REVIEW

Check for updates

Design and synthesis of metabolic chemical reporters for the visualization and identification of glycoproteins

Nichole J. Pedowitz[a] and Matthew R. Pratt [iD] *[ab]

Glycosylation events play an invaluable role in regulating cellular processes including enzymatic activity, immune recognition, protein stability, and cell–cell interactions. However, researchers have yet to realize the full range of glycan mediated biological functions due to a lack of appropriate chemical tools. Fortunately, the past 25 years has seen the emergence of modified sugar analogs, termed metabolic chemical reporters (MCRs), which are metabolized by endogenous enzymes to label complex glycan structures. Here, we review the major reporters for each class of glycosylation and highlight recent applications that have made a tremendous impact on the field of glycobiology.

1. Introduction

Glycosylation is a post-translational modification (PTM) referring to the addition of one or more carbohydrate molecules to specific amino acid residues on proteins. Glycosylated proteins, termed glycoproteins, are present in all cells. The majority of glycoproteins are found on the cell surface where their oligosaccharide chain extends out into the extracellular matrix (ECM). Glycans are involved in many cellular processes including immune recognition, cellular trafficking, protein stability, and development[1] and are associated with disease symptoms including inflammation, pathogen invasion, and malignant transformation.[2]

Our overall understanding of the structure–function relationship of glycosylation events has been hindered by a lack of practical chemical tools to study these processes. Unlike other biopolymers, oligosaccharides do not have a homogeneous arrangement of building blocks. Even within a single glycan subtype there is a huge amount of heterogeneity.[1] Consequently, traditional biological approaches such as antibodies or affinity tags are not suitable for the study of glycans as they rely on predictable recognition epitopes not always present in complex oligosaccharide architectures. There has been some success with the development of binding lectins that recognize specific carbohydrate motifs.[3] However, these can suffer from low binding affinity and a lack of selectivity.

Fortunately, a solution has emerged in the advancement of engineered sugar analogs termed metabolic chemical reporters (MCRs). MCRs exploit a cell's endogenous carbohydrate metabolism to convert unnatural molecules into glycosyltransferase substrates which can ultimately be added on to proteins. In this review we detail the origins of MCRs and outline the contribution bioorthogonal labeling has made in advancing the field (Sections 2 and 3). We go on to describe the first generation of MCRs, whose main goal is to deconvolute the function of the various subtypes and building blocks of glycosylation (Section 4). Knowledge gained through these studies has facilitated the use of MCRs in more complicated biological contexts and allowed us to answer questions previously beyond reach (Section 5). Finally, we conclude with a summary of the limitations of MCRs and opportunities for carbohydrate chemical biology (Sections 6 and 7).

## 2. Origin of carbohydrate MCRs

All complex glycans are synthesized from common building blocks (Fig. 1A).[3] The *de novo* synthesis of these building blocks relies on common metabolites such as glucose and glutamine shuttled from tightly regulated pathways. Monosaccharides can also be synthesized *via* endogenous salvage pathways where enzymes scavenge the environment for free monosaccharides and recycle these building blocks back into high energy sugar-donor substrates of glycosyltransferases (Fig. 2).[4] Salvage pathways are uniquely adept for metabolic engineering. By starting from the monosaccharide building blocks instead of basic metabolites, there is more control over the metabolic fate of introduced biomolecules.

[a] *Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA*

[b] *Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA. E-mail: matthew.pratt@usc.edu*
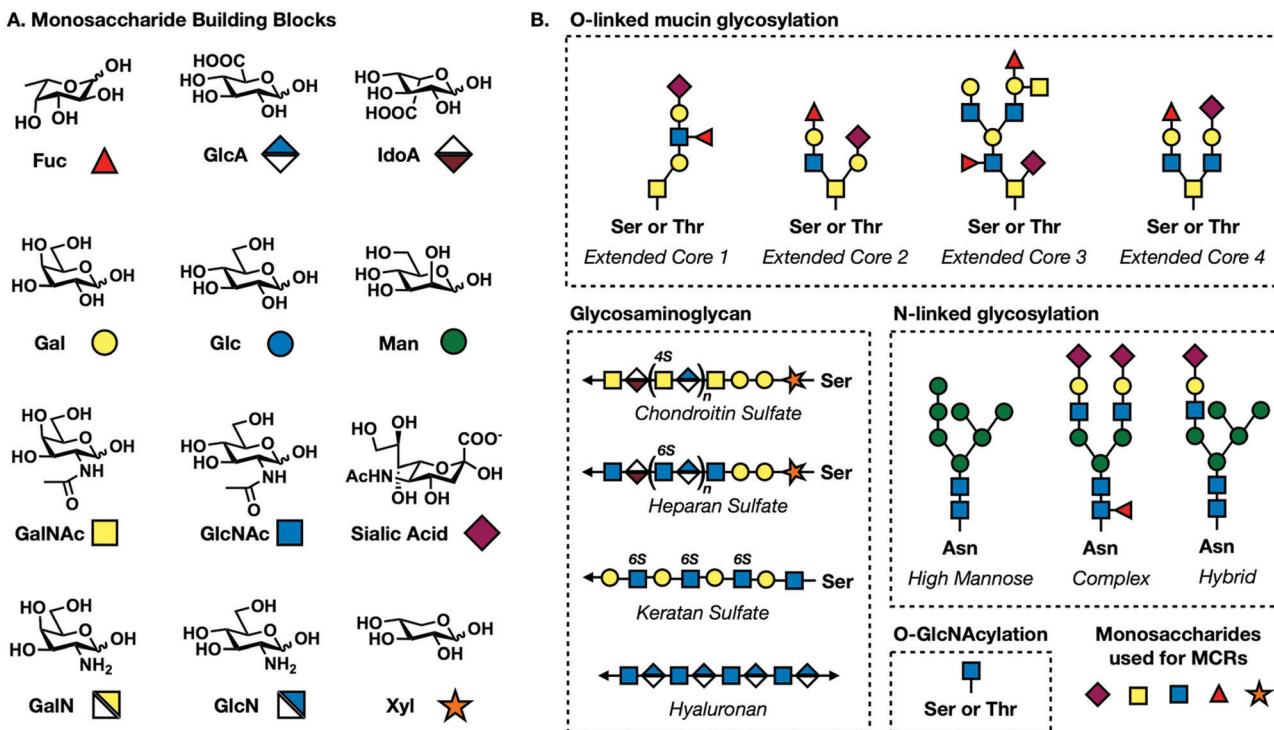
Fig. 1 (A) Common mammalian monosaccharide building blocks and their geometric codes. (B) Core structures for *O*-linked mucin, *N*-linked, GAG, and *O*-GlcNAc glycosylation.

Reutter and coworkers recognized the potential of sugar salvage pathways.[5–7] In pioneering work, Keppler *et al.* synthesized a series of *N*-acetylmannosamine (ManNAc) derivatives bearing one, two, or three additional carbons on the *N*-acetyl position.[6] These analogs were successfully transformed into the corresponding sialic acid derivatives and installed on to cell surface glycoproteins. This and follow up studies established the relaxed substrate specificity of glycosyltransferases.

Soon after, the Bertozzi group pushed harder on these salvage pathways to see if their permissive nature extended to small chemically reactive groups. They recognized that appending functional groups to glycoproteins could facilitate covalent elaboration to allow for subsequent biochemical analysis. This approach relies on a two-step process (Fig. 3). First, endogenous enzymes have to incorporate unnatural sugar molecules bearing biologically inert functional groups onto protein substrates. Next, these functional groups must be paired with selective and reactive partners to form covalent adducts.

Mahal *et al.* demonstrated the first successful semi-bioorthogonal reaction using the ketone-bearing monosaccharide *N*-levulinoylmannosamine (ManLev).[8] Though ketones are present in biological systems, they are notably absent from the cell surface where the metabolized product is deposited. Characterization of HeLa cells treated with ManLev demonstrated that it is metabolically converted into *N*-levulinoyl sialic acid (SiaLev) and incorporated onto glycoproteins. Cell-surface ketones readily react with biotin-hydrazide which can be visualized using flow cytometry. This reaction is considered only semi-bioorthogonal because there are endogenous intracellular

keto-metabolites, limiting the broad use of ketones as functional handles. Nevertheless, this study was a pivotal moment in the history of bioorthogonal labeling and chemical biology as a whole, inspiring the development of countless bioorthogonal transformations.

In the 20 years since, a large number of bioorthogonal reactions have been developed. Reactions most commonly employed for the visualization of glycans are described in the next section, but this is by no means an exhaustive list. For a more in-depth description of the available range of bioorthogonal reactions, we direct readers to other comprehensive reviews.[9–11]

3. Bioorthogonal labeling

Bioorthogonal reactions must meet the following criteria: (1) the complimentary functional groups cannot be present in living systems, (2) these reactive groups must readily react but (3) are otherwise chemically inert, and (4) the reaction must reliably occur in aqueous environments at (5) an appreciable rate.[11] The majority of bioorthogonal functional groups are small enough to be tolerated by carbohydrate salvage pathways, allowing chemists to outfit sugar molecules with their choice of reactive handle.

The first truly bioorthogonal reaction used in glycobiology was the Staudinger ligation.[12] Created by Saxon and Bertozzi, the Staudinger ligation generates an amide bond between an azide and ester-functionalized triarylphosphine (Fig. 4A).
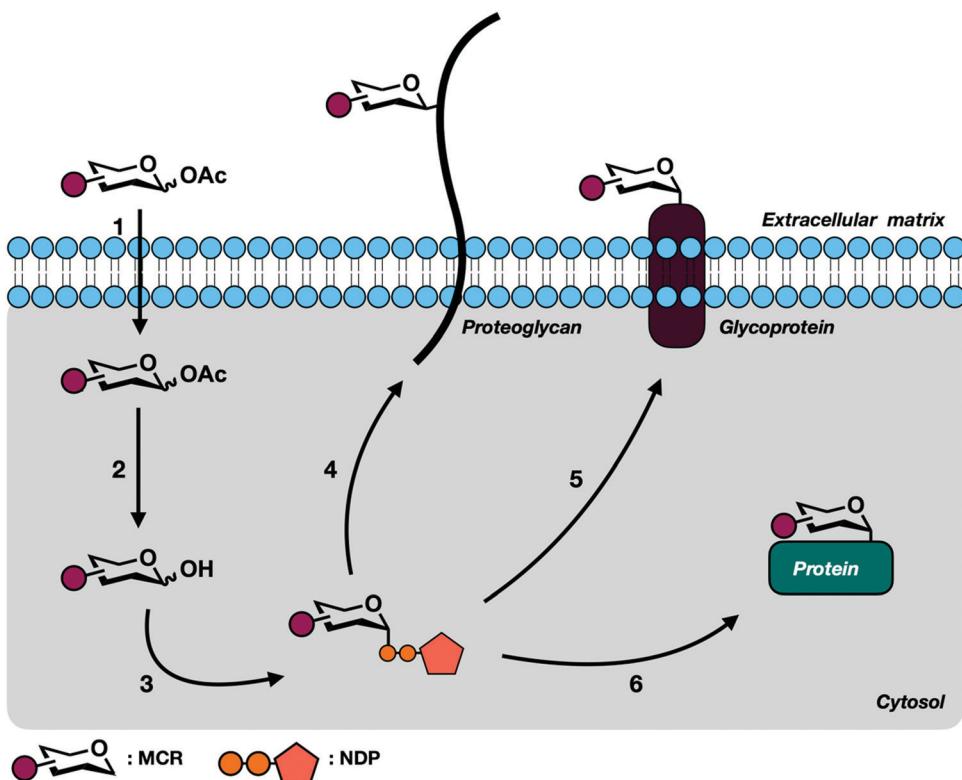
Fig. 2 MCRs take advantage of the endogenous sugar salvage pathways that exist for most monosaccharides. (1) Per-acetylated MCRs passively diffuse across the cell membrane. (2) Promiscuous deacetylase enzymes remove acyl groups revealing hydroxides. (3) MCRs are metabolically transformed into high energy sugar donors *via* covalent bonds to NDPs. These sugar donors can then be added on to protein substrates to generate proteoglycans (4), *O*- or *N*-linked glycoproteins (5), or intracellular *O*-GlcNAc modified proteins (6).
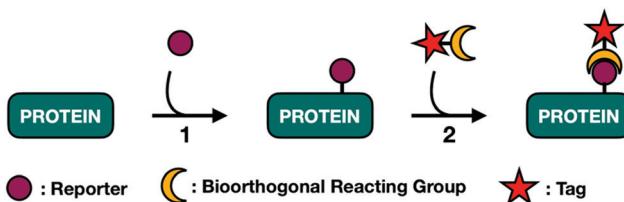

Fig. 3 Bioorthogonal labeling of glycoproteins is a two-step process. (1) Proteins are labeled with metabolically transformed sugar analogs. (2) Probes undergo orthogonal reactions with functionalized tags allowing for downstream processing.

These two reactive groups reliably form stable conjugates in cell lysate and live cells. An azide is small enough to be tolerated by the majority of carbohydrate salvage pathways leading to the development of diverse azide-functionalized unnatural sugars. Later work by Prescher *et al.* demonstrated that the Staudinger ligation proceeds in live mice.[13,14] Mice treated with the sialic acid precursor, per-acetylated *N*-azidoacetylmannosamine (Ac$_4$ManNAz), showed incorporation of the corresponding azido-sialic acid residues on their cell surfaces. Subsequent treatment with a phosphine-FLAG tag for western blot visualization demonstrated cell surface labeling in a live animal for the first time. Despite revolutionizing metabolic engineering, the Staudinger ligation suffers from slow reaction rates and

thus require excess amounts of triarylphosphine to go to completion in reasonable lengths of time.[11]

An alternative bioorthogonal reaction is copper(I)-catalyzed azide–alkyne cycloaddition (CuAAC) or click-chemistry (Fig. 4B).[15,16] As its name suggest, CuAAC refers to the reaction between an azide and terminal alkyne to generate stable triazole adducts. This reaction is highly selective, fast, and can proceed in aqueous environments lending to its emergence as a popular choice for *in vitro* visualization of cells and various enrichment workflows. Both azides and alkynes are small enough to be metabolized through sugar salvage pathways granting carbohydrate chemists some choice in how they functionalize their probes. However, in order to proceed at a reasonable rate this reaction requires a toxic Cu(I) catalyst impeding *in vivo* applications. Copper's toxicity can be mitigated through the tandem use of copper chelating ligands. A range of such ligands exist including TBTA, THPTA, BTTAA, and BTTES. Of these, BTTES and BTTAA are considered superior due to their increased reaction kinetics.[11] Continued development of chelating ligands will further expand the utility of this chemistry in living systems.

Amidst the search for a copper-free substitute, Bertozzi and colleagues revisited a reaction between an azide and cyclooctyne.[17,18] The reaction between azides and cyclooctynes, termed strain-promoted azide–alkyne cycloaddition (SPAAC), is driven by ring-strain (Fig. 4C). The cyclooctyne scaffold (OCT)

A. Staudinger Ligation

- ↑ Robust cell surface labeling
- ↑ Compatible *in vivo*
- ↓ Slow reaction rates
- ↓ Rapidly oxidize

Saxon
2000

B. Copper-catalyzed azide-alkyne cycloaddition (CuAAC)

- ↑ Smallest bioorthogonal handles
- ↑ Increased reaction rates
- ↓ Requires toxic Cu(I) limiting *in vivo* applications

Tornoe
2002

C. Strain-promoted azide-alkyne cycloaddition (SPAAC)

- ↑ No metal catalyst requirement
- ↑ Cell permeable
- ↓ Slow reaction rates
- ↓ Poor solubility in some cases

Agard
2004

D. Inverse Electron-Demand Diels-Alder (IED-DA)

- ↑ Fastest reaction rates
- ↑ Cell permeable
- ↓ Functional groups are too large for sugar salvage pathways

Yang
2012
Patterson
2012

Fig. 4 Bioorthogonal reactions commonly used for labeling glycans.

has a ''bent'' conformation along the triple bond.[19] This bond deformation generates strain that releases enough free energy upon 1,3-dipolar cycloaddition to proceed without a metal catalyst in aqueous environments. However, the first iteration of SPAAC suffered from the same slow rates as the Staudinger ligation. Efforts to improve the reaction rate led to the development of various functionalized cyclooctyne molecules including difluoronated cyclooctyne (DIFO) and dibenzocyclooctyne (DIBO) derivatives.[10] These improved cyclooctynes lead to $1000\times$ faster reaction kinetics. Unfortunately, some cyclooctynes suffer from poor solubility in aqueous environments limiting their use in biological settings.

The final major class of bioorthogonal reactions that shows great promise for *in vivo* applications are inverse electron-demand Diels–Alder (IED-DA) ligations (Fig. 4D).[20] IED-DA reactions occur between strained *trans*-cyclooctenes (TCOs) and tetrazines. In biological settings, this reaction has the fastest kinetics of any of the previously described transformations.[21,22] Early investigations demonstrated successful employment of this reaction in live mice. Sadly, this method relies on bulky reagents not tolerated by carbohydrate salvage enzymes. Alternative alkenes including methylcyclopropene and straight-chain alkenes are being investigated.[23–28] Methylcyclopropenes maintain enough ring strain to readily react with tetrazines in aqueous environments. Monosaccharides modified with straight-chain alkenes have shown promise. For example, mannosamine functionalized with a butenoyl group (Ac$_4$ManNBtl) successfully labeled HEK293T cells.[28] However, in both cases these reactions are significantly slower and less efficient than those with TCOs.

In addition to the above mentioned reactions, other bioorthogonal groups have potential for widespread use in monosaccharide MCRs. One such example from the Prescher group are cyclopropenones.[29] Cyclopropenones are small in size and react readily with triaryl phosphines making them attractive motifs for bioorthogonal chemistries. First iterations of cyclopropenones suffered from instability in aqueous environments limiting their use in biological systems. Row *et al.* stabilized these molecules by derivatizing them with electron withdrawing groups.[29] Though there are no current example of monosaccharide analogs outfitted with cyclopropenones, they serve as an attractive alternative to established reactions.

4. First generation MCRs

Early advances in monosaccharide reporters aimed to develop probes selective enough to target individual subtypes of glycosylation. This goal is complicated by the fact that the major forms of mammalian glycosylation all use several of the same monosaccharide building blocks (Fig. 1A).[1] The following section provides an introduction to each major type of mammalian glycosylation followed by a description of the MCRs that have been developed to target specific glycans or monosaccharides and probe their functions.

4.1. Core *N*-linked glycosylation

N-Linked glycans or simply *N*-glycans are branched oligosaccharides covalently bound to asparagine residues on cell

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | **309**

surface proteins *via* a GlcNAcβ1-Asn linkage.[30,31] The process of *N*-linked glycosylation begins when a GlcNAc residue is linked to a lipid anchoring molecule called a dolichol. Additional sugar molecules are added in a stepwise manner by various glycosyltransferases generating a common core found in all mammalian *N*-glycans consisting of two GlcNAc and three mannose residues (Fig. 1B). This pentasaccharide core is then further elaborated before enzymatic transfer of the corresponding oligosaccharide to proteins being synthesized into the Endoplasmic Reticulum. During protein folding and subsequent movement through the secretory system the glycan can be trimmed and further elaborated resulting in three extended core structures: high mannose consisting of the addition of four mannose molecules, complex consisting of the addition of GlcNAc molecules, and hybrid which is a combination of high mannose and complex (Fig. 1B). *N*-Linked glycans can be transferred onto proteins with the basic amino acid motif Asn-X-Ser/Thr with X being any amino acid except proline.

N-Linked glycosylation occurs co-translationally while a nascent peptide is being synthesized in the ER.[30] The extended oligosaccharide structures are recognized by lectin chaperones that aid in proper protein folding. Proteins that are not folded properly undergo specific *N*-glycan trimming, generating truncated oligosaccharides which mark the protein to be retro-translocated into the cytoplasm and degraded. Protein folding and other important functions of *N*-linked glycosylation have been shown to be critical for normal cell functioning. Misregulation of this form of glycosylation has been implicated in metabolic disorders and some cancers.

The design and synthesis of selective MCRs for *N*-linked glycans has proven to be challenging due to the overlap of common monosaccharide building blocks and inefficient turnover by salvage pathway enzymes. One successful example comes from the Bertozzi group who incorporated GlcNAlk and GlcNAz onto *N*-linked glycans in yeast.[32] To do so, they engineered a strain of *Saccharomyces cerevisiae* to be dependent on exogenous GlcNAc. Their reliance on an engineered strain of yeast suggests that salvage pathways that feed to *N*-linked glycan synthesis compete ineffectively with endogenous UDP-GlcNAc. However, the results indicate that GlcNAc-based MCRs could be successfully incorporated on to *N*-linked glycans in mammalian cells. Unfortunately, both of these reporters have been shown to be incorporated as intracellular *O*-GlcNAc modifications. To date, there are no reporters that can selectively label *N*-linked glycans' core pentasaccharide structure.

4.2. Mucin *O*-linked glycosylation

Mucin *O*-linked glycosylation is the most abundant form of glycosylation.[1,33] It is characterized by the addition of GalNAc to Ser and Thr residues on cell surface proteins that can then be further elaborated. GalNAc is metabolically transformed into the high energy sugar-donor UDP-GalNAc. It is then recognized by a family of enzymes known as polypeptide *N*-acetyl-α-galactosaminyltransferases (ppGalNAcTs)[34] and added to proteins in the Golgi apparatus. Additional sugar molecules can continue to be added *via* C3 or C6 linkages to the core GalNAc residue generating one of six core structures, four of which are expressed broadly in mammalian cells (Fig. 1B).[35] These can be further elaborated into complex and heterogenous linear or branched oligosaccharide trees. Mucin *O*-linked glycans serve a variety of functions including promoting or repelling cell-surface interactions, protecting against infection, maintaining cellular homeostasis, and serving as a barrier for cells from physical or chemical damage.[36]

Mucin MCRs have primarily been designed to target the initial GalNAc residue present in all mucin *O*-linked glycans. The first example came from Hang *et al.* who synthesized a GalNAc analog functionalized with an azide tag on the *N*-acetyl position, *N*-azidoacetylgalactosamine (GalNAz, Fig. 5) and demonstrated that it can label cells.[37] CHO cells were treated
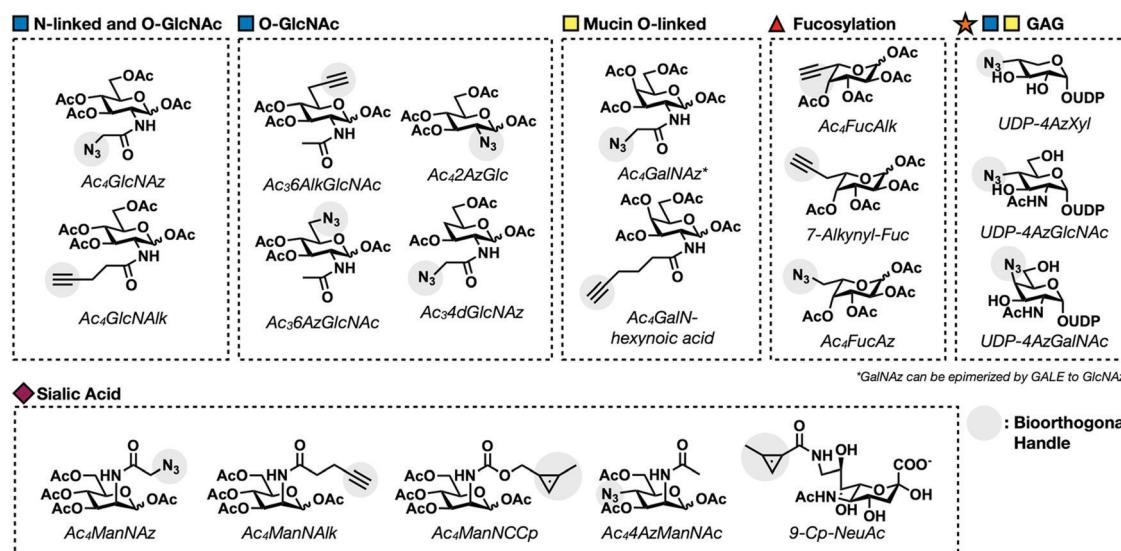
Fig. 5 Examples of first generation metabolic chemical reporters for different glycosylation events.

310 | *RSC Chem. Biol.*, 2021, **2**, 306–321

© 2021 The Author(s). Published by the Royal Society of Chemistry

with Ac$_4$GalNAz and reacted with phosphine-FLAG. Following the reaction, cells were labeled with FITC-conjugated anti-FLAG and analyzed *via* flow cytometry. GalNAz fluorescence was 30-fold higher than background fluorescence demonstrating efficient incorporation into cell surface glycoproteins. Two other reporters, Ac$_4$2AzGal and Ac$_3$6AzGalNAc, showed no significant fluorescence labeling over background suggesting that azide functionalities at these positions are not tolerated by the GalNAc salvage pathway. GalNAz labeling was further characterized through competition experiments. Co-treatment with 5 mM GalNAc completely removed cell surface labeling demonstrating that GalNAz was being metabolically incorporated through the GalNAc salvage pathway. Competition with GlcNAc resulted in partial inhibition of labeling consistent with UDP-GlcNAc's potential to be epimerized to UDP-GalNAc by the enzyme UDP-glucose 4-epimerase (GALE).

Zaro *et al.* continued characterizing GalNAz by examining its incorporation into known glycoproteins, to determine that GalNAz is preferentially incorporated into mucin *O*-linked glycans over *N*-linked glycans.[38] GlyCAM-IgG, a glycoprotein with both *N*-linked and *O*-linked glycosylation, was expressed in the presence of Ac$_4$GalNAz. The IgG domain was isolated with Sepharose beads and the enriched protein was subjected to CuAAC with an alkyne-fluorophore tag. In-gel fluorescence scanning showed robust labeling by GalNAz. Subsequent treatment with PNGase F, an enzyme that removes *N*-linked glycans, exhibited minimal decrease in signal, indicating that the majority of GalNAz residues are incorporated into mucin *O*-linked glycans. Despite achieving biased labeling for mucin *O*-linked over *N*-linked glycans, GalNAz has been shown to be incorporated as an *O*-GlcNAc modification.[35] UDP-GalNAz serves as a substrate for GALE and can be epimerized into UDP-GlcNAz. More details of GalNAz as an *O*-GlcNAc MCR can be found in the O-GlcNAc modification Section (4.5). Dube *et al.* evaluated GalNAz labeling in mice.[40] Ac$_4$GalNAz was injected into live mice for 7 days. Following treatment, their organs were harvested and GalNAz incorporation was evaluated using the Staudinger ligation with phosphine-FLAG. GalNAz labeling varied depending on the cell-type. Of particular interest was inconsistent labeling between different types of spleen cells. Splenocytes could be separated into two distinct populations based on labeling efficiency. Further investigation determined that Ac$_4$GalNAz has higher incorporation in B-cells than T-cells, a property that can be leveraged to target specific splenocyte populations.

Despite the success of GalNAz as a global mucin *O*-linked labeler, assessing the specific activity of individual ppGalNAcT isozymes has remained a major challenge. Recently, a collaboration between the Schumann and Bertozzi groups has started to address this limitation by developing bump–hole sugar–protein pairs.[41–44] Their strategy involves synthesizing GalNAc analogs with extended carbon chains ending in a bioorthogonal group on the *N*-acetyl position (Fig. 5). The additionally carbons add a "bump" that prevents endogenous ppGalNAcTs from turning over the corresponding UDP sugar-donor. These reporters were partnered with mutated ppGalNAcT

enzymes with a complementary "hole" in the active site. Initial characterization *in vitro* proved mutant ppGalNAcT enzymes selective modify peptides with UDP-GalNAc bearing a hexynoic acid substitution on the *N*-acetyl position in the presence of wild-type enzymes.[41]

Moving this system into cells is complicated by two factors: mutant ppGalNAcTs must properly function and the bulky GalNAc analogs rely on other enzymes to be converted into UDP sugar-donors. Fortunately, mutant ppGalNAcTs were shown to function properly when stably introduced in HepG2 cells. To circumvent potential issues with the biosynthesis of bulky UDP-GalNAc, the Bertozzi group synthesized a cell permeable protected GalNAc-1-phosphate precursor which was able to bypass the majority of enzymes in the GalNAc salvage pathway leaving only the final step catalyzed by the enzyme AGX1.[42] Mutant AGX1 that can tolerate bulky substituents on the *N*-acetyl position has been developed by the Kohler lab.[45] Introducing these mutants resulted in the successful synthesis of bumped UDP-GalNAz sugar-donors. This engineered system successfully incorporates reporters on cell surface mucin-type glycans, introducing the first gain-of-function strategy toward deconvolution ppGalNAcT functions. However, the need to generate double mutant cell lines expressing exogenous AGX1 and ppGalNAcT is as a major limitation for implementing this strategy in more biologically relevant settings.

4.3. Sialylation

Sialylation is the addition of members a family of 9-carbon monosaccharides called sialic acids on to cell surface glycoproteins.[1] This addition terminates further elongation of the glycan chain. Sialic acid residues are known to have important roles in mitigating cell–cell, cell–glycan, and cell–pathogen interactions. This is due in part to the fact that sialic acids are negatively charged at neutral pH. Their electronegative character also contributes to protein stabilization and ion binding.

As mentioned above, sialic acid metabolism was the first to be interrogated using engineered monosaccharides with the invention of ManLev.[8] Success with ManLev inspired Bertozzi and others to continue developing sialic acid reporters with increasingly selective functionality. The first truly bioorthogonal MCR was created by Saxon and Bertozzi who applied the Staudinger ligation to a reaction between the functionalized ManNAc analog, Ac$_4$ManNAz, and a phosphine tag.[12] Ac$_4$ManNAz is metabolized, forming the corresponding sialic acid bearing an azide, SiaNAz. Incorporation into the cell surface was visualized using the phosphine-FLAG tag followed by flow cytometry. Prescher *et al.* demonstrated that Ac$_4$ManNAz successfully labels live mice.[13] This is the first example of MCR incorporation in a live animal, inspiring the application of Ac$_4$ManNAz to visualize more complicated disease models such as cancer. Increases in sialic acid on the cell surface has been shown to correlate with malignancy of cancer types. Toward this goal, Neves *et al.* investigated the possibility that labeling with Ac$_4$ManNAz could serve as a noninvasive way to monitor tumor progression.[46] Mice with lung carcinoma xenografts

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | 311

were injected intraperitoneally with Ac₄ManNAz and labeled *in vivo* with phosphine tags. Xenograft tissue had elevated fluorescence signal attributed to sialic acid labeling compared to healthy tissue indicating selective labeling.

Azide-containing unnatural sugars were the first to be characterized. However, the advent of CuAAC click-chemistry reactions between azide and terminal alkynes open the possibility for reversing the orientation of reporter and tag functionalities as alkynes are small enough to transduce most carbohydrate salvage pathways. Wong and coworkers developed an alkyne-modified ManNAc probe termed *N*-pentynoylmannosamine (Ac₄ManNAlk), reasoning that it would be converted to the corresponding SiaNAlk sugar and displayed on the cell surface.[47] Treatment of Hep3B cells with Ac₄ManNAlk followed by CuAAC with a fluorogenic azido-hydroxycoumarin probe resulted in robust labeling over background. Switching the orientation of functionalized handle on the probe and tag is advantageous. Even the most efficient bioorthogonal reactions require excess tag to go to completion. Though azides and alkynes are both thought to be abiotic, alkynes exhibit some amount of reactivity in biological environments leading to higher background labeling when using excess alkyne-tag. Therefore, it is worthwhile to generate alkyne-reporters to ensure that the reagent used in excess is the completely inert azide.

MCRs for sialylation so far have been functionalized on solely on the *N*-acetate of the metabolic precursor ManNAc resulting in sialic residues functionalized at the C5 position. To investigate the efficiency of functionalizing other positions, Hackenberger and coworkers synthesized a C4-substituted ManNAc analog, Ac₃-4-azido-ManNAc (Ac₃4AzManNAc).[48] Ac₃4AzManNAc is converted to sialic acid molecules functionalized at the C7 position which can be added to cell surface glycoproteins to label cells. Characterization of Ac₃4AzManNAc labeling showed that it selectively labeled *O*-linked glycans demonstrating an early example of structural changes biasing the metabolic fate of functionalized sugars. Recently, sydnone functionalized sialic acid MCRs have emerged as an alternative to traditional azide bearing reporters. Chinoy *et al.* synthesized sialic acid derivatives with sydnones on either the C5 or *N*-acetyl position, termed Neu5SydCl and Neu9NSydCl respectively.[49] Interestingly, Neu5SydCl was not metabolically efficiently transferred on to protein substrates while Neu9NSydCl showed robust labeling. In addition, Neu9NSydCl could only be tolerated a few STs indicating that Neu9NSydCl labeling could differentiate among various types of sialosides.

4.4. Fucosylation

ʟ-Fucose is a 6-deoxyhexose found on the termini of *N*-linked and *O*-linked glycans (Fig. 1A and B). Structurally, fucose distinguishes itself from other 6-carbon sugars due to its lack of a hydroxyl group at the 6-position and ʟ-configuration. The addition of fucose is catalyzed by a class thirteen of enzymes known as fucosyltransferases (FucTs). FucTs transfer fucose as a high energy sugar-donor, GDP-fucose. Fucosylation has been shown to play a role in a variety of cellular functions including cell–cell and cell-pathogen interactions as well as being elevated during development.[50]

Researchers have exploited the fucose salvage pathway to incorporate chemically reactive groups on cell surface glycans.[4] The first example of a fucose MCR came from the Sawa *et al.* as an azide-bearing fucose analog, Ac₄FucAz (Fig. 5).[51] Ac₄FucAz enters the fucose salvage pathway and is converted to GDP-FucAz. From here, GDP-FucAz is transferred onto cell surface fucosylated glycans. Ac₄FucAz incorporation can be visualized using CuAAC labeling with an alkyne-based fluorescence tag. Concurrently, Rabuka *et al.* synthesized fucose analogs with azide substitutions at the C2, C4, and C6 positions (Fig. 5).[52] 2- and 4-Azido fucose are not tolerated by the fucose salvage pathway and therefore not metabolically incorporated onto glycans. 6-Azido-fucose was successfully incorporated but showed high levels of cytotoxicity. Hsu *et al.* addressed this limitation by switching the orientation of the probe and tag, generating FucAlk probes and observed reduction in cytotoxicity.[47]

All of the aforementioned fucose MCRs suffer from low levels of metabolic labeling in mammalian cells. This limitation prompted the transition to zebrafish to determine the limiting step in fucose metabolism. The first example of this came from Wu and coworkers who synthesized UDP-FucAlk and injected it directly in to zebrafish embryos.[53] To test for incorporation, labeled embryos were subjected to CuAAC with a biocompatible Cu(ɪ)-complex, bis(*tert*-butyltriazolyl) ligand (BTTES) which serves as a copper chelating agent, preventing toxic copper(ɪ) from passing through the cell membrane. 2.5 Hours post injection, fucosylated glycans could be visualized in the enveloping layer indicating that limitations in fucose metabolism occur during the initial steps of the fucose salvage pathway and not the transfer of GDP sugar-donors. Okeley *et al.* were able to improve FucAlk's labeling efficiency using a rate-accelerating, chelating azide-assisted CuAAC reaction leading to the discovery that FucAlk also acts as a metabolic inhibitor of fucosylation.[54] Haltiwanger and coworkers capitalized on the inhibitory effects of fucose MCRs to attenuate Notch signaling in zebrafish.[55] Specifically, Zebrafish embryos treated with FucAlk prevented Notch signaling resulting in the prevention of T-cell differentiation.

Building on these findings, Kizuka *et al.* generated a fucose MCR with an extension off of the C5 position to give 7-alkynyl-fucose, eliminating these inhibitory affects (Fig. 5).[56] Subsequent analysis showed that FucAlk is selectively incorporated into the core position of *N*-linked glycans. Recently, Kizuka and coworkers compared the labeling efficiency of FucAlk and 7-alkynyl-fucose (Fig. 5) to learn more about fucose metabolism.[57] A comparison of these reporters across different cell lines showed differential incorporation that was both cell and protein dependent. Subsequent *in vitro* assays demonstrated that tolerance of FucTs dictates the labeling efficiency of fucose analogs. Cell lines expressing FucTs with larger binding pockets had a higher tolerance for the additional carbon present in 7-alkynyl-fucose and resulting in higher overall labeling efficiency.

4.5. *O*-GlcNAc modification

O-GlcNAcylation refers to the addition of the monosaccharide *N*-acetylglucosamine (GlcNAc) to serine and threonine residues of intracellular and mitochondrial proteins (Fig. 1B).[58–60] *O*-GlcNAc distinguishes itself from other forms of glycosylation in three ways; (1) *O*-GlcNAc substrates are primarily cytosolic, (2) *O*-GlcNAcylation is dynamic, and (3) GlcNAc residues are not further elaborated. The cycling of *O*-GlcNAc is catalyzed by two enzymes, OGT and OGA, which add and remove the modification respectively.[61,62] The majority of GlcNAc molecules are synthesize from glucose *via* the hexosamine biosynthetic pathway (HBP) to be transformed into UDP-GlcNAc. Because GlcNAc is primarily synthesized from the HBP, overall modification levels are sensitive to the amount of circulating glucose. This allows *O*-GlcNAcylation to serve as a nutrient sensor.

Both its dynamic nature and sensitivity to circulating metabolites establishes *O*-GlcNAcylation as a vital regulator of numerous cellular functions including gene expression, signal transduction, stress response, and protein stability. Taken together, these functions solidify *O*-GlcNAcylation as essential to normal growth and development in nearly all multicellular organisms. This is further demonstrated by experiments showing that genetic knockout of OGT is embryonically lethal in Drosophila and mice and tissue specific knockouts of OGT in T-cells causes apoptosis.[63–66] Further, loss of OGT in neurons is associated with neurodegeneration in mice. *O*-GlcNAc levels are consistently seen to be upregulated in cancer and diabetic patients and lowered in neurodegenerative disorders.[67–71]

Efforts have been made by our lab and others to develop a selective *O*-GlcNAc reporters. This is challenging because UDP-GlcNAc is a substrate for other glycosyltransferases and is incorporated into both *N*-linked and *O*-linked glycoproteins. Hang and Vocadlo *et al.* developed the first *O*-GlcNAc reporters termed Ac₄GlcNAz and Ac₄GalNAz (Fig. 5).[37,39] These two molecules are C4 epimers that can be interconverted by the enzyme UDP-glucose 4-epimerase (GALE). They synthesized UDP-GlcNAz and all upstream metabolites of the GlcNAc salvage pathway in order to demonstrate that all relevant enzymes including OGT could tolerate the azide analogs *in vitro*. Additionally, they showed that OGA cleaves GlcNAz at similar rates as GlcNAc indicating that these reporters do not perturb *O*-GlcNAc dynamics. Treated Jurkat cells showed that these reporters could label the constitutively *O*-GlcNAc modified protein family nucleoproteins. A later study by Boyce *et al.* aimed to investigate the cross talk between the GalNAc and GlcNAc salvage pathways.[72] They showed that Ac₄GalNAz did label nucleocytoplasmic proteins and this labeling increased in cells overexpressing OGT. This served as confirmation that GALE was able to interconvert UDP-GalNAz to UDP-GlcNAz. The same study also showed that UDP-GalNAz had superior labeling to GlcNAz presumably due to GalNAz having a more tolerable salvage pathway allowing for more efficient conversion of Ac₄GalNAz to UDP-GalNAz.

Zaro *et al.* improved labeling by generating Ac₄GlcNAlk, an alkyne substituted reporter that can be labeled with CuAAC (Fig. 5).[38] GlcNAlk showed robust labeling with an improved

signal-to-noise ratio compared to GlcNAz. This study further demonstrated that both GlcNAlk and GlcNAz were incorporated and removed as similar rates indicating that GlcNAlk does not affect modification dynamics. Unfortunately, both GlcNAz and GlcNAlk are not specifically incorporated on *O*-GlcNAcylated proteins. Enrichment of lysates labeled with these reporters identified the reporter protein GlyCAM-IgG which is known to be have *N*- and *O*-linked glycosylation. Removal of *N*-linked glycans significantly reduced labeling of both GlcNAz and GlcNAlk, further demonstrating that they are incorporated into *N*-linked glycans.

In an effort to generate an MCR specific for *O*-GlcNAcylation, we sought to exploit OGT's promiscuity to make more exotic sugar analogs. Inspired by work demonstrating that UDP-6-azido-6-deoxy-GlcNAc was a substrate for OGT *in vitro*, we synthesized the reporter Ac₃6AzGlcNAc (Fig. 5).[73,74] Treatment of cells with 6AzGlcNAc showed similar labeling patterns and intensity compared to cells treated with GlcNAz demonstrating that the GlcNAc salvage pathway tolerates modifications made on the C6 position. Enrichment of cell lysate treated with 6AzGlcNAc pulled-down known *O*-GlcNAc modified proteins NEDD4, pyruvate kinase, nucleoporin 62, and FoxO1 but not GlyCAM-IgG demonstrating the selective incorporation into *O*-GlcNAcylation.

The success of Ac₃6AzGlcNAc lead us to seek to improve the labeling efficiency by generating an alkyne derivative also modified at the C6 position: Ac₃6AlkGlcNAc (Fig. 5).[75] Fluorescent labeling with 6AlkGlcNAc showed a similar banding pattern and intensity compared to 6AzGlcNAc but with a better signal-to-noise ratio. Cells treated with 6AlkGlcNAc and then submitted to a proteomic workflow identified caspase-3 and caspase-8 as *O*-GlcNAc modified proteins. Both were confirmed *via* enrichment and western blotting. It was ultimately determined that caspase-8's modification sites were adjacent to its cleavage/activation site. Through a series of experiments, we were able to show that *O*-GlcNAc modifications at these sites blocks cleavage and therefore activation, preventing the activation of the caspase cascade.

To test the breadth of OGT's promiscuity, two additional reporters have been developed; Ac₄2AzGlc and 4-deoxy-GlcNAz (Ac₃4dGlcNAz) (Fig. 5).[76,77] Both are successfully turned over by OGT and appear to be selective for *O*-GlcNAcylation. However, the per-acetylated form of 2AzGlc, Ac₄2AzGlc, is toxic in cells treated for longer periods of time (200 mM for 16 hours). This was consistent with work published by the Yarema lab showing that per-acetylated *versus* partially deprotected MCRs have different metabolic fates and exhibit different toxicities.[78] To address this, we removed the 6-*O*-acetyl group to form Ac₄2AzGlc which reduced the toxicity. 4dGlcNAz was proposed to be more selective for *O*-GlcNAcylation over *N*- and *O*-linked glycosylation because linkages at the C4 position of GlcNAc residues are important for extension of cell surface glycans. Characterization 4dGlcNAz showed selective labeling of intracellular *O*-GlcNAc modified substrates.[77] However, 4dGlcNAz is resistant of removal by OGA.[79] While this increases labeling efficiency, perturbing *O*-GlcNAc dynamics limits the physiological relevance of cellular and *in vivo* experimental results.

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | 313

4.6. Glycosaminoglycan

Glycosaminoglycans (GAGs) are linear polysaccharides consisting of repetitive disaccharide building blocks. GAGs are covalently bound to protein substrates to form glycosylated proteins termed proteoglycans.[80–82] GAGs are classified into four broad subtypes; hyaluronan (HA), keratin sulfate (KS), chondroitin sulfate/dermatan sulfate (CS/DS), and heparin/heparan sulfate (HS) (Fig. 1B). All GAG subtypes share a characteristic pattern of alternating hexosamine and uronic acid subunits with the exception of KS which alternates between hexosamine and galactose. Subtypes differ in their linkages to proteins. KS polysaccharides are linked to underlying N-linked or mucin O-linked glycans while CS/DS and HS chains are attached to proteins at serine residues through a unique core linkage. First, a xylose monosaccharide is attached followed by step-wise addition of two galactose and one glucosamine residue to generate a tetrasaccharide core with the following linkages: GlcAβ(1–3)Galβ(1–3)Galβ(1–3)Xylβ-serine. CS/DS subtypes distinguish themselves from HS through the addition of a GlcNAc residue following this core structure. HS proteoglycans are characterized by GalNAc. Proteoglycans are ubiquitous, having been identified in almost all mammalian cell types. Proteoglycans are secreted into the ECM and either inserted in the plasma membrane or stored in secretory granules.[83] GAGs can be further modified through the addition of sulfate groups on GlcNAc or GalNAc residues resulting in an increased negative charge[80] which facilitates interactions with proteases,[84] growth factors,[84] cytokines,[85] lectins,[85] and structural proteins.[86]

To date, there are no MCRs that label a fully formed proteoglycan. However, a few reporters that are incorporated as modified core monosaccharides leading to premature truncation of the GAG polysaccharide chain but allowing for downstream tagging. One example comes from Linhardt's lab in the synthesis of UDP sugar-donors of the hexosamine subunits, 4AzGlcNAc and 4AzGalNAc (Fig. 5).[87] Linhardt was able to incorporate these probes onto HS and HA polysaccharide chains in vitro. Unfortunately, azide modifications on the C4 position of hexosamine sugars are not tolerated by salvage pathway enzymes, limiting cellular applications for these MCRs.[88] In a similar approach, the Bertozzi lab generated C4 analogs of xylose by synthesizing the UDP donor-sugar of 4-azido-4-deoxy-xylse (4AzXyl) (Fig. 5).[89] To bypass salvage pathway limitations, UDP-4-AzXyl was directly injected into zebrafish embryos resulting in its incorporation into the core position of CS/DS and HS proteoglycans preventing GAG extension but allowing for tagging and analysis of the abundance and distribution proteoglycans of zebrafish cells at different stages of development.

5. Next generation MCRs

In recent years, efforts have been made toward probing glycoproteins in more complicated biological contexts. The following section provides examples of the diverse applications of

MCRs and how they have contributed to major breakthroughs toward a complete understanding of global glycosylation events.

5.1. Cell-specific labeling

The exploratory phase of glycan engineering mainly took place in homogenous cell populations. Preliminary MCR investigations into live animals saw reporters readily incorporated into a myriad of cell types to indiscriminately label tissue.[13] Systemic labeling of whole organisms offers valuable information about glycan density and distribution. However, cell- and tissue-specific labeling in heterogenous cellular environments expands the utility of reporter molecules.[90] For example, selective incorporation of azide-clad monosaccharides on to the surface of cancer cells enables orthogonal conjugation of an anti-cancer drug for proximity based targeted drug delivery. The majority of work toward cell-specific labeling falls into two categories: selective de-caging and liposomal delivery.

Different types of cells express different proteins. Cancer cells in particular have unique expression profiles with many protease enzymes constitutively overexpressed. Researchers have exploited this feature to unleash caged probes into complex biological environments. Bertozzi and coworkers reported the first example of a caged sugar MCR.[91] $Ac_4ManNAz$ was modified with a linker attached to a peptide sequence masking the C6-hydroxyl group. The peptide sequence can be cleaved by a prostate-specific antigen protease (PSA). Once released, the de-caged mannose analog permeates proximal cell membranes leading to selective labeling of PSA-expressing cells. Caged $Ac_4ManNAz$ was also used by Wang et al. in a metastatic breast cancer model.[92] Here, $Ac_4ManNAz$ was caged at the anomeric position via an ether bond linked to an enzyme cleavable moiety. This caged MCR was designed to require sequential cleavage events to be released. Histone deacetylase and Cathepsin L must both be present, adding an additional level of rigor to ensure cell specific activation. This process successfully labeled triple-negative breast cancer; a form notoriously hard to treat in clinical settings.

The caging strategy's major limitation is its reliance on endogenous proteases. This restricts targetable cells to those with unique or overexpressed proteases. Targeting unique proteases also demands the synthesis of different caged monosaccharides for each protease/sugar-analog pairing. In an effort to develop a more general approach, Chen and coworkers have adapted well defined liposomal delivery systems for MCR delivery (Fig. 6A).[93–95] The first iteration involved encapsulating 9AzSia in liposomes functionalized with folate generating a complex termed f-LP-9AzSia.[93] Folate receptors (FRs) are overexpressed in many epithelial tumor tissues including ovarian, breast, lung, and colorectal cancers. Treatment in HeLa cells, an FR expressing cell line, followed by CuAAC assisted with ligand BTTAA conjugation to a fluorescent probe exhibited concentration dependent cell surface labeling. When administered in a mixed population of FR expressing (FR^+) and non-expressing cells, f-LP-9AzSia discriminately labeled FR^+ cells. Chen has since been able to successfully employ this strategy to
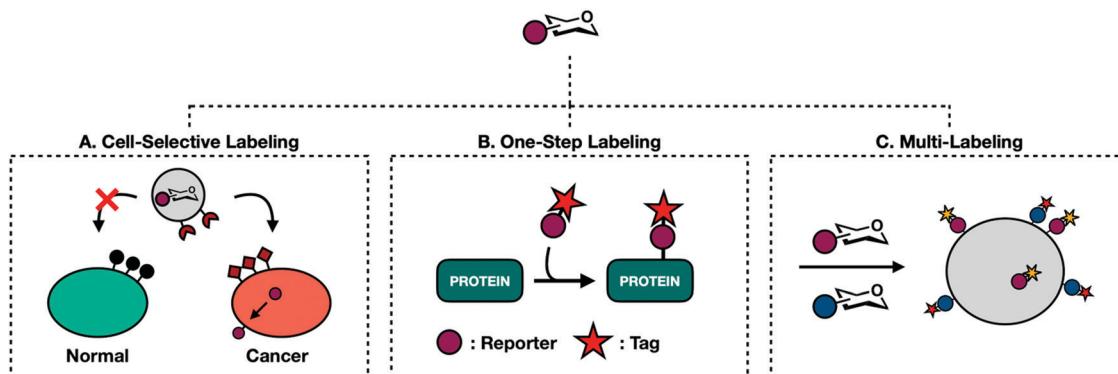
Fig. 6 Second generation MCRs expand their applications. (A) Encapsulating reporters in liposomes allows for targeting of a specific class of cells of tissues. (B) One-step labeling eliminates secondary incubation with tags allowed for direct detection of reporters on protein substrates. (C) The diverse range of bioorthogonal reactions allows for multi-labeling experiments where two bioorthogonal reactions can be performed concurrently without risking cross-labeling.

discriminately label organ tissues and mouse brain cells *in vivo*.[94,95]

5.2. One-step labeling

Traditional MCR labeling relies on a two-step procedure where reporters are first installed onto cellular substrates followed by a reaction with a bioorthogonal probe (Fig. 3). The product is a covalent tertiary structure composed of substrate-reporter-tag. Efficient labeling therefore relies on the rate of both metabolic incorporation and the reaction between reporter and tag. To simplify this process, efforts have been made to generate one-step labeling mechanisms allowing for the direct visualization of unnatural glycans without further conjugation (Fig. 6B). Toward this goal two different techniques have emerged: Raman scattering and pre-conjugated sugar analogs.

Raman spectroscopy can directly detect metabolic probes on the cell surface. Raman spectroscopy uses Raman scattering which can detect specific vibrational wavelengths of molecules of any size. Each type of bond has a specific signal, analogous to a vibrational fingerprint.[96] Chemical reporters containing abiotic functional groups introduce distinct vibrational patterns into cellular environments. Conveniently, the most commonly used bioorthogonal functional groups, azides and alkynes, exhibit vibrational signatures in the Raman-silent region of a cell (between $1800–2800$ cm^{-1}) allowing the signal from these molecules to rise above noise from cellular milieu. The major limitation of Raman spectroscopy is the inherently low signal produced with traditional MCRs.[96] Passive Raman spectroscopy relies on an accumulation of probes in a concentrated location. Given that one goal of glycan engineering is to reflect endogenous activity of glycoproteins and the fact that monosaccharides are generally dispersed throughout the entirety of the cell, they cannot be imaged based on a single concentration point. Chen and coworkers addressed this limitation with surface-enhanced Raman spectroscopy (SERS).[97] SERS employs gold plasmonic nanoparticles (AuNPs) conjugated with 4-mercaptophenylboronic acids (MPBA). MPBA has a high affinity for sialic acids. Cells treated with Ac$_4$ManNAlk

followed by incubation with MPBA-AuNPs gave modified sialic acid residues that could be easily distinguished from their endogenous counterparts using dark field microscopy. This workflow can transition back to a one-step process by making MCRs whose functionalized handle is a small Raman reporter. Unfortunately, metallic nanoparticles are required to boost Raman signal, limiting their use for intracellular imaging.

An alternative approach is to generate pre-conjugated sugar analogs outfitted with fluorescent or biotinylated tags. These can be directly incorporated on to cell surface and intracellular glycans using recombinant glycosyltransferases expressed extracellularly. These enzymes are typically mutated so that their active site can accommodate bulky substituents. This method, termed one-step selective exoenzymatic labeling (SEEL), was pioneered by the Boons group.[98] Initial work with SEEL demonstrated improved labeling efficiency and allowed for the enrichment of a large number of glycoproteins. Hong *et al.* employed a similar approach to transfer fucose analogs conjugated with fluorescent dyes directly onto the surface of zebrafish embryos.[99] These chemoenzymatic methods bypass a sugar's salvage pathway. While this results in robust labeling, it also leads to the loss of valuable information including endogenous enzyme activity and could potentially perturb regular biological functioning. Vocadlo and coworkers have demonstrated that some endogenous glycosyltransferases are tolerant bulky functional groups. Specifically, *O*-GlcNAc transferase (OGT), the enzyme responsible for incorporating GlcNAc on to intracellular proteins, has an incredibly tolerant binding pocket.[61] So tolerant that GlcNAc molecules functionalized with the small fluorophore 4-nitro-2,1,3-benzoxadiazole (NBD) is accepted as an OGT substrate and added onto proteins.[100]

5.3. Multi-labeling

So far, this review has covered singular combinations of reporter-tag ligations. While useful, this approach oversimplifies glycosylation events that are heterogenous in both structure and function. The advent of new bioorthogonal reactions has led to a large enough portfolio of reactive handles that it is

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | **315**

now possible to synthesize MCRs with orthogonal reactive groups. These can be administered at once to perform multi-labeling experiments to probe two or more cellular events concurrently without experiencing cross reactivity (Fig. 6C).

The first example of glycan multi-labeling involved co-treating Jurkat cells with $Ac_4ManLev$ and $Ac_4ManNAz$ to gain insight into sialic acid metabolism. These results demonstrated a higher turnover to SiaNAz than SiaLev.[101] Another example comes from the Wittman group who demonstrated concurrent labeling with $Ac_4ManPtl$, an alkene functionalized IED-DA probe, and $Ac_4GalNAz$.[26] HeLa cells were grown in the presence of both reporters and co-stained with tetrazine and cyclooctyne conjugates allowing for the detection of both on the cell surface. More recently, Schart et al. successfully utilized three orthogonal reactions to triply label HEK293T cells. Mannosamine derivatives, $Ac_4ManNAz$, $Ac_4ManAcryl$, and Ac_4ManN-Cyoc, were able to orthogonally participate in copper-free click chemistry, IED-DA, and photoclick respectively. HEK293T cells treated with all three molecules at once showed independent labeling of each, demonstrating that all three reactions can occur simultaneously with limited background.[102]

Some salvage pathways are so tolerant of structural changes that the same sugar can be functionalized with two different bioorthogonal handles. Feng et al. demonstrated this possibility with 9AzSiaNAlk and 9AzSiaNAz.[103] 9AzSiaNAz has particularly useful applications as a photo-cross linking metabolic reporter. Previously, diazirine functionalized MCRs such as SiaDAz were limited to examining known glycan binding partners that can be readily detected. With 9AzSiaNAz, concurrent incorporation of an azide and diazirine allows for UV-mediated cross-linking products to be identified through biotin enrichment. The initial proof-of-concept study demonstrated that 9AzSiaDAz was able to capture the dimerization of CD22 whereas 9AzSia treated cells only enriched for monomeric CD22.

Arguably the most exciting application of multi-labeling is the potential to probe two separate biological events at once, allowing researchers to answer complicated questions. A seminal example comes from Vocadlo and coworkers. Until recently, O-GlcNAc modified substrates have been thought to be modified exclusively post-translationally. However, a study conducted by Zhu et al. demonstrated that O-GlcNAcylated substrates are modified both post-translationally and co-translationally and that these co-translational modifications have functional roles.[104] Previous labeling strategies offered no way to selectively enrich for O-GlcNAc modifications on nascent peptide chains. Recently, Zhu et al. illustrated that tandem labeling could solve this problem.[105] Cells were treated with two reporters, $Ac_4GalNAz$ to label O-GlcNAc analog and O-propargyl-puromycin (OPP) to prematurely truncate poly-peptides. Because these reporters bear orthogonal tags, dual enrichment can isolate only double-tagged substrates for subsequent analysis. This strategy allowed the Vocadlo group to identify 175 endogenous nascent peptide O-GlcNAc substrates and presented a new covalent labeling technique to isolate and enrich for specific subsets of glycosylated biomolecules.

6. Considerations and limitations of MCRs

Despite the simultaneous advancement selective reporters and improved bioorthogonal reactions, MCRs are not perfect. The following section describes some of the major limitations of reporter molecules.

6.1. Physiological relevance

Whenever an exogenous species is introduced into a biological environment, the system will experience some sort of perturbance. The extent of this perturbance can convolute the results of the experiment, casting doubt as to the physiological relevance of any observations. This is particularly true for MCRs, which rely on the support of a series of endogenous enzymes for incorporation on proteins or oligosaccharide chains (Fig. 2). The resulting data is interpreted as a metabolic fingerprint that lends insights into the fate of monosaccharide building blocks and functions of labeled glycoproteins.[106] The major short-coming of this system is that these unnatural sugars directly compete with native metabolites. Minimally, this means that MCRs turned over by salvage pathway enzymes at the same rate as their analogous sugar moieties disrupt monosaccharide metabolism. More often, MCRs are incorporated less efficiently than their monosaccharide counterparts.[76] While this means that cellular metabolism may not be as greatly affected, it leads to substoichiometric incorporation that misses valuable information. The Hsieh-Wilson lab developed a chemoenzymatic labeling technique that mitigates this limitation.[107,108] They engineered a mutant β-1,4-galactosyltransferase enzyme (Y289L GalT) that adds UDP-GalNAz onto endogenous O-GlcNAc modifications. This method affords quantitative labeling of O-GlcNAcylated proteins. However, chemoenzymatic enrichment is limited to in vitro applications in cell lysates or fixed cells due to the requirement of exogenous GalT and UDP-GalNAz. Together, MCR incorporation and chemoenzymatic labeling offer complimentary techniques that can validate results leading to higher confidence in any conclusions.

6.2. Selectivity

One of the main objectives when designing monosaccharide MCRs is to establish a degree of selectivity. The ability to predict the fate of an MCR allows researchers to definitively draw conclusions from the results of any given experiment. Though an incredible amount of work has been put toward this goal (see Section 4) it still remains a challenge due to the fact that all mammalian glycans are composed of the same few building blocks (Fig. 1A). For example, GlcNAc is a major component of all the main forms of glycosylation (Fig. 1B). Accordingly, the first reported GlcNAc MCR, $Ac_4GlcNAz$, was shown to be incorporated into N-linked, O-linked, and O-GlcNAc modified glycoproteins.[38,39] Introducing selectivity is further complicated by metabolic cross talk. Endogenous salvage pathway enzymes are able to interconvert monosaccharides and high energy sugar-donors.[38,72,109] For instance, GlcNAc can be reversibly converted to both GalNAc and

ManNAc. This is useful for studying endogenous sugar metabolism.[110] However, it complicates most applications of MCRs. Fortunately, individual salvage pathways exhibit some amount of substrate selectivity. These inherent differences can be exploited by generating MCRs with subtle structural changes, biasing labeling toward a single type of glycosylation. Our lab has successfully implemented this approach to develop reporters that specifically label O-GlcNAc modified proteins[74,75] which should encourage the development of increasingly diverse MCRs.

The preceding paragraph describes non-specific enzymatic labeling where the promiscuity of glycosyltransferases and other relevant enzymes leads to unintended incorporation events. A second selectivity issue stems from non-enzymatic chemical labeling of proteins. Chemical labeling was recently discovered as a result of a routine proteomics profiling of HeLa cells with Ac₄ManNAz.[111] MCRs are usually administered as their per-acetylated derivatives. This allows for passive diffusion across the cell membrane.[112,113] Once inside, nonspecific esterase enzymes cleave the acetyl groups to release the reporter molecule (Fig. 2).[114] Masking the polar hydroxyl groups greatly enhances cellular uptake lending to robust labeling with low treatment concentrations. However, Qin et al. identified a significant number of modified cysteine residues in a process termed S-glycosylation.[111] A subsequent mechanistic investigation determined that S-glycosylation follows an elimination-addition mechanism facilitated by the presence of acetyl groups on the C2, C3, and C4 positions of the functionalized sugar[111,115] (Fig. 7). Briefly, per-acetylated sugars are first deacetylated at the anomeric position. The resulting hemiacetal exists in equilibrium between its open and closed confirmations in aqueous environments. When open, the C3 acetyl undergoes base promoted β-elimination to form an α,β-unsaturated aldehyde. This intermediate is then susceptible to Michael-addition with endogenous thiols followed by ring closer to generate 3-thiolated pyranose and furanose adducts. The resulting products are indistinguishable from legitimate enzymatic labeling events without advanced MS-based analysis. The extent of this "off-target" labeling varies by cell type and reporter but has been observed to account for up to 25% of total glycosylation events. Fortunately, lowering the treatment times and concentration when using per-acetylated sugars reduces cysteine labeling.[116] In an effort to circumvent this issue altogether, Qin et al. synthesized two GalNAz derivatives, 1,3-Pr₂GalNaz and 1,6-Pr₂GalNAz (Fig. 7). Both demonstrate high labeling efficiency with low background S-glycosylation.[115,117] Both chemical and enzymatic selectivity issues should be taken into consideration whenever performing an experiment with a MCR. Any identified glycoprotein should be validated using an orthogonal technique such as chemoenzymatic enrichment to rule out artificial labeling or unexpected incorporation.

6.3. Detection and interpretation of labeled proteins

The regulation of glycosylation events is an important part of maintaining proper cellular functions. As such, there has been increasing interest in proteomic profiling to assess a system's
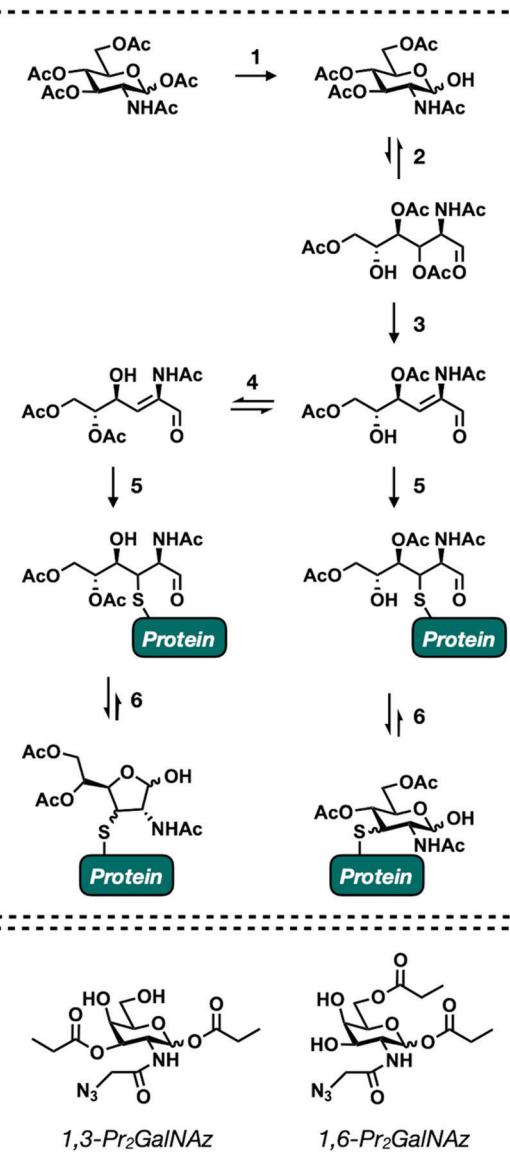

Fig. 7 Mechanistic insights into S-glycosylation. (top) Proposed mechanism for non-enzymatic S-glycosylation. (1) Per-acetylated sugars are deacetylated at the anomeric position. (2) The resulting hemiacetal exists in equilibrium between its open and closed confirmations. (3) Open sugars can undergo β-elimination reaction with a proximal bases. (4) Acyl migration between C4 and C5 position generates two isomers. (5) α,β-Unsaturated aldehydes are susceptible to Michael-addition with endogenous thiols. (6) Ring closure generates 3-thiol furanose and pyranose adducts. (bottom) Structures of next partially acetylated MCRs.

overall glycosylation state. Typically, these studies use mass-spectrometry (MS)-based methods which combine sensitive detection with computational identification. To work effectively, a standard MS workflow relies on predictable metabolic structures which will be searched against database algorithms for peptide identification. Unfortunately, glycans are not predictable. In fact, glycosylation has the greatest amount of proteomic diversity compared with all other PTMs. Glycopeptides can vary in both site occupancy (macroheterogeneity) and oligosaccharide structure (microheterogeneity) (Fig. 8).[118]

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, 2, 306–321 | 317
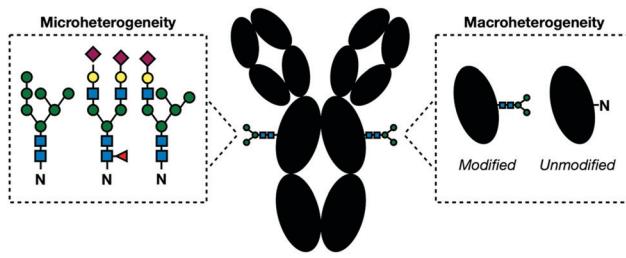
Fig. 8 In a given glycoprotein, glycans can vary in their structure and linkages (microheterogeneity) and presence (macroheterogeneity) complicating their detection and identification.

The resulting heterogenous population of glycans produces a number of substoichiometirc modifications on glycoproteins requiring large amounts of sample to deliver a suitable signal-to-noise ratio. This, in combination with inherently low ionization efficiency of glycoconjugates compared to their unmodified counterparts, makes proper detection of glycoproteins challenging. Even with enough sample, the non-templated nature of glycosylation makes the unambiguous identification of both modification site and context nearly impossible.

With no universal solution, early glycoproteomic investigators adapted a "divide-and-conquer" approach, gathering site and context data in separate MS experiments.[118] These protocols rely on glycoprotein enrichment strategies to fish out relevant proteins from the mix of modified and unmodified substrates. The emergence of MCRs greatly improved enrichment techniques. Functionalized sugars can be selectively reacted with appropriate chemical partners conjugated to an affinity tag such as biotin. Initial investigations relied on the reduction of glycans to a predictable and uniform mass for proper site identification.[119,120] These experiments offer binary "yes-or-no" data sets that are missing information associating glycan structure to modification site undermining its biological significance. Woo *et al.* developed a strategy to solve this problem through the invention of isotope-targeted glycoproteomics (IsoTaG).[121,122] The IsoTaG platform consists the following major components: (1) labeled glycoproteins with MCRs, (2) tagging and enrichment with an isotope recoding affinity probe, and (3) directed tandem MS. Using a dibrominated tag which can be recognized by previously developed IsoStamp software, intact glycopeptides are recognized in real time and selectively submitted to a second ionization event to generate paired glycan and peptide fragments. IsoTaG allows for simultaneous site and context identification. In a proof of concept study, Woo *et al.* labeled Jurkat, PC-3, and MCF-7 cell lines with Ac₄GalNaz and Ac₄ManNAz, illustrating that both are incorporated into a range of different glycosylation subtypes.[121]

Even with improvements in MS methods and enrichment techniques, the field of glycoproteomics still has some major challenges. First, IsoTaG and other technologies are only as selective as the chosen MCR. As such, they suffer from the same limitations as those addressed in Section 6.2. This was exemplified in recent work from the Kohler lab.[110] Here, efforts to generate a probe-free technique for the identification of glycoproteins illuminated the vast metabolic cross talk that

interconverts GalNaz, GlcNAz and ManNAz. Additionally, there is still no universal MS workflow that generates site, structure, and biological context in one mega experiment. Native structures, PTM crosstalk, and protein–glycan interactions are commonly lost through the multiple enrichment and digestion steps that yield glycopeptides.

7. Conclusions and future outlook

Metabolic chemical reporters revolutionized the study of glycosylation. Improvements in both MCR design and bioorthogonal chemistries have expanded the types of imaging and characterization experiments possible. These advancements allow for *in vivo* studies, offering simple solutions to answer complicated questions. Though limitations exist, advancements in MS-based technology coupled with a greater understanding of monosaccharide metabolism help to solve these problems. The future of MCRs shows great promise in both basic and industrial research. From a fundamental standpoint, MCRs are making great headway toward mapping the total glycome. These findings, in combination with advancements in medicinal chemistry, can advise advancements in pharmaceutical development.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

References

1 M. E. E. Ajit Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi and G. W. Hart, *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York, NY, 2nd edn, 2009.

2 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.

3 L. L. Kiessling and R. A. Splain, *Annu. Rev. Biochem.*, 2010, **79**, 619–653.

4 E. C. H. Hanako Ishihara and D. J. Massaro, *J. Biol. Chem.*, 1968, **243**, 1103–1109.

5 H. Kayser, R. Zeitler, C. Kannicht, D. Grunow, R. Nuck and W. Reutter, *J. Biol. Chem.*, 1992, **267**, 16934–16938.

6 M. P. Oliver, T. Keppler, P. Stehling, M. Herrmann, H. Kayser, D. Grunow and W. Reutter, *J. Biol. Chem.*, 1995, **270**, 1308–1314.

7 O. T. Keppler, R. Horstkorte, M. Pawlita, C. Schmidt and W. Reutter, *Glycobiology*, 2001, **11**, 11–18.

8 C. R. Mahal, L. K. Yarema and K. J. Bertozzi, *Science*, 1997, **276**, 1125–1128.

9 D. M. Patterson, L. A. Nazarova and J. A. Prescher, *ACS Chem. Biol.*, 2014, **9**, 592–605.

10 R. D. Row and J. A. Prescher, *Acc. Chem. Res.*, 2018, **51**, 1073–1081.

11 T. Cañeque, S. Müller and R. Rodriguez, *Nat. Rev. Chem.*, 2018, **2**, 202–215.

12 E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007–2010.

13 J. A. Prescher, D. H. Dube and C. R. Bertozzi, *Nature*, 2004, **430**, 873–877.

14 J. A. Prescher and C. R. Bertozzi, *Chem. Living Syst.*, 2005, **1**, 13–21.

15 M. Meldal and C. W. Tomøe, *Chem. Rev.*, 2008, **108**, 2952–3015.

16 C. G. Parker and M. R. Pratt, *Cell*, 2020, **180**, 605–632.

17 E. M. Sletten and C. R. Bertozzi, *Acc. Chem. Res.*, 2011, **44**, 666–676.

18 J. C. Jewett and C. R. Bertozzi, *Chem. Soc. Rev.*, 2010, **39**, 1272–1279.

19 N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046–15047.

20 M. L. Blackman, M. Royzen and J. M. Fox, *J. Am. Chem. Soc.*, 2008, **130**, 13518–13519.

21 R. D. Bach, *J. Am. Chem. Soc.*, 2009, **131**, 5233–5243.

22 R. Selvaraj and J. M. Fox, *Curr. Opin. Chem. Biol.*, 2013, **17**, 753–760.

23 Z. Yu, Y. Pan, Z. Wang, J. Wang and Q. Lin, *Angew. Chem., Int. Ed.*, 2012, **51**, 10600–10604.

24 D. M. Patterson, L. A. Nazarova, B. Xie, D. N. Kamber and J. A. Prescher, *J. Am. Chem. Soc.*, 2012, **134**, 18638–18643.

25 C. M. Cole, J. Yang, J. ŠečkuteC and N. K. Devaraj, *ChemBioChem*, 2013, **14**, 205–208.

26 A. Niederwieser, A. K. Späte, L. D. Nguyen, C. Jüngst, W. Reutter and V. Wittmann, *Angew. Chem., Int. Ed.*, 2013, **52**, 4265–4268.

27 J. E. G. A. Dold, J. Pfotzer, A. K. Späte and V. Wittmann, *ChemBioChem*, 2017, **18**, 1242–1250.

28 J. E. G. A. Dold and V. Wittmann, *ChemBioChem*, 2020, 1–10.

29 R. D. Row, H. W. Shih, A. T. Alexander, R. A. Mehl and J. A. Prescher, *J. Am. Chem. Soc.*, 2017, **139**, 7370–7375.

30 A. Helenius and M. Aebi, *Annu. Rev. Biochem.*, 2004, **73**, 1019–1049.

31 F. Schwarz and M. Aebi, *Curr. Opin. Struct. Biol.*, 2011, **21**, 576–582.

32 M. A. Breidenbach, J. E. G. Gallagher, D. S. King, B. P. Smart, P. Wu and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 3988–3993.

33 H. C. Hang and C. R. Bertozzi, *Bioorg. Med. Chem.*, 2005, **13**, 5021–5034.

34 K. G. Ten Hagen, T. A. Fritz and L. A. Tabak, *Glycobiology*, 2003, **13**, 1–16.

35 P. H. Jensen, D. Kolarich and N. H. Packer, *FEBS J.*, 2010, **277**, 81–94.

36 R. Bansil, E. Stanley and J. T. LaMont, *Annu. Rev. Physiol.*, 1995, **57**, 635–657.

37 H. C. Hang, C. Yu, D. L. Kato and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14846–14851.

38 B. W. Zaro, Y. Y. Yang, H. C. Hang and M. R. Pratt, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 8146–8151.

39 D. J. Vocadlo, H. C. Hang, E. J. Kim, J. A. Hanover and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9116–9121.

40 D. H. Dube, J. A. Prescher, C. M. Quang and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4819–4824.

41 J. Choi, L. J. S. Wagner, S. B. P. E. Timmermans, S. A. Malaker, B. Schumann, M. A. Gray, M. F. Debets, M. Takashima, J. Gehring and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2019, **141**, 13442–13453.

42 B. Schumann, S. A. Malaker, S. P. Wisnovsky, M. F. Debets, A. J. Agbay, D. Fernandez, L. J. S. Wagner, L. Lin, Z. Li, J. Choi, D. M. Fox, J. Peh, M. A. Gray, K. Pedram, J. J. Kohler, M. Mrksich and C. R. Bertozzi, *Mol. Cell*, 2020, **78**, 824–834.e15.

43 M. F. Debets, O. Y. Tastan, S. P. Wisnovsky, S. A. Malaker, N. Angelis, L. K. R. Moeckl, J. Choi, H. Flynn, L. J. S. Wagner, G. Bineva-Todd, A. Antonopoulos, A. Cioce, W. M. Browne, Z. Li, D. C. Briggs, H. L. Douglas, G. T. Hess, A. J. Agbay, C. Roustan, S. Kjaer, S. M. Haslam, A. P. Snijders, M. C. Bassik, W. E. Moerner, V. S. W. Li, C. R. Bertozzi and B. Schumann, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 25293–25301.

44 A. Cioce, S. A. Malaker and B. Schumann, *Curr. Opin. Chem. Biol.*, 2021, **60**, 66–78.

45 S. H. Yu, M. Boyce, A. M. Wands, M. R. Bond, C. R. Bertozzi and J. J. Kohler, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4834–4839.

46 A. A. Neves, H. Stöckmann, R. R. Harmston, H. J. Pryor, I. S. Alam, H. Ireland-Zecchini, D. Y. Lewis, S. K. Lyons, F. J. Leeper and K. M. Brindle, *FASEB J.*, 2011, **25**, 2528–2537.

47 T. L. Hsu, S. R. Hanson, K. Kishikawa, S. K. Wang, M. Sawa and C. H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 2614–2619.

48 H. Möller, V. Böhrsch, J. Bentrop, J. Bender, S. Hinderlich and C. P. R. Hackenberger, *Angew. Chem., Int. Ed.*, 2012, **51**, 5986–5990.

49 Z. S. Chinoy, C. Bodineau, C. Favre, K. W. Moremen, R. V. Durán and F. Friscourt, *Angew. Chem.*, 2019, **131**, 4325–4329.

50 D. J. Becker and J. B. Lowe, *Glycobiology*, 2003, **13**(7), 41R–53R.

51 M. Sawa, T. L. Hsu, T. Itoh, M. Sugiyama, S. R. Hanson, P. K. Vogt and C. H. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 12371–12376.

52 D. Rabuka, S. C. Hubbard, S. T. Laughlin, S. P. Argade and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2006, **128**, 12078–12079.

53 D. Soriano Del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow and P. Wu, *J. Am. Chem. Soc.*, 2010, **132**, 16893–16899.

54 N. M. Okeley, S. C. Alley, M. E. Anderson, T. E. Boursalian, P. J. Burke, K. M. Emmerton, S. C. Jeffrey, K. Klussman, C. L. Law, D. Sussman, B. E. Toki, L. Westendorf, W. Zeng, X. Zhang, D. R. Benjamin and P. D. Senter, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 5404–5409.

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | 319

55 M. Schneider, V. Kumar, L. U. Nordstrøm, L. Feng, H. Takeuchi, H. Hao, V. C. Luca, K. C. Garcia, P. Stanley, P. Wu and R. S. Haltiwanger, *Nat. Chem. Biol.*, 2018, **14**, 65–71.

56 Y. Kizuka, S. Funayama, H. Shogomori, M. Nakano, K. Nakajima, R. Oka, S. Kitazume, Y. Yamaguchi, M. Sano, H. Korekane, T. L. Hsu, H. Y. Lee, C. H. Wong and N. Taniguchi, *Cell Chem. Biol.*, 2016, **23**, 782–792.

57 C. Ma, H. Takeuchi, H. Hao, C. Yonekawa, K. Nakajima, M. Nagae, T. Okajima, R. S. Haltiwanger and Y. Kizuka, *Int. J. Mol. Sci.*, 2020, **21**, 1–18.

58 M. R. Bond and J. A. Hanover, *J. Cell Biol.*, 2015, **208**, 869–880.

59 N. E. Zachara, *FEBS Lett.*, 2018, **592**, 3950–3975.

60 X. Yang and K. Qian, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 452–465.

61 Z. G. Levine and S. Walker, *Annu. Rev. Biochem.*, 2016, **85**, 631–657.

62 D. T. King, A. Males, G. J. Davies and D. J. Vocadlo, *Curr. Opin. Chem. Biol.*, 2019, **53**, 131–144.

63 N. O'Donnell, N. E. Zachara, G. W. Hart and J. D. Marth, *Mol. Cell. Biol.*, 2004, **24**, 1680–1690.

64 R. Shafi, S. P. N. Iyer, L. G. Ellies, N. O'Donnell, K. W. Marek, D. Chui, G. W. Hart and J. D. Marth, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5735–5739.

65 D. A. R. Sinclair, M. Syrzycka, M. S. Macauley, T. Rastgardani, I. Komljenovic, D. J. Vocadlo, H. W. Brock and B. M. Honda, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13427–13432.

66 A. C. Wang, E. H. Jensen, J. E. Rexach, H. V. Vinters and L. C. Hsieh-Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 15120–15125.

67 Z. Ma and K. Vosseller, *Amino Acids*, 2013, **45**, 719–733.

68 C. M. Ferrer, V. L. Sodi and M. J. Reginato, *J. Mol. Biol.*, 2016, **428**, 3282–3294.

69 G. M. Trinca and C. R. Hagan, *J. Bioenerg. Biomembr.*, 2018, **50**(3), 199–204.

70 P. S. Banerjee, O. Lagerlöf and G. W. Hart, *Mol. Aspects Med.*, 2016, **51**, 1–15.

71 X. Ma, H. Li, Y. He and J. Hao, *Cell. Mol. Life Sci.*, 2017, **74**, 3667–3686.

72 M. Boyce, I. S. Carrico, A. S. Ganguli, S. H. Yu, M. J. Hangauer, S. C. Hubbard, J. J. Kohler and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3141–3146.

73 A. Mayer, T. M. Gloster, W. K. Chou, D. J. Vocadlo and M. E. Tanner, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1199–1201.

74 K. N. Chuh, B. W. Zaro, F. Piller, V. Piller and M. R. Pratt, *J. Am. Chem. Soc.*, 2014, **136**, 12283–12295.

75 K. N. Chuh, A. R. Batt, B. W. Zaro, N. Darabedian, N. P. Marotta, C. K. Brennan, A. Amirhekmat and M. R. Pratt, *J. Am. Chem. Soc.*, 2017, **139**, 7872–7885.

76 B. W. Zaro, A. R. Batt, K. N. Chuh, M. X. Navarro and M. R. Pratt, *ACS Chem. Biol.*, 2017, **12**, 787–794.

77 D. L. Shen, T. W. Liu, W. Zandberg, T. Clark, R. Eskandari, M. G. Alteen, H. Y. Tan, Y. Zhu, S. Cecioni and D. Vocadlo, *ACS Chem. Biol.*, 2017, **12**, 206–213.

78 U. Aich, C. T. Campbell, N. Elmouelhi, C. A. Weier, S. G. Sampathkumar, S. S. Choi and K. J. Yarema, *ACS Chem. Biol.*, 2008, **3**, 230–240.

79 J. Li, J. Wang, L. Wen, H. Zhu, S. Li, K. Huang, K. Jiang, X. Li, C. Ma, J. Qu, A. Parameswaran, J. Song, W. Zhao and P. G. Wang, *ACS Chem. Biol.*, 2016, **11**, 3002–3006.

80 D. Soares Da Costa, R. L. Reis and I. Pashkuleva, *Annu. Rev. Biomed. Eng.*, 2017, **19**, 1–26.

81 H. E. Bülow and O. Hobert, *Annu. Rev. Cell Dev. Biol.*, 2006, **22**, 375–407.

82 T. M. Handel, Z. Johnson, S. E. Crown, E. K. Lau, M. Sweeney and A. E. Proudfoot, *Annu. Rev. Biochem.*, 2005, **74**, 385–410.

83 V. H. Pomin and B. Mulloy, *Pharmaceuticals*, 2018, **11**, 1–9.

84 J. Douaiher, J. Succar, L. Lancerotto, M. F. Gurish, D. P. Orgill, M. J. Hamilton, S. A. Krilis and R. L. Stevens, *Development of mast cells and importance of their tryptase and chymase serine proteases in inflammation and wound healing*, Elsevier Inc., 1st edn, 2014, vol. 122.

85 A. Korpetinou, S. S. Skandalis, V. T. Labropoulou, G. Smirlaki, A. Noulas, N. K. Karamanos and A. D. Theocharis, *Front. Oncol.*, 2014, **4**, 1–12.

86 R. V. Iozzo, *Annu. Rev. Biochem.*, 1998, **67**, 609–652.

87 X. Zhang, D. E. Green, V. L. Schultz, L. Lin, X. Han, R. Wang, A. Yaksic, S. Y. Kim, P. L. Deangelis and R. J. Linhardt, *J. Org. Chem.*, 2017, **82**, 9910–9915.

88 W. Guan, L. Cai, J. Fang, B. Wu and P. G. Wang, *Chem. Commun.*, 2009, 6976–6978.

89 B. J. Beahm, K. W. Dehnert, N. L. Derr, J. Kuhn, J. K. Eberhart, D. Spillmann, S. L. Amacher and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2014, **53**, 3347–3352.

90 R. Xie, S. Hong and X. Chen, *Curr. Opin. Chem. Biol.*, 2013, **17**, 747–752.

91 P. V. Chang, D. H. Dube, E. M. Sletten and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2010, **132**, 9516–9518.

92 H. Wang, R. Wang, K. Cai, H. He, Y. Liu, J. Yen, Z. Wang, M. Xu, Y. Sun, X. Zhou, Q. Yin, L. Tang, I. T. Dobrucki, L. W. Dobrucki, E. J. Chaney, S. A. Boppart, T. M. Fan, S. Lezmi, X. Chen, L. Yin and J. Cheng, *Nat. Chem. Biol.*, 2017, **13**, 415–424.

93 R. Xie, S. Hong, L. Feng, J. Rong and X. Chen, *J. Am. Chem. Soc.*, 2012, **134**, 9914–9917.

94 R. Xie, L. Dong, R. Huang, S. Hong, R. Lei and X. Chen, *Angew. Chem.*, 2014, **126**, 14306–14310.

95 Y. Sun, S. Hong, R. Xie, R. Huang, R. Lei, B. Cheng, D. E. Sun, Y. Du, C. M. Nycholat, J. C. Paulson and X. Chen, *J. Am. Chem. Soc.*, 2018, **140**, 3592–3602.

96 L. Lin, X. Tian, S. Hong, P. Dai, Q. You, R. Wang, L. Feng, C. Xie, Z. Q. Tian and X. Chen, *Angew. Chem., Int. Ed.*, 2013, **52**, 7266–7271.

97 M. Xiao, L. Lin, Z. Li, J. Liu, S. Hong, Y. Li, M. Zheng, X. Duan and X. Chen, *Chem. – Asian J.*, 2014, **9**, 2040–2044.

98 T. Sun, S. H. Yu, P. Zhao, L. Meng, K. W. Moremen, L. Wells, R. Steet and G. J. Boons, *J. Am. Chem. Soc.*, 2016, **138**, 11575–11582.

99 S. Hong, P. Sahai-Hernandez, D. G. Chapla, K. W. Moremen, D. Traver and P. Wu, *Angew. Chem.*, 2019, **131**, 14465–14471.

100 H. Y. Tan, R. Eskandari, D. Shen, Y. Zhu, T. W. Liu, L. I. Willems, M. G. Alteen, Z. Madden and D. J. Vocadlo, *J. Am. Chem. Soc.*, 2018, **140**, 15300–15308.

101 E. Saxon, S. J. Luchansky, H. C. Hang, C. Yu, S. C. Lee and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2002, **124**, 14893–14902.

102 V. F. Schart, J. Hassenrück, A. K. Späte, J. E. G. A. Dold, R. Fahrner and V. Wittmann, *ChemBioChem*, 2019, **20**, 166–171.

103 L. Feng, S. Hong, J. Rong, Q. You, P. Dai, R. Huang, Y. Tan, W. Hong, C. Xie, J. Zhao and X. Chen, *J. Am. Chem. Soc.*, 2013, **135**, 9244–9247.

104 Y. Zhu, T. W. Liu, S. Cecioni, R. Eskandari, W. F. Zandberg and D. J. Vocadlo, *Nat. Chem. Biol.*, 2015, **11**, 319–325.

105 Y. Zhu, L. I. Willems, D. Salas, S. Cecioni, W. B. Wu, L. J. Foster and D. J. Vocadlo, *J. Am. Chem. Soc.*, 2020, **142**, 15729–15739.

106 N. J. Pedowitz, B. W. Zaro and M. R. Pratt, *Curr. Protoc. Chem. Biol.*, 2020, **12**, 1–20.

107 N. Khidekel, S. Arndt, N. Lamarre-Vincent, A. Lippert, K. G. Poulin-Kerstien, B. Ramakrishnan, P. K. Qasba and L. C. Hsieh-Wilson, *J. Am. Chem. Soc.*, 2003, **125**, 16162–16163.

108 P. M. Clark, J. F. Dweck, D. E. Mason, C. R. Hart, S. B. Buck, E. C. Peters, B. J. Agnew and L. C. Hsieh-Wilson, *J. Am. Chem. Soc.*, 2008, **130**, 11576–11577.

109 L. A. Bateman, B. W. Zaro, K. N. Chuh and M. R. Pratt, *Chem. Commun.*, 2013, **49**, 4328–4330.

110 A. Shajahan, N. T. Supekar, H. Wu, A. M. Wands, G. Bhat, A. Kalimurthy, M. Matsubara, R. Ranzinger, J. J. Kohler and P. Azadi, *ACS Chem. Biol.*, 2020, **15**, 2692–2701.

111 W. Qin, K. Qin, X. Fan, L. Peng, W. Hong, Y. Zhu, P. Lv, Y. Du, R. Huang, M. Han, B. Cheng, Y. Liu, W. Zhou, C. Wang and X. Chen, *Angew. Chem.*, 2018, **130**, 1835–1838.

112 A. K. Sarkar, T. A. Fritz, W. H. Taylor and J. D. Esko, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 3323–3327.

113 S. T. Lauehlin and C. R. Bertozzi, *Nat. Protoc.*, 2007, **2**, 2930–2944.

114 M. P. Mathew, E. Tan, S. Shah, R. Bhattacharya, M. Adam Meledeo, J. Huang, F. A. Espinoza and K. J. Yarema, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 6929–6933.

115 K. Qin, H. Zhang, Z. Zhao and X. Chen, *J. Am. Chem. Soc.*, 2020, **142**, 9382–9388.

116 N. Darabedian, B. Yang, R. Ding, G. Cutolo, B. W. Zaro, C. M. Woo and M. R. Pratt, *Front. Chem.*, 2020, **8**, 1–14.

117 Y. Hao, X. Fan, Y. Shi, C. Zhang, D. En Sun, K. Qin, W. Qin, W. Zhou and X. Chen, *Nat. Commun.*, 2019, **10**, 4065.

118 H. Zhu, A. Aloor, C. Ma, S. M. Kondengaden and P. G. Wang, *ACS Symp. Ser.*, 2020, **1346**, 169–203.

119 J. M. Smeekens, W. Chen and R. Wu, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 604–614.

120 C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T. B. G. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, R. Gupta, E. Paul Bennett, U. Mandel, S. Brunak, H. H. Wandall, S. B. Levery and H. Clausen, *EMBO J.*, 2013, **32**, 1478–1488.

121 C. M. Woo, A. T. Iavarone, D. R. Spiciarich, K. K. Palaniappan and C. R. Bertozzi, *Nat. Methods*, 2015, **12**, 561–567.

122 C. M. Woo, P. J. Lund, A. C. Huang, M. M. Davis, C. R. Bertozzi and S. J. Pitteri, *Mol. Cell. Proteomics*, 2018, **17**, 764–775.

© 2021 The Author(s). Published by the Royal Society of Chemistry

RSC Chem. Biol., 2021, **2**, 306–321 | **321**