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

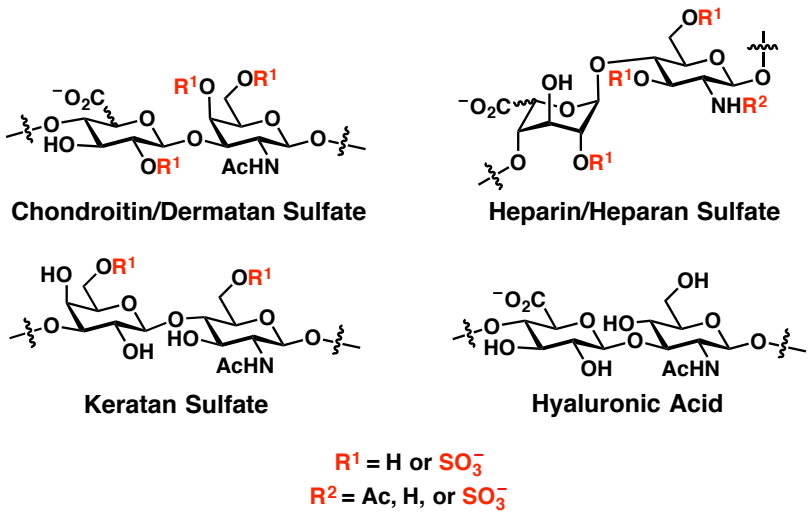

Figure 1.1 Structures of GAG family members. GAGs are made of repeating disaccharide units. Three of the four families can be sulfated at different hydroxyl groups along the polysaccharide.

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.