vortex mixing after each addition. Protein was pelleted at the aqueous-organic interface by centrifuging at 21,000 x *g* for 5 min. The top,

aqueous layer was removed, and one volume of methanol was added with mixing. The protein was pelleted again, and all liquid was removed. After the pellets were air-dried, samples were redissolved at 5 mg/mL (100 μL) in 1% SDS, 20 mM HEPES pH 7.9 by sonication. To each sample, the following were added in the given order: 10 μL of 50x PIC-EDTA, 112.5 μL of $ddH_2O$, 200 μL of 2.5x labeling buffer (50 mM HEPES pH 7.9, 125 mM NaCl, 5% NP-40), and 27.5 μL of 100 mM MnCl2. Samples were briefly mixed by pipetting and placed on ice. Next, 25 μL of 0.5 mM UDP-GalNAz was added followed by pipetting to mix. Finally, 25 μL of 2 mg/mL Y289L GalT was added, and samples were rotated end-over-end for 16 h at 4 °C. Proteins were then precipitated again, and pellets were air dried. Pellets were next dissolved at 4 mg/mL (125 μL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of the GalT-labeled sample was removed (25 μL), and the remaining sample was split into two equal portions (50 μL each) to be labeled with either alkyne-Dde-biotin **5-4** or alkyne-PC-biotin **5-1**. All manipulations with the photocleavable linker were performed in the dark. To each sample, the following were added in the given order with mixing: 4 μL of 50x PIC-EDTA, 78 μL of $ddH_2O$, and 10 μL of 20x PBS pH 7.4. Next, the CuAAC reagents were added with vortex mixing after each addition: 4 μL of 5 mM alkyne-Dde-biotin **5-4** or alkyne-PC-biotin **5-1** (stock in DMSO), 4 μL of 100 mM sodium ascorbate (freshly prepared), 10 μL of 2 mM THPTA (stock in 4:1 *t*BuOH/DMSO), and 4 μL of 50 mM $CuSO_4$ (freshly prepared). Samples were rotated endover-end for 1 h at RT, and the reaction was halted by the addition of 25 μL EDTA pH 8.0. Samples were

acetone precipitated, and the pellet was washed once with 1 mL MeOH to remove residual, unreacted linker. The pellet was then air-dried and redissolved at 4 mg/mL (50 μL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of each sample was reserved (25 μL), and the remaining sample was cleaved. For the Dde sample, the mixture was diluted to 1 mg/mL with 2% hydrazine monohydrate, and the sample was rotated end-over-end for 1 h at RT. For the PC sample, the protein was diluted to 1 mg/mL with $ddH_2O$, and the liquid was irradiated from the open top of the tube (2 cm distance) with 365 nm UV light (UVGL-25 handheld UV lamp, 1.5 $mW/cm^2$) for 1 h at RT with mixing every 10 min. Both samples were then precipitated by addition of four volumes of -20 °C acetone and storage at -20 °C for 1 h. Samples were air-dried and then redissolved at 4 mg/mL (25 μL) in 1% SDS, 20 mM HEPES pH 7.9.

*Coomassie staining and streptavidin blotting.* Aliquots corresponding to 20 μg of protein (5 μL) for each sample were resolved by SDS-PAGE as follows. Samples were added to 10 μL $ddH_2O$ and 5 μL 4x SDS-PAGE loading buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) and were then used directly without boiling to avoid cleaving the linkers. The mixtures were loaded in duplicate on NuPAGE Novex 4-12% Bis-Tris protein gels (ThermoFisher Scientific, 1.0 mm, 10-well). One duplicate was then stained with Imperial protein stain (ThermoFisher Scientific) according to the manufacturer's specifications and imaged by an Odyssey scanner (LI-COR Biosciences). The other set of samples was transferred onto an Immobilon-FL PVDF membrane

(EMD Millipore, 0.45 μM) and blocked for 1 h at RT with 5% bovine serum albumin (BSA) in 1x TBST (19 mM Tris pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween 20). The blot was then incubated with 1:20,000 AlexaFluor-680-conjugated streptavidin (ThermoFisher Scientific) in 5% BSA/TBST for 1 h at RT, washed thrice with TBST for 5 min, and imaged with an Odyssey scanner.

*α-Crystallin and sOGT labeling.* Labeling was conducted as described above using 200 μg of cortical lysate, 20 μg of α-crystallin, and 5 μg of sOGT.

*Enrichment and elution of labeled proteins.* Labeled samples were diluted to 1 mL using 1x PBS pH 7.4, 1x PIC-EDTA. For each sample, 25 μL (settled volume) of high-capacity Neutravidin agarose was washed twice with 500 μL of 1x PBS in spin columns, and samples were added to the washed beads. Mixtures were rotated end-over-end for 1 h at RT. Lysate was removed by centrifugation at 2,000 x *g* for 30 s. The beads were washed with 1% SDS (5 x 0.5 mL), 6 M urea (5 x 0.5 mL), and 1x PBS pH 7.4 (5 x 0.5 mL). Beads were then resuspended in 50 μL of 2% hydrazine monohydrate in dd$H_2O$ and rotated end-over-end for 1 h at RT. The elution volume was removed, and beads were washed with 50 μL of 1x PBS pH 7.4. The wash volume was combined with the elution volume, and samples were flash-frozen and stored at -80 °C prior to analysis.

*Processing proteins for MS analysis.* Samples were thawed and precipitated by addition of four volumes of - 20 °C acetone. Samples were stored at -20 °C for 1 h and centrifuged at 21,000 x *g* for 5 min. Pellets were redissolved in 20 μL of 8 M urea, 100 mM Tris pH 8.0, 10 mM DTT and incubated at 60 °C with shaking for 20

min. Cysteine residues were blocked by addition of 25 mM IAA for 45 min. Samples were diluted four-fold with 100 mM Tris pH 8.0. Samples were split in two and digested with 0.01 mg/mL trypsin or chymotrypsin for 4-16 h at 37 °C. A portion of the trypsin digests was further digested with 7 μg/mL AspN for 6 h at 37 °C. Digests were acidified to a final concentration of 0.5% formic acid, 0.05% TFA.

*LC separation and MS analysis.* The digests were analyzed by nanoLC/MS on the LTQ-Velos with a 0 to 30% B in 120 min gradient with top 5 MS/MS (A: dd$H_2O$, 0.1% formic acid; B: acetonitrile, 0.1% formic acid). Samples were desalted on a 360 x 100 μm Kasil fritted pre-column (2 cm Monitor C18) prior to separation on a 360 x 75 μm (10 cm BEH130 C18, 1.7 μm) analytical column/tip. Full scan MS was acquired at 60,000 resolution followed by top 5 tandem MS in the linear ion trap alternating between ETD and CID modes of the same precursor. The ETD reaction time was 100 ms with supplemental activation. RAW files were converted to MGF files for Mascot searching using Proteome Discoverer with CID and ETD spectra extracted to separate MFG files. Data was searched against a custom database with fixed modifications of carbamidomethyl (C) and variable mods of oxidation (M) and a custom modification for the tagged *O*-GlcNAc. The custom modification was defined as addition of $C_{19}H_{30}N_6O_{10}$ to Ser or Thr (net addition of 502.202341 Da). For CID, a scoring neutral loss of $C_{19}H_{30}N_6O_{10}$ was included, but this was omitted for ETD. Enzyme specificity was trypsin (KR), chymotrypsin (FLYW), or trypsin-AspN_ND (cleave C-term KR and N-term

ND). Mass tolerances were 25 ppm and 0.8 Da for precursor and fragments ions, respectively. The instrument type was chosen as either ESI-TRAP or ETD-TRAP. Search results were combined in Scaffold 4.4, filtered for 80% peptide confidence and modifications manually evaluated.

*Testing linker in on-bead digestion conditions.* Cy3-Dde-biotin **5-5** was synthesized using conditions similar to protein labeling. To 15 µL 20x PBS and 259 µL dd$H_2O$, the following reagents were added in order: 6 µL 5 mM alkyne-Dde-biotin **5-4**, 5 µL 5 mM Cy3-azide (0.83 eq), 6 µL 100 mM NaAsc, 3 µL 10 mM THPTA, and 6 µL 50 mM $CuSO_4$. The reaction was incubated for 1 h at RT, and the crude product was used directly without purification. Eight 40 µL aliquots of the reaction were diluted to 500 µL in PBS and were added to 20 µL settled volume high-capacity Neutravidin resin. The mixture was incubated for 1 h at RT, and then the beads were washed five times with PBS to remove excess dye. An image of the samples was taken. Samples were then incubated in 500 µL buffer (PBS; 2 M urea, PBS; 0.1% SDS, PBS; 0.1% SDS, 20 mM HEPES pH 7.9) overnight at RT or 37 °C with end-over-end rotation. Beads were pelleted the following day, and a second image of the samples was taken. Because of the obvious difference between samples by eye, Cy3 released into solution was not quantified. To test if the beads were playing a role in linker cleavage, four 40 µL aliquots were diluted to 500 µL in PBS, and only two were immobilized on Neutravidin beads as described above. The beads were then rotated end-over-end overnight at RT or 37 °C. Samples that had not been incubated with beads overnight were then immobilized on Neutravidin

beads as described above to capture all linker-bound Cy3. All beads were washed twice with PBS to remove all released Cy3, and an image of the samples was taken.

*Cell culture.* All reagents used for cell culture were obtained from ThermoFisher Scientific unless otherwise noted. The HEK-293T cell line and HUVECs were obtained from the American Type Culture Collection. HEK-293T cells were cultured in DMEM with 10% fetal bovine serum (FBS) and 1x penicillin/streptomycin (P/S) (referred to now on as complete DMEM).

*Protein digestion.* All work prior to trypsinization was performed in a laminar flow hood to minimize keratin contamination. Protein lysate from HEK-293T cells in 1% SDS, 50 mM Tris pH 7.6, 150 mM NaCl, 1x PIC- was quantified using the BCA assay (ThermoFisher Scientific), and 20 mg was diluted to 2 mg/mL in lysis buffer. Samples were reduced by incubation for 5 min at 95 °C with 20 mM DTT. Two 15-mL Amicon centrifugal filter units (10 kDa NMWL, UFC901024, EMD Millipore) were centrifuged at 4,000 x *g* with 5 mL dd$H_2O$ to wet the filters, and both the retentate and filtrate were discarded. The reduced sample was split in two (10 mg per filter unit) and then concentrated to ~1 mL by centrifugation at 4,000 x *g* at RT. All remaining centrifugation steps occur at 4,000 x *g* at RT. No time is specified since the rate at which buffer exchange occurs is not always equivalent. All samples should be concentrated to ~1 mL before proceeding to the next step. Solutions of 8 M urea, 50 mM Tris pH 7.9 and 50 mM Tris pH 7.9 were sterile filtered to remove all particulate matter. Samples were diluted to 15

mL in the filter units with 8 M urea, 50 mM Tris pH 7.9 containing 20 mM IAA and allowed to alkylate for 30 min at RT in the dark. The samples were then concentrated and buffer exchanged twice with 15 mL 8 M urea, 50 mM Tris pH 7.9 and twice with 15 mL 50 mM Tris pH 7.9. Alternative 1: The final Tris exchange step was replaced with 15 mL 50 mM Tris pH 7.9, 0.2% w/v sodium deoxycholate that had been sterile filtered. Alternative 2: Prior to the first urea exchange step, an additional exchange step was added using 15 mL 8 M urea, 50 mM Tris pH 7.9, 4% w/v sodium deoxycholate that had been sterile filtered. Samples were then diluted up to 5 mL with 50 mM Tris pH 7.9, 1 mM $CaCl_2$ in the filter units and treated with 1:50 w/w sequencing-grade trypsin (90305, ThermoFisher Scientific). Samples were rotated end-over-end overnight at 37 °C. Peptides were separated from undigested protein by centrifugation for 5 min at 4,000 x *g*, and the filter was washed twice with 2 mL 50 mM Tris pH 7.9 and centrifuged. All filtrates were combined and acidified to pH 3 with TFA. Samples were centrifuged for 5 min at 15,000 x *g* without the filter cup to pellet any residual sodium deoxycholate. Solutions were combined and desalted using a 20 cc HLB cartridge. Peptide concentrations were measured using a NanoDrop 2000C (ThermoScientific) with absorbance at 205 nm ($\varepsilon_{205}{}^{1\ mg/mL} = 31$).

*HLB desalting.* Oasis HLB cartridges (3 cc, WAT094226; Plus short, 186000132; 20 cc, 186000117; Waters Corp.) were first conditioned with methanol (1 mL, 5 mL, 10 mL) and then equilibrated with 0.1% TFA (2 mL, 10 mL, 20 mL) on a vacuum manifold (WAT200677, Waters Corp.). Acidified samples were then

applied to the cartridge and allowed to flow through at a rate of ~1 drop per second. Columns were then washed with 0.1% TFA (2 mL, 10 mL, 20 mL), and samples were eluted in 70% acetonitrile, 0.1% TFA (1 mL, 5 mL, 10 mL). 0.1% FA was used for the wash and elution steps for the clean up step immediately prior to MS analysis. Peptides were aliquoted as 5-mg samples and were dried by either vacuum centrifugation or freeze-drying.

*GalT labeling of peptides*. The peptide sample (5 mg) was resuspended in 1 mL dd$H_2O$. Peptides may not dissolve readily and may require sonication and vortexing. To the solution was added 1 mL 100 mM HEPES pH 7.9, 275 μL 100 mM $MnCl_2$, 50 μL 5 M NaCl, and 2.15 mL dd$H_2O$. The pH was checked at this point to ensure that any residual TFA from HLB elution did not acidify the solution lower than pH 7.5. Then, to the solution was added 250 μL 0.5 mM UDP-GalNAz, 250 μL 2 mg/mL Y289L GalT, and 20 μL PNGase F (P0705L, New England Biolabs). The reaction was rotated at overnight at RT. Some Y289L GalT may precipitate out, but this does not affect labeling. The following day, the reaction was acidified to pH 3 with TFA. Any precipitated protein was pelleted by centrifugation, and the solution was desalted using an HLB Plus short cartridge.

*CuAAC reaction of peptides*. Dried peptides were resuspended in 1 mL dd$H_2O$ with sonication and vortexing, if necessary. A 2x CuAAC reagent mixture was produced in the following order: 860 μL dd$H_2O$, 40 μL 5 mM alkyne-Dde-biotin **5-4** or 5 mM alkyne-PC-biotin **5-1**, 40 μL 100 mM NaAsc, 20 μL 10 mM THPTA, and 40 μL 50 mM $CuSO_4$. 1 mL of the 2x mixture was immediately added to the

peptide solution, and the reaction was rotated end-over-end for 2 h at RT. 100 μL settled volume azide agarose resin (1038-2, Click Chemistry Tools) was washed thrice with $ddH_2O$, resuspended in 1 mL $ddH_2O$, and then added to the reaction. The bead-reaction mixture was rotated end-over-end for 1 h at RT. The reaction was filtered, and beads were rinsed twice with 1 mL $ddH_2O$. Samples were acidified to pH 3 with TFA and desalted using an HLB Plus short cartridge.

*Peptide enrichment and elution.* Dried peptides were resuspended in 1 mL $ddH_2O$ with sonication and vortexing, if necessary. The sample was diluted to 10 mL with PBS. 0.5 mL settled volume of high-capacity Neutravidin beads were washed thrice with PBS and then added to the sample. The suspension was rotated end-over-end for 2 h at RT. The suspension was transferred to a column, and beads were rinsed with 5 mL PBS, 5 mL PBS + 2 M NaCl, and 5 mL PBS. Beads were transferred to a clean tube for elution. For **5-4**, samples were rotated for 2 h at RT with 2 mL 100 mM sodium phosphate pH 6.6, 2% w/v hydroxylamine (final pH 7.2). For **5-1**, samples were irradiated (2 cm distance) with 365 nm UV light (UVGL-25 handheld UV lamp, 1.5 $mW/cm^2$) for 1 h at RT with mixing every 10 min. Elution buffer was removed, and beads were washed twice with 1 mL 100 mM sodium phosphate pH 7.2. Washes were combined with elution buffer as the first elution fraction. The elution procedure was repeated as described above and combined as the second elution fraction. Samples were acidified to pH 3 with TFA and desalted using and HLB 6 cc cartridge. Dried samples were stored at -20 °C prior to MS analysis as described above.

## 5.7 References


1. J. Ma, G. W. Hart. *O*-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics*. **2014**, *11*: 8.
2. N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA*. **2004**, *101*: 13132-13137.
3. P. M. Clark *et al.* Direct in-gel fluorescence detection and cellular imaging of *O*-GlcNAc-modified proteins. *J Am Chem Soc*. **2008**, *130*: 11576-11577.
4. N. Khidekel *et al.* Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol*. **2007**, *3*: 339-348.
5. J. C. Trinidad *et al.* Global identification and characterization of both *O*-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics*. **2012**, *11*: 215-229.
6. B. W. Zaro, Y.-Y. Yang, H. C. Hang, M. R. Pratt. Chemical reporters for fluorescent detection and identification of *O*-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc Natl Acad Sci USA*. **2011**, *108*: 8146-8151.
7. J. F. Alfaro *et al.* Tandem mass spectrometry identifies many mouse brain *O*-GlcNAcylated proteins including EGF domain-specific *O*-GlcNAc transferase targets. *Proc Natl Acad Sci USA*. **2012**, *109*: 7280-7285.
8. Z. Wang *et al.* Enrichment and site mapping of *O*-Linked *N*-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage,

and electron transfer dissociation mass spectrometry. *Mol Cell Proteomics.* **2010**, *9*: 153-160.

9. N. Khidekel *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc.* **2003**, *125*: 16162-16163.

10. B. Ramakrishnan, P. K. Qasba. Structure-based design of beta 1,4-galactosyltransferase I (β4Gal-T1) with equally efficient *N*-acetylgalactosaminyltransferase activity: point mutation broadens β4Gal-T1 donor specificity. *J Biol Chem.* **2002**, *277*: 20833-20839.

11. C. S. McKay, M. G. Finn. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol.* **2014**, *21*: 1075-1101.

12. J. Szychowski *et al.* Cleavable biotin probes for labeling of biomolecules via azide–alkyne cycloaddition. *J Am Soc Chem.* **2010**, *132*: 18351-18360.

13. C. M. Woo, A. T. Iavarone, D. R. Spiciarich, K. K. Palaniappan, C. R. Bertozzi. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact *N*- and *O*-glycopeptide discovery and analysis. *Nat Methods.* **2015**, *12*: 561-567.

14. S. A. Myers, S. Daou, E. B. Affar, A. L. Burlingame. Electron transfer dissociation (ETD): the mass spectrometric breakthrough essential for *O*-GlcNAc protein site assignments—a study of the *O*-GlcNAcylated protein host cell factor 1. *Proteomics.* **2013**. *12*: 981-991.

15. C. M. Woo *et al.* Development of IsoTaG, a chemical glycoproteomics technique for profiling intact *N*- and *O*-glycopeptides from whole cell proteomes. *J Proteome Res.* **2017**, *16*: 1706-1718.

16. M. Boyce *et al.* Metabolic cross-talk allows labeling of *O*-linked β-*N*-acetylglucosamine-modified proteins via the *N*-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci USA*. **2011**, *108*: 3141-3146.

17. B. W. Bycroft, W. C. Chan, S. R. Chhabra, N. D. Hone. A novel lysine-protecting procedure for continuous-flow solid-phase synthesis of branched peptides. *J Chem Soc Chem Commun.* **1993**: 778-779.

18. S. R. Chhabra, H. Parekh, A. N. Khan, B. W. Bycroft, B. Kellam. An appraisal of new variants of Dde amine protecting group for solid phase peptide synthesis. *Tetrahedron Lett*. **1998**, *42*: 2189-2192.

19. Y. Yang, S. H. L. Verhelst, Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem Comm*. **2013**, *49*: 5366-5368.

20. R. J. Chalkley, A. L. Burlingame. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J Am Soc Mass Spectrom*. **2001**, *12*: 1106-1113.

21. J. A. Hanover *et al.* Mitochondrial and nucleocytoplasmic isoforms of *O*-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys*. **2003**, *409*: 287-297.

22. B. D. Lazarus, D. C. Love, J. A. Hanover. Recombinant *O*-GlcNAc transferase isoforms: identification of *O*-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology*. **2006**, *16*: 415-421.

23. E. P. Roquemore *et al.* Vertebrate lens alpha-crystallins are modified by *O*-linked *N*-acetylglucosamine. *J Biol Chem*. **1992**, *267*: 555-563.

24. H.-C. Tai, N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson. Parallel identification of *O*-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc*. **2004**, *126*: 10500-10501.

25. Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu. TET2 promotes histone *O*-GlcNAcylation during gene transcription. *Nature.* **2012**, *493*: 561-564.

26. N. Zeytuni, R. Zarivach. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure.* **2012**, *20*: 397-405.

27. S. P. Iyer, G. W. Hart. Roles of the tetratricopeptide repeat domain in *O*-GlcNAc transferase targeting and protein substrate specificity. *J Biol Chem.* **2003**, *278*: 24608-24616.

28. M. E. Griffin *et al.* Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag. *Mol Biosyst.* **2016**, *12*: 1756-1759.

29. C. M. Woo, A. Felix, L. Zhang, J. E. Elias, C. R. Bertozzi. Isotope-targeted glycoproteomics (IsoTaG) analysis of sialylated *N*- and *O*-glycopeptides on an Orbitrap Fusion Tribrid using azido and alkynyl sugars. *Anal Bioanal Chem.* **2017**, *409*: 579-588.

30. K. Augustyns, W. Kraas, G. Jung. Investigation on the stability of the Dde protecting group used in peptide synthesis: migration to an unprotected lysine. *J Pept Res.* **1998**, *51*: 127-133.

31. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann. Universal sample preparation method for proteome analysis. *Nat Methods.* **2009**, *6*: 359-362.

32. J. Erde, R. R. Loo, J. A. Loo. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res.* **2014**, *13*: 1885-1895.

33. D. Pellerin, H. Gagnon, J. Dubé, F. Corbin. Amicon-adapted enhanced FASP: an in-solution digestion-based alternative sample preparation method to FASP. *F1000Research.* **2015**, *4*: 140.

34. J. Yang *et al.* Global, *in situ*, site-specific analysis of protein S-sulfenylation. *Nat Protoc.* **2015**, *10*: 1022-1037.

35. P. J. J. Lee, B. J. Compton. U. S. Patent Office. **2007**, 7,229,539.

36. H. C. Hang, C. Yu, M. R. Pratt, C. R. Bertozzi. Probing glycosyltransferase activities with the Staudinger ligation. *J Am Chem Soc*. **2004**, *126*: 6-7.

*Appendix A*

# Expedient Synthesis of Chondroitin Sulfate Glycomimetic Polymers

## A.1 Rationale

The activity of sulfated glycosaminoglycans (GAGs) depends on protein binding, which occurs through multiple protein-carbohydrate interactions along the length of an oligosaccharide. However, the chemical synthesis of defined oligosaccharides is extremely difficult, necessitating alternative approaches to obtain higher order, chemically pure carbohydrates. One method used by our lab is the production of glycomimetic polymers based on ring-opening metathesis polymerization (ROMP). In this procedure, disaccharide units of known sulfation pattern are appended to a ROMP monomer.[1-4] After polymerization with a ruthenium-based catalyst, the formed structure displays a number of disaccharide units on a single backbone as controlled by the catalyst loading, providing a macromolecular scaffold that mimics the natural multivalent presentation of GAG sulfation patterns. Unfortunately, production of the disaccharide starting material was still an arduous task, requiring nine and eight steps for the glucuronic acid (GlcA) and galactosamine (GalNAc) monomers, respectively, followed by a coupling step to reach the fully protected dissacharide building block for further sulfation steps.[1, 2, 5] Here, we describe the adaptation of an efficient hydrolysis of natural chondroitin sulfate (CS) polysaccharides[6] to produce an equivalent disaccharide building block in only four steps on a multi-gram scale and its elaboration into CS sulfation epitope monomers and the respective polymers.

## A.2 General Chemical Methods

Unless otherwise noted, reactions were performed under Ar in sealed, flame-dried glassware using dry, deoxygenated solvents by passing through a column of activated alumina with Ar. Commercially obtained reagents were used as received. Thin-layer chromatography (TLC) was performed using E. Merck silica gel 60 F254 precoated plates (0.25 mm) and visualized by UV fluorescence quenching or staining with potassium permanganate, ceric ammonium molybdate, *p*-anisaldehyde, or ninhydrin. Flash column chromatography was performed using ICN silica gel (particle size 0.032 – 0.063 mm). Ion exchange chromatography was performed using Amberlite IR-120 $H^+$ form (Sigma Aldrich, washed with 1 M HCl, dd$H_2O$, MeOH, then $Et_2O$) or $Na^+$ form (Sigma Aldrich). For sulfated compounds, gel filtration chromatography was used with Sephadex LH-20 gel resin (GE Healthcare Life Sciences) preswollen in 1:1 $CH_2Cl_2$/MeOH. Polymers were purified using gel filtration chromatography with Sephadex G-25 Fine gel resin (GE Healthcare Life Sciences) preswollen in dd$H_2O$.

$^1$H NMR spectra were recorded on either a Varian Inova 500 (500 MHz) or a Varian Mercury 300 (300 MHz) and are reported relative to the solvent residual peak ($CDCl_3$ at δ 7.26, $CD_3OD$ at 3.31, and $D_2O$ at 4.79). Data for NMR spectra are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, b = broad), coupling constant in Hz, and integration. HRMS were acquired using an Agilent 6200 Series TOF with an Agilent G1978A Multimode source in electrospray or mixed mode ionization.

## A.3 Norbornenyl Linker Synthesis

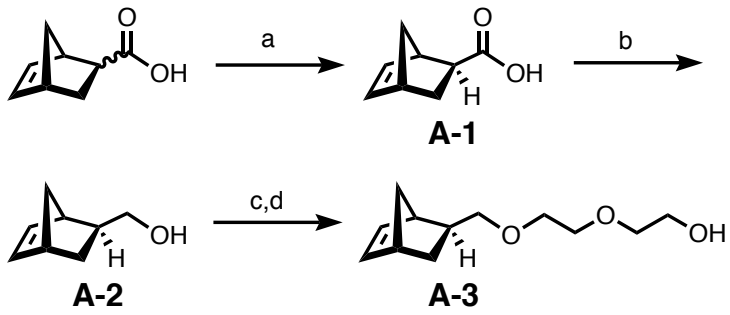



**Figure A-1. Synthesis of norbornene linker.**[2] (a) KI (3 eq), $I_2$ (1 eq), 0.75 M $NaHCO_3$, 1 h, RT, 90% recovery of *exo* isomer; (b) $LiAlH_4$ (3 eq), $Et_2O$, 0° to 50°C, 97% yield; (c) NaH (1.33 eq), 2-(2-((*tert*-butyldimethylsilyl)oxy)ethoxy) ethyl methanesulfonate (0.83 eq), 18-crown-6 (0.17 eq), 4 Å mol. sieves, THF, 0° to RT, 3 h, then 0.5 M TBAF, THF, 0° to RT, 3 h, 43% over two steps.

*(1R, 2S, 4R)-Bicyclo[2.2.1]hept-5-ene-2-carboxylic acid* (**A-1**). A 4:1 mixture of *endo*- and *exo*-5-norbornene-2-carboxylic acid (25.0 g, 181 mmol) was dissolved in 820 mL 0.75 M $NaHCO_3$. KI (90.12 g, 543 mmol, 3 eq) and $I_2$ (45.94 g, 181 mmol, 1 eq) were dissolved separately in 115 mL $H_2O$. This brown solution was added dropwise to the stirring solution of the carboxylic acid using an addition funnel. Once completely added, the solution was extracted with $Et_2O$ (5 x 150 mL), and the organic layers were discarded. After decolorizing with 10% $Na_2SO_3$, the pH of the aqueous layer was lowered to 2 using 1 N $H_2SO_4$, and the solution was extracted again with $Et_2O$ (5 x 150 mL). The combined organic layers were dried ($MgSO_4$), filtered, and concentrated to give the pure *exo* enantiomer (4.52 g, 90%) as a clear oil. If desired, the product could be further purified by precipitating in pentane at -78 °C; however, the product was

generally used directly in the next step. $^1$H NMR (300 MHz, $CDCl_3$) δ 6.14 (tt, *J* = 7.1, 4.1 Hz, 2H), 3.12 (d, *J* = 2.3 Hz, 1H), 2.95 (s, 1H), 2.28 (ddt, *J* = 8.4, 4.2, 1.7 Hz, 1H), 2.03 – 1.89 (m, 1H), 1.54 (dd, *J* = 8.3, 1.7 Hz, 1H), 1.41 (td, *J* = 9.4, 1.9 Hz, 2H).

*(1R, 2S, 4R)-Bicyclo[2.2.1]hept-5-ene-2-ylmethanol* (**A-2**). *Exo*-5-norbornene-2-carboxylic acid **A-1** (4.52 g, 32.7 mmol) was dissolved in dry $Et_2O$ (80 mL) and cooled to 0 °C. Lithium aluminum hydride (3.75 g, 98.8 mmol, 3 eq) was then added portionwise with vigorous stirring between additions. The resulting suspension was refluxed at 50 °C. After 5 h, the suspension was cooled down to 0 °C and excess lithium aluminum hydride was quenched (4 mL $H_2O$, 8 mL 10% NaOH, then 12 mL $H_2O$). The white aluminum salts were filtered washed, and the solution was concentrated. Column chromatography (5:1 to 3:1 hexanes/EtOAc) gave the product (3.93 g, 97%) as a volatile, clear oil. $R_f$ 0.24 (5:1 hexanes/EtOAc); $^1$H NMR (300 MHz, $CDCl_3$) δ 6.09 (qd, *J* = 5.6, 2.8 Hz, 2H), 3.70 (dd, *J* = 10.6, 6.4 Hz, 1H), 3.54 (dd, *J* = 10.6, 8.8 Hz, 1H), 2.82 (s, 1H), 2.75 (s, 1H), 1.68 – 1.55 (m, 1H), 1.38 – 1.19 (m, 3H), 1.11 (ddd, *J* = 11.7, 4.5, 3.5 Hz, 1H).

*2-(2-((1R,2S,4R)-Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)ethoxy)ethanol* (**A-3**). *Exo*-5-norbornenyl-2-methanol **A-2** (11.10 g, 89 mmol, 1.2 eq) was added to dry THF (80 mL) with 4 Å molecular sieves under Ar. The mixture was stirred at RT for 30 minutes then cooled to 0 °C. NaH (2.83 g, 118 mmol, 1.6 eq) was added portionwise, and the resulting mixture was stirred for 30 minutes at 0 °C. 2-(2-

((*tert*-butyldimethylsilyl)oxy)ethoxy) ethyl methanesulfonate (22.23 g, 74 mmol) and 18-crown-6 (3.94 g, 15 mmol, 0.2 eq) were then added. The reaction was allowed to warm to RT while stirring for 3 h. After quenching with a minimum of $H_2O$, the mixture was diluted with $CH_2Cl_2$ and filtered through Celite. The organic layer was washed with saturated $NaHCO_3$, dried by $MgSO_4$, filtered, and concentrated. The crude was then redissolved in dry THF (120 mL) and cooled to 0 °C. 1 M TBAF in THF (150 mL) was added dropwise, and the reaction was stirred for 3 hours while warming to RT. The mixture was diluted with EtOAc and washed with $H_2O$. The organic layer was dried ($MgSO_4$), filtered, and concentrated. Column chromatography (1:1 hexanes/EtOAc) yielded the pure product (6.78 g, 43% over two steps) as a yellow oil. $R_f$ 0.26 (1:1 hexanes/EtOAc); $^1$H NMR (500 MHz, $CDCl_3$) δ 6.10 (dd, *J* = 5.7, 3.1 Hz, 1H), 6.05 (dd, *J* = 5.7, 2.9 Hz, 1H), 3.76 – 3.72 (m, 2H), 3.70 (t, *J* = 4.7 Hz, 2H), 3.66 – 3.57 (m, 4H), 3.53 (dd, *J* = 9.4, 6.3 Hz, 1H), 3.39 (t, *J* = 9.2 Hz, 1H), 2.80 (s, 1H), 2.74 (s, 1H), 2.17 (bs, 1H) 1.75 – 1.67 (m, 1H), 1.36 – 1.21 (m, 3H), 1.10 (ddd, *J* = 11.7, 4.3, 3.4 Hz, 1H); HRMS: calculated $[M+H]^+$ 213.1491, determined: 213.1473.

## A.4 Production of the CS Disaccharide Building Block

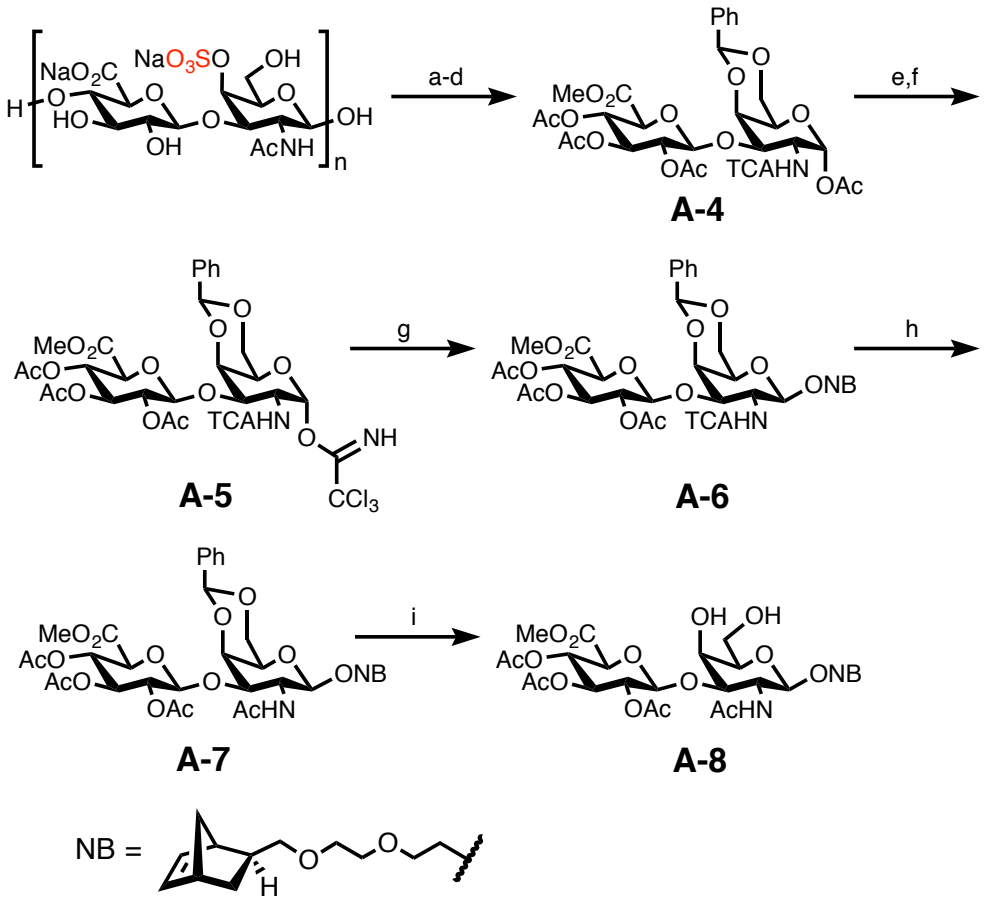



**Figure A-2. Synthesis of core CS monomer building block.** (a) 0.5 M $H_2SO_4$, $H_2O$, reflux, 6 h; (b) 0.02 M HCl, MeOH, 0 °C, 4 d; (c) TCACl (11.3 eq), pyridine, 0 °C, 1 h, then MeOH, DCM, pyridine, RT, 4 h; (d) benzaldehyde (36.5 eq), TFA (2.64 eq), neat, RT, 1 h, then $Ac_2O$ (27.3 eq), NaOAc (3.9 eq), pyridine, RT, 16 h, 11.5% yield over four steps; (e) hydrazine acetate (1.5 eq), DMF, RT, 1 h; (f) trichloroacetonitrile (8 eq), DBU (0.6 eq), DCM, RT, 1 h, 56% yield over two steps; (g) **A-3** (1.6 eq), 0.5 M TMSOTf (0.1 eq), 4 Å mol. sieves, DCM, -65 to -40 °C, 30 min, 95% yield; (h) $Bu_3SnH$ (6 eq), AIBN (0.17 eq), benzene, reflux, 3 h, 76% yield; (i) 80% AcOH, 80 °C, 99% yield.

*O-(Methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-1-O-acetyl-4,6-O-benzylidene-2-deoxy-2-trichloroacetamido-α-D-galactopyranose* (**A-4**).[6] 100.0 g of CS-A polymer (Federal Laboratories, Corp., Alden, NY, USA) was dissolved in 1

L deionized $H_2O$, and the pH of the solution was lowered to 1.6 by addition of Amberlite IR-120 $H^+$ resin. The solution was filtered, and the resin was washed with deionized $H_2O$ (4x100 mL). The final volume was adjusted to 1.94 L and 55.6 mL conc. $H_2SO_4$ was added. The solution was boiled at 100 °C for 6 h. $Ba(OH)_2{\cdot}8H_2O$ was added while vigorously stirring until the pH reached 3.5. The solids were allowed to settle overnight, then the solution was filtered with Celite. The solid was washed with deionized $H_2O$ until the filtrate was clear. The solution was concentrated to 1 L then applied to a column of Amberlite IR-120 $H^+$ resin (1 L settled volume). The column was flushed with $H_2O$ (2 L), 3:1 AcOH/$H_2O$ (2 L), and 1 M HCl (6 L). The positively staining fractions via ninhydrin were concentrated and then coconcentrated with $H_2O$ (2x500 mL). The crude solid was dried overnight *in vacuo*, then dissolved in 1 L 0.02 M methanolic HCl, and stirred at 0 °C for 4 d. This solution was then concentrated and co-concentrated with EtOH (2x100 mL) to give the crude disaccharide (55.93 g) as the hydrochloride salt. A portion of this crude disaccharide (20.0 g, 49.3 mmol) was then dissolved in 220 mL pyridine and cooled to -30 °C. Trichloroacetyl chloride (62 mL, 0.555 mol, 11.3 eq) was added dropwise. Once the addition was complete, the reaction was stirred at 0 °C for 1 hour then cooled to -30 °C again. Deionized $H_2O$ (4.2 mL) was added very slowly while maintaining temperature. This was then diluted with $CH_2Cl_2$, washed with deionized water, brine then deionized water, dried over $MgSO_4$, and concentrated. The residue was redissolved in 150 mL 1:1:1

$CH_2Cl_2$/MeOH/pyridine and stirred for 4 hours at RT. The solution was concentrated and dried *in vacuo*. A silica plug (4:1 $CH_2Cl_2$/MeOH) was used to yield the crude trichloroacetamide (10.54 g), which was again used directly as well. The crude trichloroacetamide (10.0 g, 19.4 mmol) was dissolved in benzaldehyde (72 mL, 709 mmol, 36.5 eq) along with TFA (3.8 mL, 51.2 mmol, 2.64 eq). This mixture was for 24 hours at RT. Directly to the mixture, NaOAc (6.20 g, 75.6 mmol, 3.90 eq), pyridine (80 mL, 989 mmol, 51.0 eq), and acetic anhydride (50 mL, 529 mmol, 27.3 eq) were added sequentially. This was then stirred for 16 hours at RT. The solution was then added to ice-cold deionized $H_2O$ and stirred for an additional 2 h. This was then extracted with $CH_2Cl_2$ (2x150 mL). The organic layers were combined, washed with deionized $H_2O$, saturated $NaHCO_3$ (aq.), then deionized $H_2O$, dried over $MgSO_4$, and concentrated. Column chromatography (100% $CH_2Cl_2$ →15:1 $CH_2Cl_2$/acetone) afforded the pure alpha anomer (6.10 g, 11.5% yield from polymer) as a white powder. $R_f$ 0.44 (15:1 $CH_2Cl_2$/acetone); $^1$H NMR (500 MHz, $CDCl_3$) δ 7.55 – 7.49 (m, 2H, Ph*H*), 7.42 – 7.31 (m, 3H, Ph*H*), 6.77 (d, *J* = 7.7 Hz, 1H, GalN N*H*), 6.49 (d, *J* = 3.4 Hz, 1H, GalN H-1), 5.55 (s, 1H, PhC*H*), 5.30 – 5.19 (m, 2H, GlcA H-3, H-4), 5.10 (dd, *J* = 8.7, 7.8 Hz, 1H, GlcA H-2), 4.95 (d, *J* = 7.8 Hz, 1H, GlcA H-1), 4.65 (ddd, *J* = 11.1, 7.7, 3.4 Hz, 1H, GalN H-2), 4.51 (dd, *J* = 3.3, 1.2 Hz, 1H, GalN H-4), 4.47 (dd, *J* = 11.1, 3.3 Hz, 1H, GalN H-3), 4.30 (dd, *J* = 12.6, 1.6 Hz, 1H, GalN H-6a), 4.12 (d, *J* = 9.7 Hz, 1H, GlcA H-5), 4.08 (dd, *J* = 12.6, 1.8

Hz, 1H, GalN H-6b), 3.85 (m, 1H, GalN H-5), 3.74 (s, 3H, $CO_2CH_3$), 2.19 (s, 3H, $OCH_3$), 2.04 (m, 9H, $OCH_3$).

*O-(Methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-4,6-O-benzylidene-2-deoxy-2-trichloroacetamido-1-O-trichloroacetimidoyl-α-D-galactopyranose* (**A-5**). Disaccharide **A-4** (8.10 g, 10.51 mmol) was dissolved in 65 mL dry DMF. To this solution, hydrazine acetate (1.48 g, 16.07 mmol, 1.5 eq) was added. The reaction was stirred at RT for 2 hours then diluted with EtOAc and washed with deionized $H_2O$, brine, then deionized $H_2O$. The organic layer was dried over $MgSO_4$ and concentrated. The residue was redissolved in 65 mL dry $CH_2Cl_2$, and DBU (0.25 mL, 1.67 mmol, 0.16 eq) and trichloroacetonitrile (8.51 mL, 84.87 mmol, 8 eq) were added. Upon addition of DBU, the reaction turned bright yellow and slowly darkened to brown. After stirring for 1 hour at RT, the reaction was concentrated to afford an orange solid. This solid was passed through a silica plug (15:1 $CH_2Cl_2$/acetone). The crude product was then recrystallized with $Et_2O$ to afford the pure imidate (5.32 g, 56%) as white crystals. $R_f$ 0.78 (15:1 $CH_2Cl_2$/acetone); $^1$H NMR (500 MHz, $CDCl_3$) δ 8.76 (s, 1H, =N*H*), 7.57 – 7.50 (m, 2H, Ph*H*), 7.42 – 7.32 (m, 3H, Ph*H*), 6.79 (d, *J* = 8.1 Hz, 1H, -N*H*), 6.65 (d, *J* = 3.4 Hz, 1H, GalN H-1), 5.57 (s, 1H, C*H*Ph), 5.25 (t, *J* = 9.4 Hz, 1H, GlcA H-3), 5.20 (t, *J* = 8.8 Hz, 1H, GlcA H-4), 5.09 (dd, *J* = 8.6, 7.6 Hz, 1H, GlcA H-2), 4.95 (d, *J* = 7.6 Hz, 1H, GlcA H-1), 4.77 (ddd, *J* = 11.4, 8.1, 3.4 Hz, 1H, GalN H-2), 4.56 (d, *J* = 4.1 Hz, 1H, GalN H-4), 4.40 (dd, *J* = 11.0, 3.4 Hz, 1H, GalN H-3), 4.34 (dd, *J* = 12.6, 1.6 Hz, 1H, GalN H-6a), 4.17 – 4.05 (m,

2H, GalN H-6b, GlcA H-5), 3.93 (s, 1H, GalN H-5), 3.72 (s, 3H, -$CO_2CH_3$), 2.05 – 1.98 (m, 9H, -$OCH_3$); HRMS: calculated $[M+Na]^+$ 894.9802, determined: 894.9742.

*2-(2-((1R,2S,4R)-Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)ethoxy)ethyl O-(methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-4,6-O-benzylidene-2-deoxy-2-trichloroacetamido-β-D-galactopyranose* (**A-6**). Disaccharide **A-5** (1.84 g, 2.11 mmol, 1 eq) was dissolved in 30 mL dry $CH_2Cl_2$ with 4 Å molecular sieves under Ar, and the solution was cooled to –60 °C. To this was added the norbornenyl linker **3** (0.70 g, 3.30 mmol, 1.6 eq) dissolved in 7 mL dry $CH_2Cl_2$. The mixture was stirred at –60 °C for 1 h, then 440 μL 0.5 M TMSOTf in $CH_2Cl_2$ (0.1 eq) was injected by syringe. This was then stirred at –60 °C for 30 minutes then allowed to warm to –40 °C over 45 m. The reaction was subsequently quenched with $NEt_3$, filtered using Celite, and concentrated. Column chromatography (2:1 to 1:2 hexanes/EtOAc) yielded the final product as a white solid (1.82 g, 95%). $R_f$ 0.95 (1:2 hexanes/EtOAc); $^1$H NMR (500 MHz, $CDCl_3$) δ 7.56 – 7.50 (m, 2H, Ph*H*), 7.40 – 7.30 (m, 3H, Ph*H*), 7.08 (dd, *J* = 6.7, 1.4 Hz, 1H, N*H*), 6.07 (ddd, *J* = 22.4, 5.7, 2.9 Hz, 2H, C*H*=C*H*), 5.59 (s, 1H, C*H*Ph), 5.23 (dd, *J* = 9.8, 9.3 Hz, 1H, GlcA H-3), 5.16 (t, *J* = 9.0 Hz, 1H, GlcA H-4), 5.09 (d, *J* = 8.3 Hz, 1H, GalN H-1), 5.04 (dd, *J* = 8.8, 7.6 Hz, 1H, GlcA H-2), 4.91 (d, *J* = 7.5 Hz, 1H, GlcA H-1), 4.67 (ddd, *J* = 11.0, 3.5, 1.8 Hz, 1H, GalN H-3), 4.43 (d, *J* = 3.6 Hz, 1H, GalN H-4), 4.32 (dd, *J* = 12.3, 1.6 Hz, 1H, GalN H-6a), 4.09 (d, *J* = 10.9 Hz, 1H, GalN H-6b), 4.02 (d, *J* = 9.9 Hz, 1H, GlcA H-5), 4.02 – 3.96 (m, 1H, -OC*H*), 3.85 – 3.78

(m, 1H, GalN H-2), 3.79 – 3.70 (m, 1H, OC*H*), 3.72 (s, 3H, $CO_2CH_3$), 3.65 (dd, *J* = 5.5, 4.3 Hz, 2H, 2 OC*H*), 3.63 – 3.60 (m, 2H, 2 OC*H*), 3.59 – 3.54 (m, 2H, 2 OC*H*), 3.52 – 3.50 (m, 1H, GalN H-5), 3.50 – 3.48 (m, 1H, OC*H*), 3.34 (td, *J* = 9.3, 1.1 Hz, 1H, OC*H*), 2.79 (bs, 1H, NB-*H*), 2.73 (bs, 1H, NB-*H*), 2.06 – 1.96 (m, 9H, 3 C(O)C$H_3$), 1.68 (tt, *J* = 8.8, 4.7 Hz, 1H, NB-*H*), 1.33 – 1.18 (m, 3H, NB-*H*), 1.12 – 1.04 (m, 1H, NB-*H*); HRMS: calculated $[M-H]^-$ 920.2072, determined: 920.2090.

*2-(2-((1R,2S,4R)-Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)ethoxy)ethyl O-(methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-4,6-O-benzylidene-2-deoxy-2-acetamido-β-D-galactopyranose* (**A-7**). To a solution of disaccharide glycoside **A-6** (0.405 g, 0.439 mmol, 1 eq) in 13 mL dry benzene, $Bu_3SnH$ (0.72 mL, 2.86 mmol, 6 eq) and AIBN (12 mg, 0.073 mmol, 0.17 eq) were added. The mixture was stirred at RT for 1 hour with Ar bubbling through the solution. The reaction was then refluxed at 80 °C under Ar for 3 hours then quenched with $NEt_3$. Column chromatography (100% EtOAc to 100:1 to 20:1 EtOAc/MeOH) afforded the product as a white solid (0.310 g, 86%). $R_f$ 0.45 (9:1 EtOAc/MeOH); $^1$H NMR (500 MHz, $CDCl_3$) δ 7.57 – 7.52 (m, 2H, Ph*H*), 7.41 – 7.29 (m, 3H, Ph*H*), 6.14 – 6.03 (m, 2H, C*H*=C*H*), 5.98 (s, 1H, N*H*), 5.55 (s, 1H, C*H*Ph), 5.32 – 5.17 (m, 2H, GlcA H-3, GlcA H-4), 5.09 (dd, *J* = 8.2, 1.6 Hz, 1H, GalN H-1), 5.00 (t, *J* = 8.1 Hz, 1H, GlcA H-2), 4.92 (d, *J* = 7.6 Hz, 1H, GlcA H-1), 4.71 (dt, *J* = 11.2, 3.0 Hz, 1H, GalN H-3), 4.34 (d, *J* = 3.4 Hz, 1H, GalN H-4), 4.30 (dd, *J* = 12.4, 1.6 Hz, 1H, GalN H-6a), 4.06 (dd, *J* = 12.5, 1.8 Hz, 1H, GalN H-6b), 4.02 (d, *J* = 8.3 Hz, 1H,

GlcA H-5), 4.00 – 3.96 (m, 1H, OC*H*), 3.76 – 3.70 (m, 1H, OC*H*), 3.69 (s, 3H, $CO_2CH_3$), 3.67 – 3.61 (m, 4H, 4 OC*H*), 3.61 – 3.55 (m, 2H, 2 OC*H*), 3.54 – 3.46 (m, 3H, OC*H*, GalN H-2, GalN H-5), 3.36 (td, *J* = 9.2, 3.7 Hz, 1H, OC*H*), 2.79 (bs, 1H, NB-*H*), 2.73 (bs, 1H, NB-*H*), 2.05 – 1.93 (m, 12H, 4 C(O)C$H_3$), 1.72 – 1.65 (m, 1H, NB-*H*), 1.33 – 1.19 (m, 3H, NB-*H*), 1.09 (dtd, *J* = 11.7, 3.8, 2.3 Hz, 1H, NB-*H*); HRMS: calculated $[M-H]^-$ 820.3386, determined: 820.3384.

*2-(2-((1R,2S,4R)-Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)ethoxy)ethyl O-(methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-2-deoxy-2-acetamido-β-D-galactopyranose* (**A-8**). Acetamide **A-7** (0.310 g, 0.378 mmol) was suspended in 9 mL 80% AcOH. The suspension was stirred at 80 °C for 30 minutes, during which the compound dissolved. The reaction was concentrated, and column chromatography (9:1 EtOAc/MeOH) afforded the product as a white solid (0.273 g, 99%). $R_f$ 0.18 (9:1 EtOAc/MeOH); $^1$H NMR (500 MHz, $CDCl_3$) δ 6.69 (t, *J* = 7.0 Hz, 1H, N*H*), 6.15 – 6.04 (m, 2H, C*H*=C*H*), 5.26 (td, *J* = 9.4, 1.9 Hz, 1H, GlcA H-3), 5.16 (t, *J* = 9.6 Hz, 1H, GlcA H-4), 5.01 (ddd, *J* = 9.3, 7.9, 1.2 Hz, 1H, GlcA H-2), 4.78 (dd, *J* = 8.5, 2.9 Hz, 1H, GalN H-1), 4.74 (d, *J* = 8.0 Hz, 1H, GlcA H-1), 4.25 – 4.02 (m, 3H, GlcA H-5, GalN H-3, H-4), 3.98 – 3.84 (m, 2H, OC*H*, GalN H-2), 3.86 – 3.67 (m, 4H, OC*H*, $CO_2CH_3$), 3.70 – 3.44 (m, 8H, OC*H*, GalN H-5), 3.42 (bs, 1H, GalN O*H*), 3.43 – 3.29 (m, 1H, OC*H*), 3.28 (bs, 1H, GalN O*H*), 2.81 (bs, 1H, NB-*H*), 2.72 (bd, *J* = 6.8 Hz, 1H, NB-*H*), 2.09 – 1.93 (m, 12H, C(O)C$H_3$), 1.72 – 1.61 (m, 1H, NB-*H*), 1.37 – 1.15 (m, 3H, NB-*H*), 1.17 – 1.04 (m, 1H, NB-*H*); HRMS: calculated $[M-H]^-$ 730.2928, determined: 730.2945.

## A.5 Elaboration to CS Sulfation Epitope Monomers

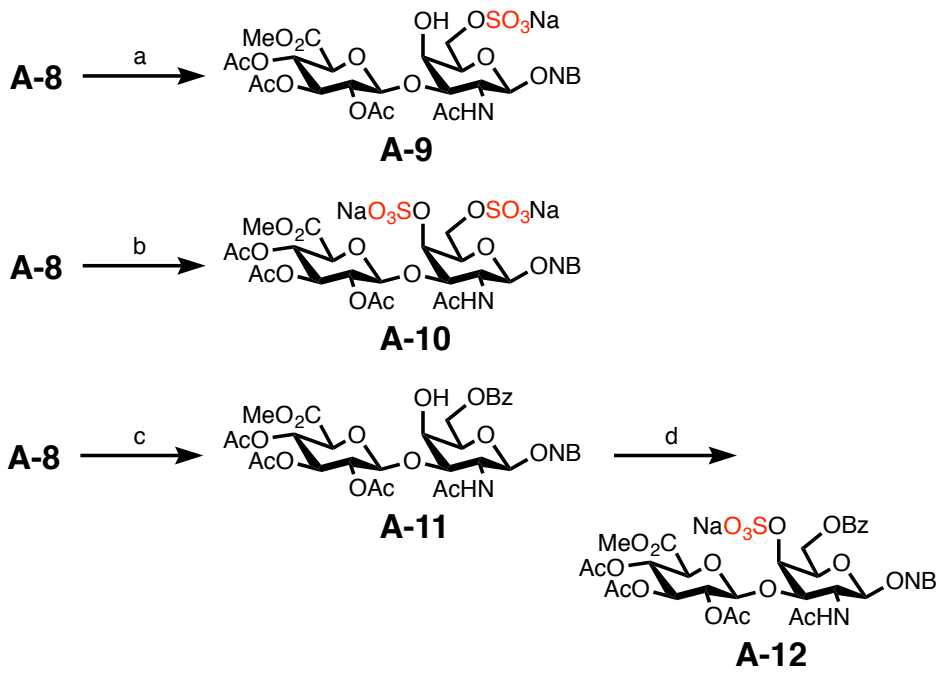



**Figure A-3. Synthesis of CS sulfation epitope monomers.**[5] (a) $SO_3$·TMA (2.7 eq), DMF, 50 °C, 40 min, 90% yield; (b) $SO_3$·TMA (25 eq), DMF, 50 °C, 1 d, 43% yield; (c) BzCN (2 eq), pyridine, RT, 16 h, 80% yield; (d) $SO_3$·TMA (6 eq), DMF, 50 °C, 1.5 d, 36% yield.

*CS-C monomer* (**A-9**). Diol **A-8** (100 mg, 0.137 mmol) was dissolved in 3.8 mL dry DMF under Ar. To this, sulfur trioxide trimethylamine complex (56 mg, 0.369 mmol, 2.7 eq) was added. The reaction was stirred at 50 °C for 40 m then cooled and added directly to an LH-20 column. After eluting the product from the LH-20 column, the crude mixture was subjected to cation exchange chromatography (Amberlite IR-120 $Na^+$ form) followed by silica column chromatography (10:2:1 EtOAc/MeOH/$H_2O$), which gave the product as a white solid (103 mg, 90%). $R_f$ 0.37 (10:2:1 EtOAc/MeOH/$H_2O$); $^1$H NMR (500

MHz, MeOD-$d_4$) δ 6.10 (ddd, *J* = 18.7, 6.3, 3.5 Hz, 2H, C*H*=C*H*), 5.32 (td, *J* = 9.5, 1.6 Hz, 1H, GlcA H-3), 5.11 (td, *J* = 9.8, 1.6 Hz, 1H, GlcA H-4), 4.99 (ddd, *J* = 9.6, 7.8, 1.6 Hz, 1H, GlcA H-2), 4.91 (dd, *J* = 7.9, 1.5 Hz, 1H, GlcA H-1), 4.48 (dd, *J* = 8.5, 1.6 Hz, 1H, GalN H-1), 4.30 (dt, *J* = 10.0, 1.7 Hz, 1H, GlcA H-5), 4.19 (d, *J* = 5.8 Hz, 2H, GalN H-6a, H-6b), 4.12 – 4.05 (m, 2H, GalN H-3, H-5), 4.01 – 3.91 (m, 1H, OC*H*), 3.87 – 3.76 (m, 2H, GalN H-2, H-4), 3.78 – 3.67 (m, 4H, OC*H*, $CO_2CH_3$), 3.70 – 3.51 (m, 7H, OC*H*), 3.43 (td, *J* = 9.3, 1.6 Hz, 1H, OC*H*), 2.81 (bs, 1H, NB-*H*), 2.74 (bs, 1H, NB-*H*), 2.14 – 1.93 (m, 12H, C(O)$CH_3$), 1.74 – 1.65 (m, 1H, NB-*H*), 1.39 – 1.21 (m, 3H, NB-*H*), 1.21 – 1.13 (m, 1H, NB-*H*); HRMS: calculated $[M-H]^-$ 810.2496, determined: 810.2511.

*CS-E monomer* (**A-10**). Diol **A-8** (250 mg, 0.343 mmol) was dissolved in 12 mL dry DMF under Ar. To this, sulfur trioxide trimethylamine complex (1.19 g, 8.55 mmol, 25 eq) was added. The reaction was stirred at 50 °C for 24 hours then cooled and added directly to an LH-20 column. After eluting the product from the column, the crude mixture was subjected to cation exchange chromatography (Amberlite IR-120 $Na^+$ form) followed by silica column chromatography (5:2:1 EtOAc/MeOH/$H_2O$), which yielded the product as a white solid (139 mg, 43%). $R_f$ 0.18 (10:2:1 EtOAc/MeOH/$H_2O$); $^1$H NMR (500 MHz, MeOD-$d_4$) δ 6.09 (ddd, *J* = 19.9, 6.0, 3.2 Hz, 2H, C*H*=C*H*), 5.31 (t, *J* = 9.4 Hz, 1H, GlcA H-3), 5.20 (t, *J* = 9.9 Hz, 1H, GlcA H-4), 5.07 (t, *J* = 8.7 Hz, 1H, GlcA H-2), 4.95 (d, *J* = 8.1 Hz, 1H, GlcA H-1), 4.86 (hidden under HDO peak, 1H, GalN H-4), 4.51 (d, *J* = 8.3 Hz, 1H, GalN H-1), 4.38 (dd, *J* = 11.8, 3.4 Hz, 1H,

GalN H-6a), 4.31 – 4.24 (m, 2H, GalN H-6b, GlcA H-5), 4.06 (t, *J* = 9.6 Hz, 1H, GalN H-2), 4.03 – 3.88 (m, 3H, OC*H*, GalN H-3, H-5), 3.75 (s, 3H, $CO_2CH_3$), 3.72 – 3.51 (m, 8H, OC*H*), 3.42 (t, *J* = 9.2 Hz, 1H, OC*H*), 2.79 (bs, 1H, NB-*H*), 2.73 (bs, 1H, NB-*H*), 2.11 – 1.95 (m, 12H, $C(O)CH_3$), 1.74 – 1.63 (m, 1H, NB-*H*), 1.38 – 1.20 (m, 3H, NB-*H*), 1.15 (dt, *J* = 11.7, 3.9 Hz, 1H, NB-*H*); HRMS: calculated $[M\text{-}Na]^-$ 912.1883; determined: 912.1905.

*2-(2-((1R,2S,4R)-Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)ethoxy)ethyl O-(methyl 2,3,4-tri-O-acetyl-β-D-glucopyranosyluronate)-(1→3)-6-O-benzoyl-2-deoxy-2-acetamido-β-D-galactopyranose* (**A-11**). Diol **A-8** (200 mg, 0.273 mmol) was dissolved in 5.0 mL dry pyridine along with benzoyl cyanide (72 mg, 0.547 mmol, 2 eq) under Ar. The reaction was stirred at RT for 16 h. MeOH (2 mL) was added to quench the reaction, after which the solution was concentrated. Column chromatography (100% EtOAc) afforded the product as a white solid (0.251 g, 80%). $R_f$ 0.16 (100% EtOAc); $^1H$ NMR (500 MHz, $CDCl_3$) δ 8.07 – 8.01 (m, 2H, Ph*H*), 7.60 – 7.53 (m, 1H, Ph*H*), 7.48 – 7.41 (m, 2H, Ph*H*), 6.14 – 6.06 (m, 2H, C*H*=C*H*), 6.00 – 5.91 (m, 1H, N*H*), 5.28 – 5.15 (m, 2H, GlcA H-3, H-4), 5.02 (dd, *J* = 9.0, 7.7 Hz, 1H, GlcA H-2), 4.93 (dd, *J* = 8.4, 4.1 Hz, 1H, GalN H-1), 4.74 (d, *J* = 7.7 Hz, 1H, GlcA H-1), 4.60 (dd, *J* = 11.4, 5.6 Hz, 1H, GalN H-6a), 4.56 (dd, *J* = 11.4, 6.9 Hz, 1H, GalN H-6b), 4.49 (td, *J* = 11.5, 3.2 Hz, 1H, GalN H-3), 4.15 – 4.11 (m, 1H, GalN H-4), 4.03 (dd, *J* = 9.4, 7.0 Hz, 1H, GlcA H-5), 3.94 (dt, *J* = 11.6, 4.1 Hz, 1H, OC*H*), 3.88 (t, *J* = 6.3 Hz, 1H, GalN H-5), 3.74 (dt, *J* = 11.2, 5.4 Hz, 1H, OC*H*), 3.66 (s, 3H, $CO_2CH_3$), 3.66 – 3.46 (m, 8H, GalN H-2,

OC*H*), 3.37 (dt, *J* = 16.0, 9.2 Hz, 1H, OC*H*), 2.81 (bs, 1H, NB-*H*), 2.73 (bd, *J* = 7.4 Hz, 1H, GalN 4-O*H*), 2.67 (bs, 1H, NB-*H*), 2.09 – 1.95 (m, 12H, C(O)C$H_3$), 1.74 – 1.64 (m, 1H, NB-*H*), 1.42 – 1.20 (m, 3H, NB-*H*), 1.11 (ddt, *J* = 11.3, 7.4, 3.9 Hz, 1H, NB-*H*); HRMS: calculated $[M+Cl]^-$ 870.2951, determined: 870.2966.

*CS-A monomer* (**A-12**). Disaccharide **A-11** (250 mg, 0.299 mmol) was dissolved in 15 mL dry DMF under Ar. To this, sulfur trioxide trimethylamine complex (250 mg, 1.79 mmol, 6 eq) was added. The reaction was stirred at 50 °C for 36 hours then cooled and added directly to an LH-20 column. After eluting the product from the LH-20 column, the crude mixture was subjected to cation exchange chromatography (Amberlite IR-120 $Na^+$ form) followed by silica column chromatography (10:2:1 EtOAc/MeOH/$H_2O$), which gave the product as a white solid (101 mg, 36%). $R_f$ 0.55 (10:2:1 EtOAc/MeOH/$H_2O$); $^1$H NMR (500 MHz, MeOD-$d_4$) δ 8.06 (dt, *J* = 9.0, 1.7 Hz, 2H, Ph*H*), 7.63 – 7.55 (m, 1H, Ph*H*), 7.47 (td, *J* = 7.6, 1.5 Hz, 2H, Ph*H*), 6.12 – 6.02 (m, 2H, C*H*=C*H*), 5.31 (td, *J* = 9.4, 2.4 Hz, 1H, GlcA H-3), 5.23 (td, *J* = 9.6, 2.3 Hz, 1H, GlcA H-4), 5.07 (ddd, *J* = 10.5, 8.4, 2.4 Hz, 1H, GlcA H-2), 5.02 – 4.90 (m, 2H, GlcA H-1, GalN H-3), 4.65 (dd, *J* = 6.5, 2.4 Hz, 2H, GalN H-6a, H-6b), 4.61 (d, *J* = 8.1 Hz, 1H, GalN H-1), 4.23 (dq, *J* = 9.9, 1.9 Hz, 1H, GlcA H-5), 4.10 (d, *J* = 11.0 Hz, 1H, GalN H-4), 4.04 – 3.96 (m, 2H, GalN H-2, H-5), 3.87 (dt, *J* = 11.4, 4.3 Hz, 1H, OC*H*), 3.77 – 3.66 (m, 4H, OC*H*, $CO_2CH_3$), 3.66 – 3.46 (m, 7H, OC*H*), 3.37 (ddt, *J* = 11.6, 9.4, 1.7 Hz, 1H, OC*H*), 2.78 (bs, 1H, NB-*H*), 2.71 (bs, 1H, NB-*H*), 2.06 – 1.93 (m, 12H, C(O)C$H_3$), 1.70 – 1.61 (m, 1H, NB-*H*), 1.36 – 1.26 (m, 2H, NB-*H*), 1.22 (ddd, *J* = 11.1, 8.4, 2.4

Hz, 1H, NB-*H*), 1.13 (dt, *J* = 11.7, 3.9 Hz, 1H, NB-*H*); HRMS: calculated $[M\text{-}Na]^-$ 914.2758, determined: 914.2770.

**A.6 CS Polymerization**

*General polymerization procedure.*[1, 2] A vial with a rubber septum was charged with disaccharide monomers (60.0 mg) under Ar. To this dry, degassed MeOH (0.25 mL) and DCE (1.25 mL) were added. The solution was heated to 55 °C, and then an appropriate amount (mol % to determine polymer size) of the fast-activating Grubbs 3$^{rd}$ generation catalyst (2.0 mg/mL in DCE) was rapidly added by syringe injection. This was stirred for 1 hours at 55 °C. The reaction was quenched with ethyl vinyl ether (0.30 mL) with stirring for 30 minutes at 55 °C and concentrated. The polymer residue was dissolved in a minimum of 1:1 DCM/MeOH and precipitated by adding dropwise to a vortexing solution of 1:1 hexanes/diethyl ether (50 mL). The precipitate was centrifuged down and the solvent decanted. This was repeated three times to give a white pellet.

*Deprotection steps.* The precipitated pellet was redissolved in 2:1 THF/$H_2O$ (3 mL). The solution was cooled to 0 °C, and LiOH (1 M, 0.50 mL) and $H_2O_2$ (30%, 0.25 mL) were added. The reaction was stirred at 0 °C for 1 hour then allowed to warm to RT while stirring for another 12 h. To this, NaOH (4 M, 0.8 mL) and MeOH (2.0 mL) were added directly. The reaction was stirred at RT for 24 hours then neutralized with Amberlite IR-120 $H^+$. The solution was filtered through a 0.2 μm syringe filter, and organic solvents were removed. The remaining $H_2O$ was removed by lyophilization. Gel filtration chromatography (G-25 Fine, dd$H_2O$) and subsequent lyophilization affording the deprotected polymers as a white solid. For long-term storage, polymers were subjected to

cation exchange with tributylammonium containing Amberlite IR-120. Amberlite IR-120 $H^+$ resin was loaded into a column and then washed with ten column volumes of 1 M TBACl, and excess reagent as removed by washing with ten column volumes of $ddH_2O$. Next, the CS polymers were dissolved at 1 mg/mL in $ddH_2O$ and flowed over the column by gravity. The flowthrough was then combined, lyophilized, and subjected to gel filtration chromatography as previously described. Polymers were stored as the lyophilized powder at -80 °C.

## A.7 References


1. M. Rawat, C. I. Gama, J. B. Matson, L. C. Hsieh-Wilson. Neuroactive chondroitin sulfate glycomimetics. *J Am Chem Soc*. **2008**, *130*: 2959-2961.
2. S.-G. Lee *et al.* End-functionalized glycopolymers as mimetics of chondroitin sulfate proteoglycans. *Chem Sci*. **2010**, *1*: 322-325.
3. G. J. Sheng, Y. I. Oh, S.-K. Chang, L. C. Hsieh-Wilson. Tunable heparan sulfate mimetics for modulating chemokine activity. *J Am Chem Soc*. **2013**, *135*: 10898-10901.
4. Y. I. Oh, G. J. Sheng, S.-K. Chang, L. C. Hsieh-Wilson. Tailored glycopolymers as anticoagulant heparin mimetics. *Angew Chem Int Ed*. **2013**, *52*: 11796-11799.
5. C. I. Gama *et al.* Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol*. **2006**, *2*: 467-473.
6. C. Lopin, J.-C. Jacquinet. From polymer to size-defined oligomers: an expeditious route for the preparation of chondroitin oligosaccharides. *Angew Chem Int Ed*. **2006**, *45*: 2574-2578.