REVIEWS

Vertebrate protein glycosylation: diversity, synthesis and function

Kelley W. Moremen, Michael Tiemeyer, and Alison V. Nairn

Abstract | Protein glycosylation is a ubiquitous post-translational modification found in all domains of life. Despite their significant complexity in animal systems, glycan structures have crucial biological and physiological roles, from contributions in protein folding and quality control to involvement in a large number of biological recognition events. As a result, they impart an additional level of 'information content' to underlying polypeptide structures. Improvements in analytical methodologies for dissecting glycan structural diversity, along with recent developments in biochemical and genetic approaches for studying glycan biosynthesis and catabolism, have provided a greater understanding of the biological contributions of these complex structures in vertebrates.

Lectins
Carbohydrate binding proteins that are involved in various biological recognition phenomena.

Sialic acid
(SA; also known as neuraminic acid (Neu)). SAs are a family of nine-carbon monosaccharides with a carboxylic acid at the C1 position and a glycerol side chain at C7-C9. Two SAs predominate in vertebrates, with either an *N*-acetyl group (NeuAc) or an *N*-glycolyl group (NeuGc) at C5. Humans do not make NeuGc, but can obtain it from their diet and incorporate it into glycoconjugates.

Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA. Correspondence to K.W.M. e-mail: moremen@uga.edu

Glycans associated with cell surface and intracellular proteins and lipids contribute to numerous biological functions in animal systems[1]. Oligosaccharide structures at the cell surface influence interactions with the extracellular environment by providing ligands for cell adhesion, macromolecule interactions and pathogen invasion[1,2]. The glycans associated with cell surface receptors and proteins also directly modulate protein function and signalling, as well as altering the dynamics of glycoprotein endocytosis and cell surface half-life through binding to multivalent lectins[3]. Glycan structures on newly synthesized glycoproteins are crucial for protein secretion, as they influence protein folding, provide ligands for lectin chaperones, contribute to quality control surveillance in the endoplasmic reticulum (ER) and mediate transit and selective protein targeting throughout the secretory pathway[4,5]. More general roles of glycan structures include contributions to the regulation of cytosolic and nuclear functions, immune surveillance, inflammatory reactions, autoimmunity, hormone action and tumour metastasis[1,6]. It is clear that these post-translational modifications confer an additional information content at the molecular and cellular level that extends beyond the simple information flow from genome-derived transcripts to encoded protein functions[7].

Protein glycosylation is broadly used by cell biologists to monitor protein transit through the secretory pathway (for example, endoglycosidase H-resistant glycan structures are used to monitor protein translocation in the Golgi[8]). Moreover, glycans detected by specific antibodies[9] and ectopically expressed fluorescently tagged glycosylation enzymes[10] are used as markers for intracellular compartments. Glycan structures also distinguish cell types in developing and mature animal tissues[2]. Despite this limited use of mammalian glycan structures or glycosylation machinery as tools for cell biology studies, few cell biologists have truly embraced the diversity and complexity of glycan modifications for their contributions and roles in biological systems. This results from an intimidating collection of challenges for studying the functions of glycan structures at the molecular, cellular and organismal level.

It has been estimated that approximately 700 proteins are required to generate the full diversity of mammalian glycans (estimated to be ≥7,000 structures), which are assembled from only ten monosaccharides: fucose (Fuc), galactose (Gal), glucose (Glc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), glucuronic acid (GlcA), iduronic acid (IdoA), mannose (Man), sialic acid (SA) and xylose (Xyl)[7,11,12]. Among these proteins, ~200 are glycosyltransferases, which are enzymes that extend acceptor glycan structures using nucleotide or lipid-linked sugars as activated donor substrates. Competition between glycosyltransferases that possess overlapping glycan acceptor preferences but different donor specificities can strongly influence the relative abundance of glycan structural features in the total glycome of a cell or tissue. Thus, the glycan biosynthetic potential of a given cell is determined by multiple factors, including relative enzyme abundance and localization, abundance and trafficking of glycoprotein substrates, and the availability of activated sugar donors in appropriate secretory compartments.

Table 1 | Consensus motifs and enzymes responsible for various glycosylation reactions*

| Type	Linkage	Enzyme	Consensus sequence	Domain	Examples
N-glycosylation	GlcNAc-β-Asn	OST	N-X-(S/T)[‡] (standard sequons)		Nascent polypeptides
			N-X-C[‡,§], N-G[§], N-X-V[‡,§] (non-standard sequons)		
O-glycosylation	GalNAc-α-Ser/Thr	ppGALNTs	Isoform specific[‖]		Mucins
	GlcNAc-β-Ser/Thr	OGT	No set consensus[¶]		Cytosolic, nuclear proteins
	GlcNAc-β-Ser/Thr	EOGT	Unknown	EGF	Notch, Dumpy
	Xyl-β-Ser	XYLT1, XYLT2	a-a-a-a-G-S-G-a-(a/G)-a ('a' represents Asp or Glu)[#]		Heparin, proteoglycan core proteins
	Fuc-α-Ser/Thr	POFUT2	C-X-X-(S/T)-C-X-X-G	TSR	Thrombospondin, ADAMTS family
		POFUT1	C^2-X-X-X-X-(S/T)-C^3	EGF	Notch, clotting factors
	Glc-β-Ser	POGLUT1	C^1-X-S-X-(A/P)-C^2	EGF	Notch, clotting factors
	Man-α-Ser/Thr	POMT1, POMT2	I-X-P-T-(P/X)-T-X-P-X-X-X-X-P-T-X-(T/X)-X-X**		α-dystroglycan
	Gal-β-Hyl	GLT25D1, GLT25D2	X-Hyl-Gly (collagen repeats)[‡‡]		Collagen, adiponectin
	Glc-α-Tyr	GYG	Tyr194 of GYG[§§]		GYG (autoglycosylation during glycogen formation)
C-mannosylation	Man-α-Trp	Unknown	W-X-X-W[‖‖]	TSR	Thrombospondin, ADAMTS family
Glypiation	Pr-C(O)EthN-6-P-Man	Transamidase	No set consensus[¶¶]		THY1, NCAM1

*References provided in the text except where noted. [‡]X cannot be Pro. [§]Determined by high-throughput proteomic analysis of lectin-enriched glycopeptides[21]. Other non-standard sequons have been observed for recombinant immunoglobulin G2 antibodies expressed in Chinese hamster ovary cells[23] or mouse laminin[22], but these structures have not been confirmed more widely in animal systems. [‖]Variation in ppGALNT specificities was determined empirically[154]. [¶]Prediction software available from the YinOYang WWW server[155]. [#]The specificity for the initiation of proteoglycan core synthesis is defined empirically[156]. **Proposed consensus[157] with variable support from mass spectrometry analysis[123]. [‡‡]Collagen domains are modified at Lys residues to form hydroxylysine (Hyl) and galactosylated, then they are extended with an α1,2-Glc residue prior to triple helix formation[144]. [§§]Autoglycosylation and extension of the initial glycogen polymer on the glycogenin backbone occurs on the Tyr194 hydroxyl group[143]. [‖‖]Mannose liked to the C2 of the indole ring of the tryptophan residue based on the corresponding consensus sequence[111]. [¶¶]Prediction of consensus sequence has been defined based on hidden Markov model and prediction software[158]. ADAMTS, a disintegrin and metalloproteinase with thrombospondin motifs; EOGT, EGF domain-specific O-linked GlcNAc transferase; GLT25D1/2, glycosyltransferase 25 family member 1/2; GYG, glycogenin; NCAM1, neural cell adhesion molecule 1; OGT, O-linked GlcNAc transferase; OST, oligosaccharyltransferase; POFUT, protein O-fucosyltransferase; POGLUT, protein O-glucosyltransferase; POMT, protein O-mannosyltransferase; ppGALNTs, polypeptide GalNAc transferases; TSR, thrombospondin type 1 repeat; XYLT, xylosyltransferase.

Glycans can be attached to polypeptide structures through amide linkages to Asn side chains (N-glycosylation), through glycosidic linkages (O-glycosylation) to side chains of Ser/Thr, hydroxylysine (collagen) or Tyr (glycogenin), or through C-C linkages to the C2 position of Trp (C-mannosylation). Alternatively, they be can attached as a linker structure bridging glycosylphosphatidylinositol anchors to protein backbones[13] (TABLE 1). Even among the O-linked structures to Ser/Thr, at least six distinct glycan types are known in animal systems (BOX 1; TABLE 1), with most found on extracellular, cell surface or intracellular membrane compartments. An O-linked monosaccharide addition (GlcNAc-β-Ser/Thr) is also found on a large number of cytosolic and nuclear proteins, where it contributes to the regulation of numerous cellular functions and can compete with Ser/Thr phosphorylation[6]. Monosaccharides and complex extended glycan structures are also found attached to many other macromolecule classes in animal systems, from glycolipids to secondary metabolites, but these latter structures are beyond the scope of this Review.

Glycan structures attached to proteins can be highly complex, with numerous possibilities for branching and anomeric linkage, and accordingly they have much greater structural diversity than linear nucleic acid or polypeptide structures[7]. Additional complexity and diversity arises from the fact that mammalian glycans are synthesized in intricate biosynthetic pathways, and the resulting products depend on potential competition among multiple enzymes for the same glycan substrates[12,14]. This competition, when combined with the relative inefficiency of the enzymatic reactions, often results in a diverse ensemble of branched glycan structures (termed 'microheterogeneity'), even at the same glycosylation site of otherwise identical proteins[15]. The diversity of glycan modification can also produce complex pleiotropy, where unique modifications on one glycosylation site may alter function or recognition within a specific cellular context, but may cause other effects or be functionally silent in other contexts. Finally, the diversity of glycan structures found in a typical cell or tissue sample presents considerable challenges for structural elucidation, even when using the most contemporary analytical approaches[16]. Thus, the goal of determining the roles and contributions of individual oligosaccharides in various biological contexts is equivalent to hunting for a handful of discrete glycan functions within a forest of

Box 1 | Major classes of vertebrate glycan structures
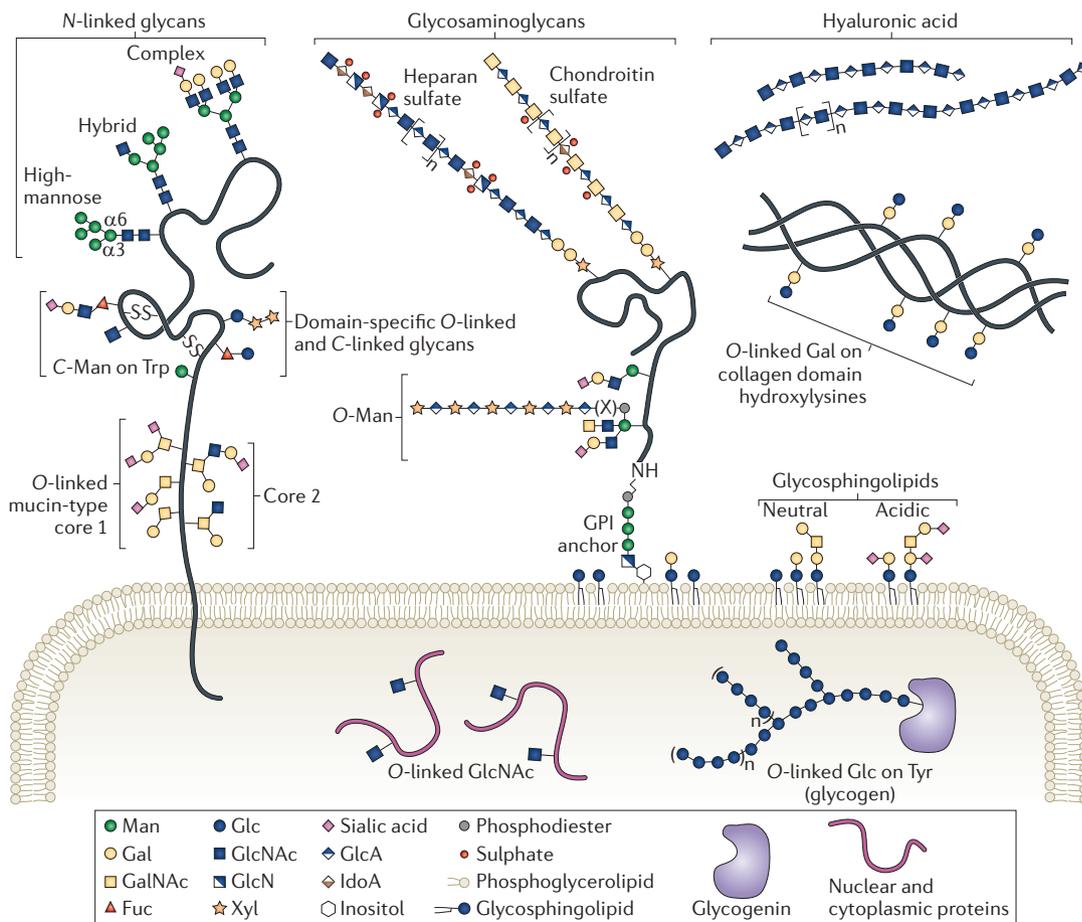
Most glycans on membrane and secreted proteins are found in *N*-linkage to Asn or in *O*-linkage to Ser/Thr. *N*-linked glycans that have undergone minimal mannosidase processing are called 'high-mannose' glycans. Addition of *N*-acetylglucosamine (GlcNAc) to the α3 arm of an acceptor containing five Man residues produces a hybrid glycan[14,18]. Hybrid glycans can be extended on the α3 arm with galactose (Gal), *N*-acetylgalactosamine (GalNAc), fucose (Fuc) and sialic acid (SA). Further processing of hybrid glycans produces complex glycans, which can also be decorated with Gal, GalNAc, Fuc and SA.

Glycosaminoglycans (GAGs) are *O*-linked glycans initiated by a conserved tetrasaccharide (GlcA-β1,3-Gal-β1,3-Gal-β1,4 Xyl-β) and classified by the composition of their disaccharide repeat. GAG chains can be post-synthetically modified by sulphation and epimerization (GlcA conversion to IdoA), producing substantial heterogeneity[140]. Glycoproteins carrying one or more GAG chains are called proteoglycans, and can be secreted or transmembrane or glycosylphosphatidiylinositol (GPI)-anchored. Hyaluronic acid, a GAG-like polymer of the extracellular matrix, is the only glycan that is not linked to a protein or lipid[141].

Other *O*-linked glycans are classified by their initiating monosaccharide. Addition of GalNAc to Ser/Thr initiates mucin-type *O*-linked glycans. Extension with Gal, GlcNAc or GalNAc produces eight different core structures[142]. For example, addition of a branching GlcNAc to a core 1 disaccharide (Gal-β1,3-GalNAc) produces core 2. Man initiates another class of *O*-linked glycan (*O*-Man glycans). A subset of *O*-Man glycans are extended with a repeating disaccharide — (-3-Xyl-α1,3-GlcA-β-)$_n$ — in phosphodiester linkage to an incompletely defined bridging moiety (X)[121,127]. Two types of *O*-linked Fuc glycan and one type of *O*-linked Glc glycan can be added to specific Cys-rich domains, as well as *C*-Man residues on Trp side chains[94,118]. *O*-linked GlcNAc is found on the extracellular domains of some proteins and on numerous cytosolic and nuclear proteins, but different enzymes mediate extracellular and nucleocytoplasmic *O*-GlcNAcylation[6]. *O*-linked glycan structures attached to other amino acids include Gal on hydroxylysine of collagen domains that is extended by the addition of an α1,2-Glc residue and the addition of a Glc on Tyr of glycogenin that is extended and branched with additional Glc residues to form glycogen[143,144]. In addition to proteins, sphingolipids can be modified by glycosylation. Eukaryotic cell surfaces are enriched in glycosphingolipids, which are ceramide-linked glycans that can be capped and branched with Fuc and SA[145].

cellular carbohydrate structures. This complexity commonly dissuades non-glycobiologists from exploring the functional information content embedded within these glycan modifications.

To facilitate broader exploration by non-glycobiologists of the functional information content embedded within glycan structures, this Review article focuses on four key areas of contemporary glycobiology

research that have a great impact within the cell biology community: the roles of glycans in quality control; the mechanisms of glycoprotein transport and modification through the Golgi complex; protein domain-specific glycosylation; and high-throughput analytical approaches for glycan structural profiling. The overall goals are to highlight the latest developments in the glycobiology field while emphasizing the unique information content and functions that glycans have in animal systems.

N-glycan synthesis and quality control

Transfer of glycan precursor to nascent polypeptides.

N-glycosylation is the most highly studied form of protein glycosylation in eukaryotic organisms. It has been estimated that approximately half of all human proteins are glycoproteins, and most of those contain N-glycan structures[17]. N-glycans are initially synthesized as a lipid-linked oligosaccharide (LLO) precursor, and the glycans are transferred from LLO to a nascent polypeptide chain during translation. Although the general details of co-translational glycan transfer have been known for decades[18], the mechanistic details of oligosaccharide transfer have only recently been revealed[19]. Eukaryotic organisms generally use a multi-subunit oligosaccharyltransferase (OST) on the lumenal face of the ER membrane to catalyse glycan transfer to acceptor peptide sequons[20], which are comprised of an Asn-X-(Ser/Thr) tripeptide (and less frequently of Asn-X-Cys[21] and other non-standard sequons[22,23]), where X can be any amino acid except for Pro (FIG. 1a; TABLE 1). One OST subunit, STT3, contains the catalytic site of the enzyme[24] and the mechanism of its interaction with the tripeptide sequon has long been hypothesized on the basis of theoretical and biochemical studies[25].

Recent work has identified a homologous form of N-glycosylation on similar acceptor sequons of periplasmic and secreted glycoproteins in a restricted set of Gram-negative bacteria[26–28]. Bacterial glycosylation also relies on the transfer of glycans from an LLO precursor to nascent polypeptide chains, but the enzyme responsible is a single polypeptide, PglB, with high sequence similarity to eukaryotic STT3 (REFS 27–29). The structure of *Campylobacter lari* PglB with a bound peptide acceptor was recently solved[19], and the co-complex provides interesting insights into the mechanistic details of sugar transfer for both bacteria and eukaryotes. PglB is anchored in the bacterial membrane via 13 transmembrane helical segments while a large globular domain faces into the periplasm. Two clefts are found at the junction between the transmembrane and globular domains, and the acceptor peptide was found to be positioned across one of the clefts with its Asn side chain extending through a hole into the adjacent cleft, which is presumed to bind the LLO donor. The nature of the interactions between the peptide acceptor and the active site cleft, the indirect catalytic role of a protein-bound divalent cation, and the conformational change in a loop region required for product release all provide novel insights into the mechanism of glycan transfer[25]. The additional subunits required for eukaryotic OST activity presumably act to facilitate coupling to the Sec61 protein translocation

pore and to select full-length Glc$_3$Man$_9$GlcNAc$_2$ LLO glycan substrates, as truncated glycans are transferred with greatly reduced efficiency in animal systems[20].

N-glycosylation in quality control of protein folding.

Following transfer of Glc$_3$Man$_9$GlcNAc$_2$ glycans to nascent polypeptide chains on the lumenal face of the ER in eukaryotes, trimming of two Glc residues results in the formation of a GlcMan$_9$GlcNAc$_2$ structure that acts as a ligand for the membrane-bound or soluble lectin chaperones, calnexin and calreticulin[30] (FIG. 1b). These lectins are tethered to ERp57 (also known as protein disulphide-isomerase A3 (PDIA3)), and they protect nascent polypeptides from hydrophobic aggregation and non-productive disulphide bonding during their folding itinerary[31]. Iterative cycles of Glc removal by glucosidase II and Glc re-addition by UDP-Glc:glycoprotein glucosyltransferase (UGGT1), followed by re-binding to the lectin chaperones, help to facilitate efficient folding of newly synthesized glycoproteins in the ER lumen[30,31] (FIG. 1c). Glucosidase II is comprised of an α (catalytic) and a β (lectin) subunit that facilitate substrate interactions, but neither glucosidase subunits nor the ER lectins calnexin and calreticulin appear to recognize the polypeptide backbone of the folding intermediate. However, UGGT1 clearly recognizes glycan structures in proximity to incompletely folded 'molten globule' peptide structures, and it acts as a folding sensor that drives re-entry into the 'calnexin cycle'[30,32].

Control of misfolded glycoprotein disposal from the ER.

In addition to recruiting chaperones during protein folding, glycan structures define the ER residence time for the downstream quality control of newly synthesized glycoproteins. Glycoproteins with slow folding kinetics owing to mutation or incomplete oligomeric assembly of subunits are targeted for degradation by the activity of Man trimming enzymes, which is the basis for the 'mannose timer' model of ER quality control[30]. Glycan trimming is followed by recognition of the trimmed structures by components of a multiprotein complex at the ER membrane that facilitates 'retro-translocation' and subsequent proteasomal degradation in a process termed ER-associated degradation (ERAD)[30,31]. Recent studies on ERAD in *Saccharomyces cerevisiae* demonstrated that two related ER mannosidases (Mns1 and Htm1 (also known as Mnl1)) are required for sequential trimming of terminal Man residues[33], with Htm1 working in a complex with Pdi1 (REF. 34). The resulting trimmed glycan is recognized by the ERAD lectin Yos9 (OS9 in humans along with its homologue XTP3B (also known as ERLEC1)), which mediates transfer to the retrotranslocon complex (FIG. 1d).

In mammals, the components involved in glycan trimming and recognition are more complex. Seven enzyme isoforms related to the yeast trimming mannosidases are found in mammalian cells, where they have roles in both ERAD and glycan maturation to complex type structures in the secretory pathway[35]. Mannosidase inhibitor and overexpression studies indicate that Man trimming is the proximal and rate-limiting step in ERAD and is tightly coupled to components of
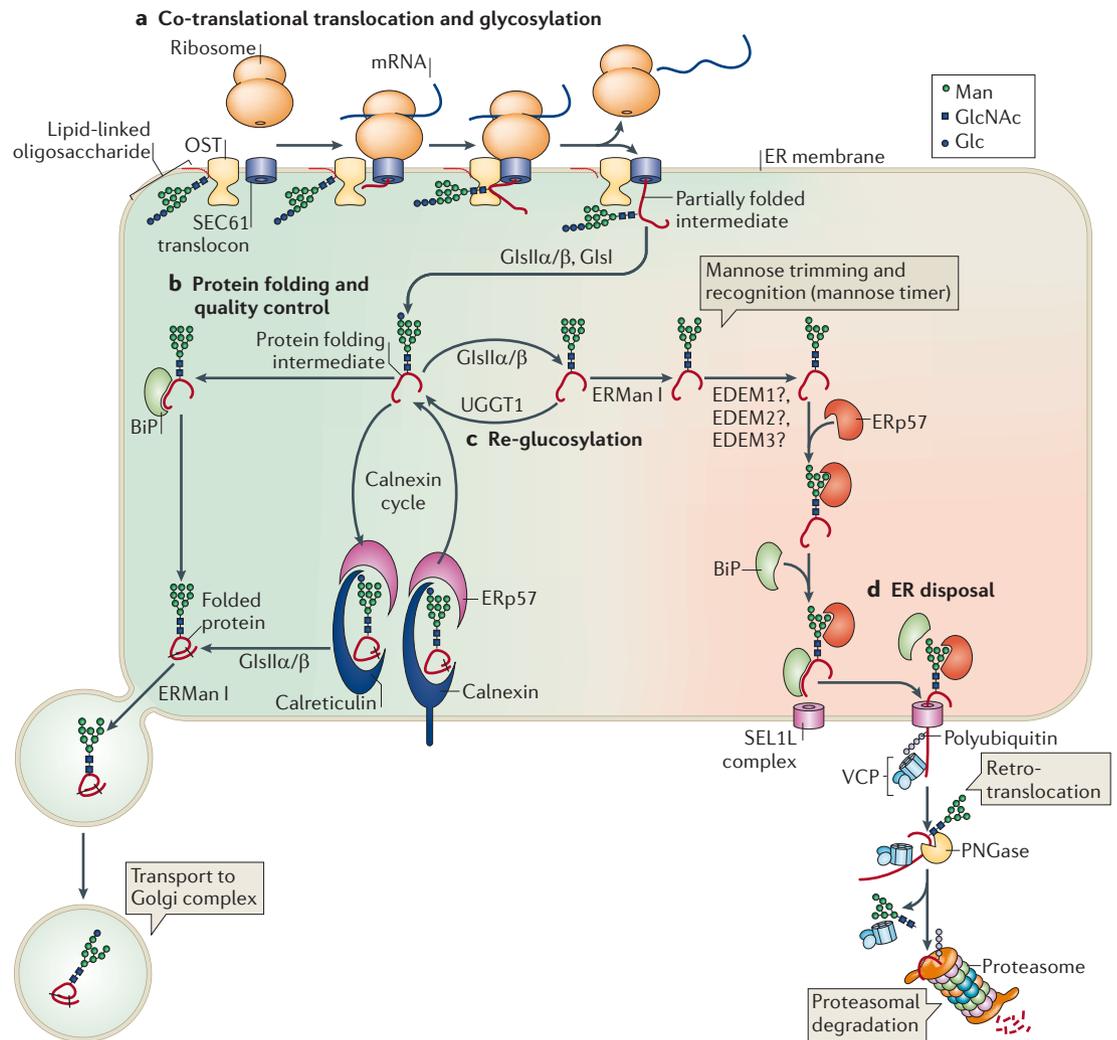

Lipid-linked oligosaccharide
(LLO). Extended carbohydrate structure attached through a phosphodiester linkage to a polyisoprenoid lipid, usually dolichol. *N*-linked glycan precursor structures are commonly assembled as a lipid-linked intermediate before transfer to a polypeptide side chain.

Oligosaccharyltransferase
(OST). A multi-enzyme complex (or single subunit in bacteria) in the endoplasmic reticulum membrane that transfers glycan structures from a lipid-linked oligosaccharide precursor to acceptor sequences on nascent polypeptides.

Acceptor peptide sequons
Short amino acid sequences on glycan-acceptor polypeptide chains that are recognized by glycosyltransferases prior to glycosylation.

Calnexin
Integral membrane endoplasmic reticulum lectin that recognizes GlcMan$_9$GlcNAc$_2$ glycans on early glycoprotein folding intermediates and, in collaboration with an associated thiol oxidoreductase, ERp57, aids in protein folding as a part of endoplasmic reticulum quality control.

Calreticulin
A soluble lectin in the lumen of the endoplasmic reticulum that contains a K-D-E-L endoplasmic reticulum retrieval sequence and, in a similar way to calnexin, binds to GlcMan$_9$GlcNAc$_2$ glycoprotein folding intermediates to facilitate protein folding and quality control.

ER-associated degradation
(ERAD). A protein quality control pathway in which misfolded lumenal or integral membrane proteins are recognized (often through trimmed glycan structures) for disposal by translocation into the cytosol for proteasomal degradation.

Figure 1 | Protein *N*-glycosylation and quality control of protein folding. **a** | During glycoprotein biosynthesis, the translation of nascent polypeptides is followed by their translocation through the SEC61 pore and the simultaneous transfer of a glycan from a lipid-linked intermediate to peptide acceptor sequons by the oligosaccharyltransferase (OST). One cleft in the STT3 subunit of OST scans for Asn-X-Ser/Thr acceptor sequons, while an adjacent cleft binds the glycan donor. **b** | Glycan trimming through Glc removal occurs immediately after transfer by α-glucosidase I (GlsI) and the α-glucosidase II α–β heterodimer (GlsIIα/β). Folding intermediates containing $Glc_1Man_9GlcNAc_2$ structures are ligands for the lectins calnexin or calreticulin, which function in complex with ERp57. Dissociation from the lectins is followed by further Glc cleavage. Additional chaperone assistance is provided by the ATP-driven chaperone BiP (also known as GRP78). Correctly folded glycoproteins are packaged for transport to the Golgi. **c** | Incompletely folded glycoproteins are recognized by the folding sensor UDP-Glc:glycoprotein glucosyltransferase (UGGT1). They are then re-glucosylated through the addition of a Glc residue back to the glycan structure and are reintegrated into the calnexin cycle. **d** | Terminally misfolded glycoproteins are subjected to endoplasmic reticulum (ER) disposal by Man trimming (through the activity of ER α-mannosidase I (ERManI) or GolgiManIA, GolgiManIB and GolgiManIC (not shown), followed by the activity of ER degradation-enhancing α-mannosidase-like 1 (EDEM1), EDEM2 and EDEM3 (which are homologues of Htm1 (also known as Mnl1) in yeast)). The trimmed glycans bind the ER lectins OS9 or XTP3B (not shown) and are translocated into the cytosol via a complex of derlin 1 (DER1), DER2 and DER3 (the SEL1L complex; Hrd3 in yeast) using the driving force of the cytosolic ubiquitin binding protein and ATPase functions of valocin-containing protein (VCP; also known as TER ATPase; known as Cdc48, Ufd1 or Npl4 in yeast). The peptide is deglycosylated by a cytosolic PNGase and degraded by the proteasome[149].

the calnexin cycle to assure effective quality control surveillance and clearance of misfolded proteins from the lumen of the ER[36]. However, the subcellular location of the ERAD compartment, the substrate specificities and the localizations of the trimming enzymes, as well as their relative contributions in ERAD versus glycan maturation, remain controversial[36,37].

Neither the trimming mannosidases nor the Yos9 homologues are believed to recognize the folding state of the underlying polypeptide, and only some glycans on ERAD substrates are required for effective disposal[38,39]. An intriguing model has recently been developed suggesting that some protein glycosylation sites mark secondary structure elements that are sensitive to the

Nocodozole
A pharmacological agent that blocks mitosis by binding to tubulin and inhibiting microtubule polymerization. Inhibition of microtubule polymerization also causes the dispersal of well-organized Golgi stacks into smaller units distributed throughout the cytoplasm.

Brefeldin A
A fungally derived antibiotic that inhibits the guanine nucleotide exchange factor (GEF) responsible for activating ARF1 GTPase. Activation of ARF1 recruits coatomer protein complex I (COPI) to Golgi membranes to form vesicles. In the absence of Golgi vesicle formation, *cis* and medial cisternae fuse with the endoplasmic reticulum and the *trans* and *trans*-Golgi network components disperse into a drug-induced entity called the brefeldin compartment.

Cisternal maturation
An early model of Golgi trafficking that proposed that Golgi cisternae formed from the fusion of vesicles leaving the endoplasmic reticulum and were subsequently matured by the import of appropriate processing enzymes as the cisternae were pushed forward towards the *trans* face. The original cisternal maturation model had its groundings in extensive electron microscopic observations of Golgi morphology.

Vesicular transport
A model for Golgi trafficking, which proposed that cisternae are stable structures and that cargo proteins move between cisternae in transport vesicles that are targeted to specific Golgi domains.

Rapid partitioning
A model for Golgi trafficking, which proposes that cargo proteins can exit the Golgi in vesicles arising from all cisternae and that new protein arriving at the Golgi rapidly gains access to the entire apparatus. Distinct transport and processing domains within each Golgi cistern are defined by their lipid content, which varies systematically from *cis*-to-*trans* and may provide favourable environments for cisternae-specific subsets of glycan processing enzymes.

overall folding state of the protein[38]. A bipartite signal that combines trimmed glycans and underlying protein conformation would allow the selection of ERAD substrates while sparing fully folded intermediates and ER-resident proteins from disposal. The identity of the ERAD components that couple the glycan and peptide recognition components still remains unclear[38]. Thus, protein glycosylation has key roles in assisting in protein folding as well as in defining the timing for protein disposal as a part of the quality control function in the early secretory pathway.

Golgi organization and glycomic diversity

Glycan processing in the post-ER secretory pathway.
After leaving the ER, glycoproteins traverse the Golgi apparatus, where their glycans are susceptible to further processing. Generally, few 'high-mannose' glycans reach the cell surface of most differentiated vertebrate cell types[40]. Almost all glycoprotein glycans are subject to trimming and extension as they transit the Golgi. Localization studies have demonstrated that glycan processing enzymes are distributed across the Golgi in a *cis*-to-*trans* topography that correlates successive glycan maturation steps with vectorial traffic through the organelle[41,42] (BOX 2). However, this generic Golgi architecture, where glycan processing enzymes serve as defining markers for specific Golgi subcompartments, is an oversimplification. Overlapping or altered enzyme distributions are common, and different cell types or organisms frequently present unique Golgi organizations[43–47]. Thus, Golgi heterogeneity, coupled with an incomplete understanding of the mechanisms that regulate membrane trafficking and sub-Golgi enzyme targeting, complicate the standard definitions of Golgi compartmentalization.

It is easy to envision that alterations in the glycan assembly line could affect glycan expression. However, it is not always possible to predict the precise effects of altered Golgi morphology on glycan production. Pharmacologic interventions that disperse the Golgi stack produce glycosylation changes that range from relatively minor glycan modifications (after low doses of nocodozole) to extreme under-processing of *N*-linked glycans (after brefeldin A administration)[48–53]. More subtle changes to the Golgi architecture — such as changes that cells might routinely use to modulate their glycan expression — have not been investigated in any significant detail in terms of their impact on the cellular glycome.

Current models of Golgi trafficking offer many intriguing candidates for proteins that may influence cellular glycan expression. Many excellent reviews have presented the evolution of evidence supporting and contradicting the three major models for Golgi trafficking — cisternal maturation, vesicular transport and rapid partitioning — and have also discussed several alternative models[54–61]. Each of these models makes specific predictions regarding the mechanisms by which glycan biosynthetic enzymes are brought into contact with their substrates (BOX 3). From a glycomic viewpoint, the most important consideration is whether these models can account for the diversity of cellular glycans and whether they accommodate the known biochemical and localization characteristics of the glycan processing machinery.
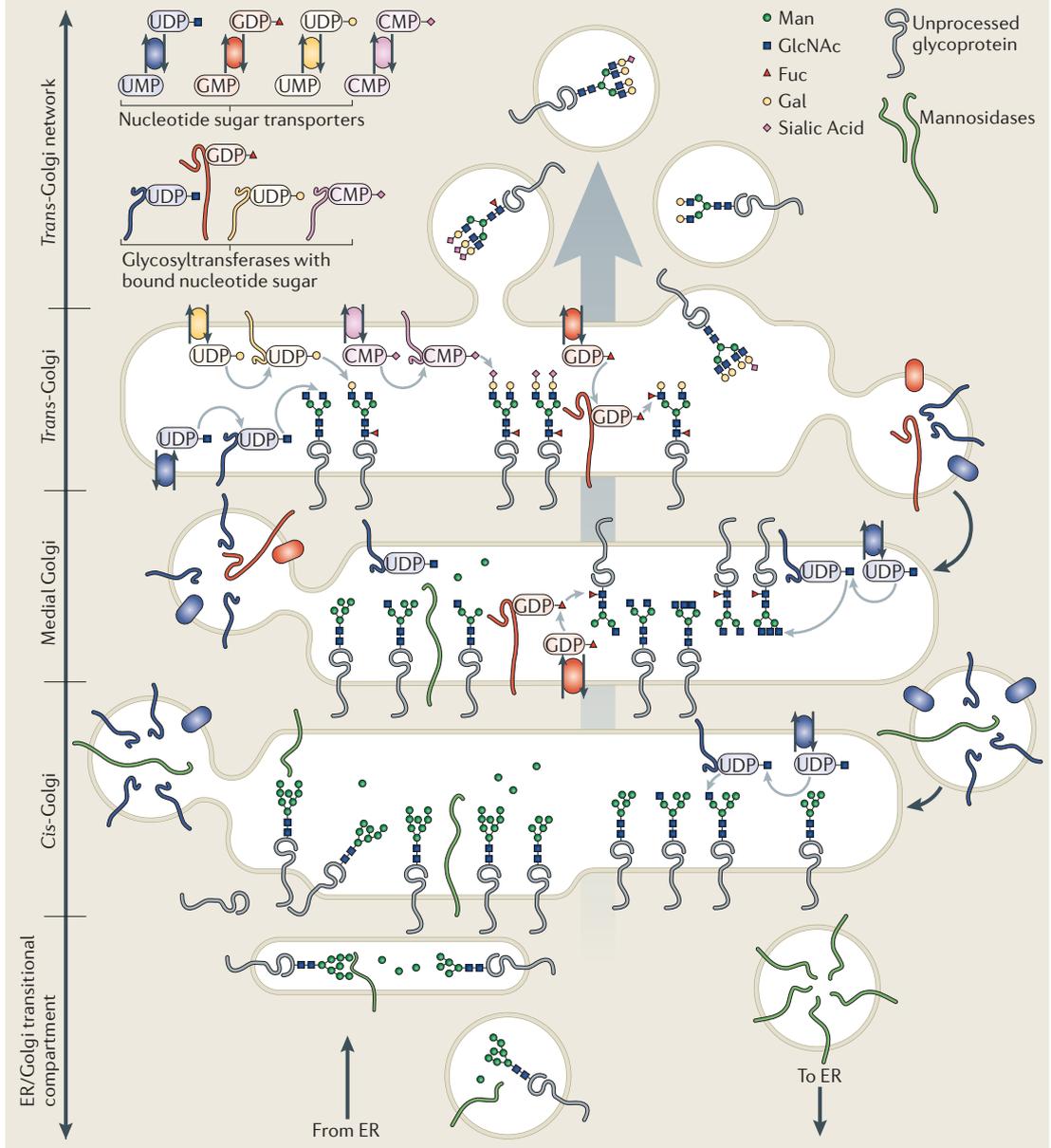
Compartmentalization of Golgi glycosylation enzymes.
Several mechanisms have been proposed for the targeting and capturing of specific glycosylation enzymes within Golgi sub-compartments that ensures orderly glycan processing. The biophysical characteristics of the transmembrane domains of some processing enzymes may stabilize their distribution within membrane domains that possess specific ratios of sphingolipids, glycosphingolipids and phosphoglycerolipids[62–64]. It is not currently clear whether sufficient resolution exists in the matching of transmembrane domains with lipid microenvironments to account for specific cisternal sorting or whether this matching serves primarily as a means of excluding enzymes from transport to the cell surface. Essentially all Golgi glycosyltransferases are type 2 transmembrane proteins with very small NH$_2$-terminal cytoplasmic domains. In yeast, a consensus sequence in the NH$_2$-terminal cytoplasmic tail that confers binding of vacuolar protein sorting 74 (Vps74) has been identified in a subset of glycosyltransferases; in turn, this sequence interacts with coatomer protein complex I (COPI), suggesting a mechanism for segregating transferases into transport vesicles[65,66]. Golgi phosphoprotein 3 (GOLPH3) has been identified as the animal homologue of Vps74, but the yeast consensus sequence does not appear in the cytoplasmic tails of any human glycosyltransferase[66], and little evidence supports a dominant role for the cytoplasmic tail in localization of animal glycosyltransferases. However, biochemical, mutagenesis and domain-swapping studies have demonstrated that amino acid sequences within the cytoplasmic, transmembrane and stalk domains of individual transferases, as well as disulphide bond formation between the lumenal domains of glycosyltransferase monomers, can contribute to targeting[67–72].

In the absence of a single clear retention or sorting mechanism for glycosyltransferases, it is reasonable to assume that these enzymes probably use a combination of multiple, weak sorting signals for localization[73,74]. The kin recognition and oligomerization model, which was originally based on co-immunoprecipitation and mislocalization studies, proposed that glycosyltransferases form homo-oligomers with each other and hetero-oligomers with functionally related enzymes[75]. As larger aggregates are formed, they reach a size at which they are excluded from transport vesicles, thus stabilizing their localization within a specific cisterna. In the context of the vesicular transport model, the aggregation hypothesis was useful. However, in the cisternal maturation and rapid diffusion models, localization by aggregation is problematic, as it does not allow for dynamic changes in cisternal enzyme content. Nonetheless, elegant recent studies using bi-molecular fluorescence complementation have demonstrated that specific transferases form homodimers and heterodimers in living cells[76]. The organization of transferases into biosynthetic complexes has also been described for *N*-linked glycan precursor biosynthesis, yeast mannan extension and

Glycosphingolipids
Glycoconjugates comprised of a ceramide lipid (Cer) carrying a glycan headgroup. The glycan is usually initiated by Glc, although GalCer is an important component of some cells. GlcCer is elongated to generate four major types of neutral cores in mammalian tissues (ganglio-, globo-, lacto- and neolacto-series), each of which can be capped and branched to produce additional structural diversity.

Type 2 transmembrane proteins
Single pass transmembrane proteins with their amino terminus oriented towards the cytosol and their carboxyl terminus facing the lumen of the secretory pathway or cell exterior.

Vacuolar protein sorting 74
(Vps74). A yeast protein identified as having a vacuolar protein sorting function. Yeast *vps74* mutants are deficient in intra-Golgi transport and exhibit mis-localized glycosyltransferases. Golgi phosphoprotein 3 (GOLPH3) is the mammalian homologue of Vps74.

Coatomer protein complex I
(COPI). A protein complex that is recruited to membranes by ARF (ADP-ribosylation factor) GTPases and that mediates intra-Golgi and Golgi-to-endoplasmic reticulum retrograde transport.

Mucin-type *O*-linked glycosylation
A form of protein glycosylation initiated by the addition of a GalNAc residue to protein Ser or Thr side chains and extended and branched with other complex termini. It is commonly found on highly glycosylated mucin glycoproteins, but also on many non-mucin polypeptides.

Box 2 | Distribution of glycan processing enzymes across a generic Golgi apparatus

Transitional endoplasmic reticulum (ER) and *cis*-Golgi compartments contain enzymes that initiate *O*-linked glycosylation and mannosidases for trimming high-mannose *N*-linked glycans. Medial compartments contain enzymes that branch *O*-linked glycans and that initiate complex *N*-linked glycan formation, branching and core fucosylation[9,74]. The *N*-linked glycans on glycoproteins that achieve this level of processing become resistant to removal by endoglycosidase H, and studies of trafficking frequently rely on this characteristic to report the transit of a protein through the Golgi. *Trans* compartments elaborate additional branching and capping reactions (galactosylation, sialylation, sulphation and external fucosylation) on complex *N*-linked and *O*-linked antennae[42]. Beyond the *trans*-Golgi, capping reactions are continued in the *trans*-Golgi network (TGN), a tubulovesicular compartment that stages glycoprotein cargoes for secretion or sorting to specific cell surface or subcellular membrane domains. Bulk flow of Golgi contents from *cis*-to-trans is counterbalanced by retrograde transport that returns resident Golgi proteins to appropriate cisterna[61]. Glycoprotein substrates carrying appropriate glycan acceptors, glycosyltransferases possessing appropriate enzymatic activities, and nucleotide sugar transporters of the appropriate specificities must all be brought together within the same functional unit of the Golgi apparatus in order to achieve glycan maturation.

glycosaminoglycan and glycosphingolipid processing, but not for mucin-type *O*-linked glycosylation[74,77–79]. Oligomeric organization is thought to confer increased efficiency, as the product of one enzyme is immediately accessible as the substrate for the next. Further work will be required to determine the extent to which glycosyltransferase localization and oligomerization are useful mechanisms for regulating cellular glycomic composition.

Box 3 | Models of Golgi trafficking in animal cells

The original cisternal maturation model proposed that glycoproteins exit the endoplasmic reticulum (ER) in vesicles or vesiculotubular structures that coalesce into *cis*-cisternae, which subsequently receive enzymes that function in the early stages of glycan processing. Cisternal maturation would then be achieved by the sequential delivery of enzymes responsible for successively terminal glycan modifications. Ongoing delivery of new ER cargo vesicles would drive forward the translocation of maturing cisternae, which would eventually bud transport vesicles at the *trans*-Golgi[146] (part **a** of the figure).

The vesicular transport model supposes the existence of stable Golgi cisternae that possess characteristic sets of glycan processing enzymes[58]. Glycoprotein cargos are moved by targeted anterograde vesicular transport from early to late cisternae (part **b** of the figure). However, the vesicular transport model is inconsistent with the enrichment of glycosyltransferases in transport vesicles and with the trafficking of macromolecular complexes that are too large to fit within vesicles[54,60]. Subsequent reassessments of the cisternal maturation model introduced retrograde transport of glycan processing enzymes as a mechanism to retrieve processing capacity into appropriate cisternae. Both the cisternal maturation and

vesicular transport models share the assumption that transported proteins progress through the Golgi as a wavefront. However, photobleaching measurements of glycoprotein progression through the Golgi in live cells have recently challenged this conveyer belt assumption by demonstrating that product disappears uniformly from all cisternae and that new protein arriving at the Golgi quickly accesses all cisternae[63] (part **c** of the figure). These results have led to the formulation of the rapid partitioning model, which proposes a major role for Golgi lipid composition in establishing and regulating the polarity of the organelle and the distribution of its resident proteins[147,148]. The rapid partitioning model proposes the coexistence of two distinct lipid domains in each stack of the Golgi, one enriched in sphingolipid and the other enriched in phosphoglycerolipid. The phosphoglycerolipid domains carry out transport functions. The sphingolipid domains are home to Golgi resident proteins, such as glycan processing enzymes that may partition into these domains based on their sphingolipid composition[62,64]. Glycoprotein processing and transport occur as substrate proteins partition between these two domains. Although the central tenants of the major models are mutually exclusive, many of the proposed mechanisms are not, and the most useful conceptualization of Golgi trafficking allows for the simultaneous existence of aspects of each model (part **d** of the figure).
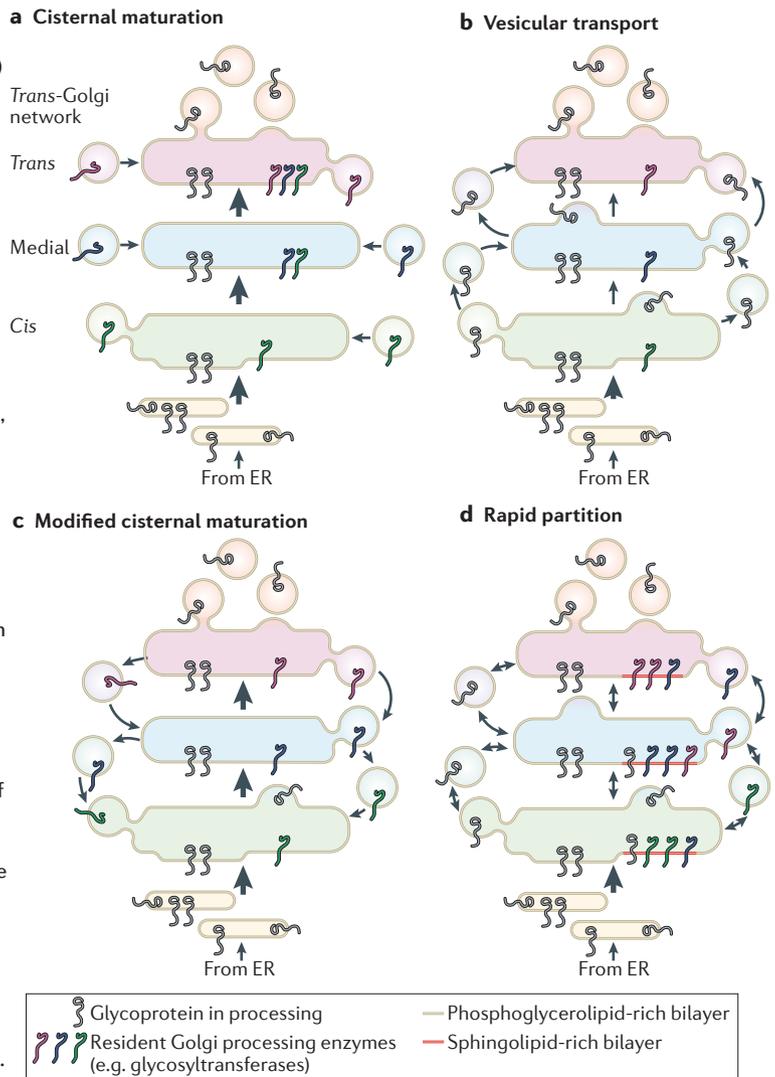
Golgins

A family of Golgi tethering factors characterized by their extended coiled-coil domains. Individual family members exhibit specific sub-Golgi localizations, GTPase interactions and effector activities. The golgins form long filaments that emanate from the cytoplasmic face of Golgi cisternae, providing access to transport vesicles in the near environment. Some golgins possess multiple binding sites for RAB proteins, suggesting that they capture transport vesicles carrying specific RAB proteins.

Conserved oligomeric Golgi complex

(COG complex). A multiprotein tethering factor comprised of eight protein subunits. The complex facilitates Golgi retrograde transport and glyco-syltransferase localization.

Golgi trafficking sculpts glycan profiles. Extensive testing of the three major models of Golgi organizational dynamics has driven substantial progress towards identifying molecular components of the Golgi trafficking machinery. Key proteins involved in intra-Golgi trafficking include vesicle COPI, tethering factors of the Golgi matrix (golgins and multiprotein complexes), small GTPases (the RAB family, ADP-ribosylation factor (ARF) and ARF-like GTPases,

and associated GTPase activating proteins (GAPs) and guanine exchange factors (GEFs)) and the membrane fusion machinery (NSF (*N*-ethylmaleimide-sensitive factor), SNAP (soluble NSF attachment protein) and vesicle- and target-SNAREs (SNAP receptors)) (FIG. 2a–c). Among these components, the conserved oligomeric Golgi complex (COG complex) has been directly implicated in the regulation of cellular glycan profiles (FIG. 2d).

The COG complex is composed of eight highly conserved protein subunits. To date, mutations in three human COG subunits (COG1, COG7 and COG8) have been identified, and each causes altered protein glycosylation and severe multi-systemic developmental defects, forming a subgroup of a growing class of inherited diseases known as congenital disorders of glycosylation (CDGs)[80]. Fibroblasts from these patients with CDG have proven to be useful experimental tools, shedding considerable light on the organization and function of the COG complex[61,81–84]. Knockdown of COG components in wild-type cells identified a set of proteins comprised of Golgi glycosylation enzymes and other related proteins (Golgi mannosidase II, four glycosyltransferases, the Golgi SNARE protein syntaxin 5 and several golgins). These proteins are collectively termed GEARs, and their stability and localization requires an intact COG complex[10,61,85]. Experimentally validated interactions between these GEARs and COG subunits support the hypothesis that Golgi vesicle tethering may occur in two stages: initial capture by golgins followed by COG-mediated stabilization that enhances vesicle association with the fusion machinery[61]. Thus, the COG complex may drive retrograde transport steps to completion and may also provide an opportunity for a final quality check before the commitment to fusion. Reduced function of the COG complex would be predicted to decrease retrograde transport, leading to mislocalization and degradation of the GEARs that mediate glycan processing.

The major models for Golgi trafficking universally require a concerted process for retention and partitioning of glycosylation enzymes across the Golgi stacks in the face of considerable anterograde flow. The glycan structures delivered to the cell surface should reflect the length of time necessary for maturation within a given cisterna, which, in turn, should be coupled to anterograde and retrograde transport efficiency. However, there is currently no reason to suspect that these transport events occur at constant rates across the organelle. Differential control of intra-organellar transport may provide cells with regulatory points for expanding, contracting or otherwise shifting their glycome. These regulatory points may include COPI coat dynamics, COG complex availability, RAB protein cycling, golgin expression and lipid biogenesis (FIG. 2a–g). Furthermore, a handful of paradigms have been described in which cellular responses to external stimuli affect Golgi form and function, presumably by altering trafficking. For example, administration of epithelial growth factor (EGF) and platelet-derived growth factor (PDGF) to HeLa cells activates a Golgi-localized SRC kinase that induces a redistribution of the enzymes responsible for initiating O-linked glycosylation from the cis-Golgi to the ER, where they drive increased mucin glycosylation[86]. Other Tyr and Ser/Thr kinases have recently been implicated as regulators of Golgi morphology and, in some cases, protein glycosylation[87–90]. These emerging paradigms, in which growth factors, cytokines, chemokines and other signalling molecules stimulate changes in Golgi dynamics, promise to reveal stronger mechanistic links between cell signalling and glycomic diversity.

Protein domain-specific glycosylation

Ubiquitous forms of N- and mucin-type peptide-O-GalNAc modifications have been extensively studied for decades. In addition, the selective modification of N-glycans on restricted families of proteins has been recognized for many years (for example, selective poly-sialylation of N-glycans on neural cell adhesion molecule 1 (NCAM1)[91] and a handful of other proteins, or GlcNAc-1-PO_4 addition to terminal Man residues on lysosomal enzymes[5]). Unconventional glycan additions have also been described for a small number of proteins[21–23]. An emerging field of glycobiology has recently focused on the selective glycosylation of distinct protein domains with specific classes of glycan structures (TABLE 1; BOX 1). These novel forms of glycosylation have been shown to have crucial roles in modulating biological recognition events in development and physiology.

The first of these domain-specific modifications was identified on the Cys-rich EGF domains of the cell surface transmembrane signalling proteins of the Notch family, where the glycans were found to modulate Notch–ligand interactions in developmental cell fate decisions[92]. Subsequent studies identified other protein domain-specific modifications, including modifications of thrombospondin type 1 repeats (TSRs) and other selective protein modifications such as those found on the mucin-type domains of α-dystroglycan[93–97] (TABLE 1; BOX 1). In each case, initial peptide modification occurs in the ER, but further extension of the glycan structure continues in the Golgi complex.

EGF domain-specific glycosylation. EGF repeats are Cys-rich domains of ~30–40 amino acids that are found in many cell surface and secreted proteins and are frequently involved in protein–protein interactions[98]. The Notch family of single-pass, transmembrane receptors has served as a paradigm for studying glycosylation of EGF repeats, but modification of other EGF domain-containing proteins has not been well studied. In contrast with classical O-GalNAc glycosylation, Ser and Thr residues between the second and third Cys residues within EGF domains are modified by O-Fuc addition by the enzyme protein O-fucosyltransferase 1 (POFUT1; also known as OFUT in *Drosophila melanogaster*)[99,100] (TABLE 1). The Fuc residue can be extended by additional glycosyltransferases to form the tetrasaccharide SA-α2,6-Gal-β1,4-GlcNAc-β1,3-Fuc-α1-O-Ser. *Pofut1* is essential in mice[101], as is *Ofut1* in *D. melanogaster*[102], and conditional knockouts of *Pofut1* in adult mice revealed roles of O-Fuc modification in haematopoietic homeostasis[103]. The Fringe family of β3-GlcNAc transferases initiate elongation of glycans on O-Fuc residues and are modulators of Notch activity[92].

EGF repeats of Notch have also been found to contain modifications with O-Glc residues added to Ser side chains before the second Cys residue of the EGF repeat by the action of protein O-glucosyltransferase 1 (POGLUT1; also known as O-glucosyltransferase Rumi in *D. melanogaster*)[104] (TABLE 1). This glycan can also be extended by the addition of two Xyl residues[105,106]. Most of the O-Glc consensus sites in mouse NOTCH1 appear to be occupied, but variability exists among cell types
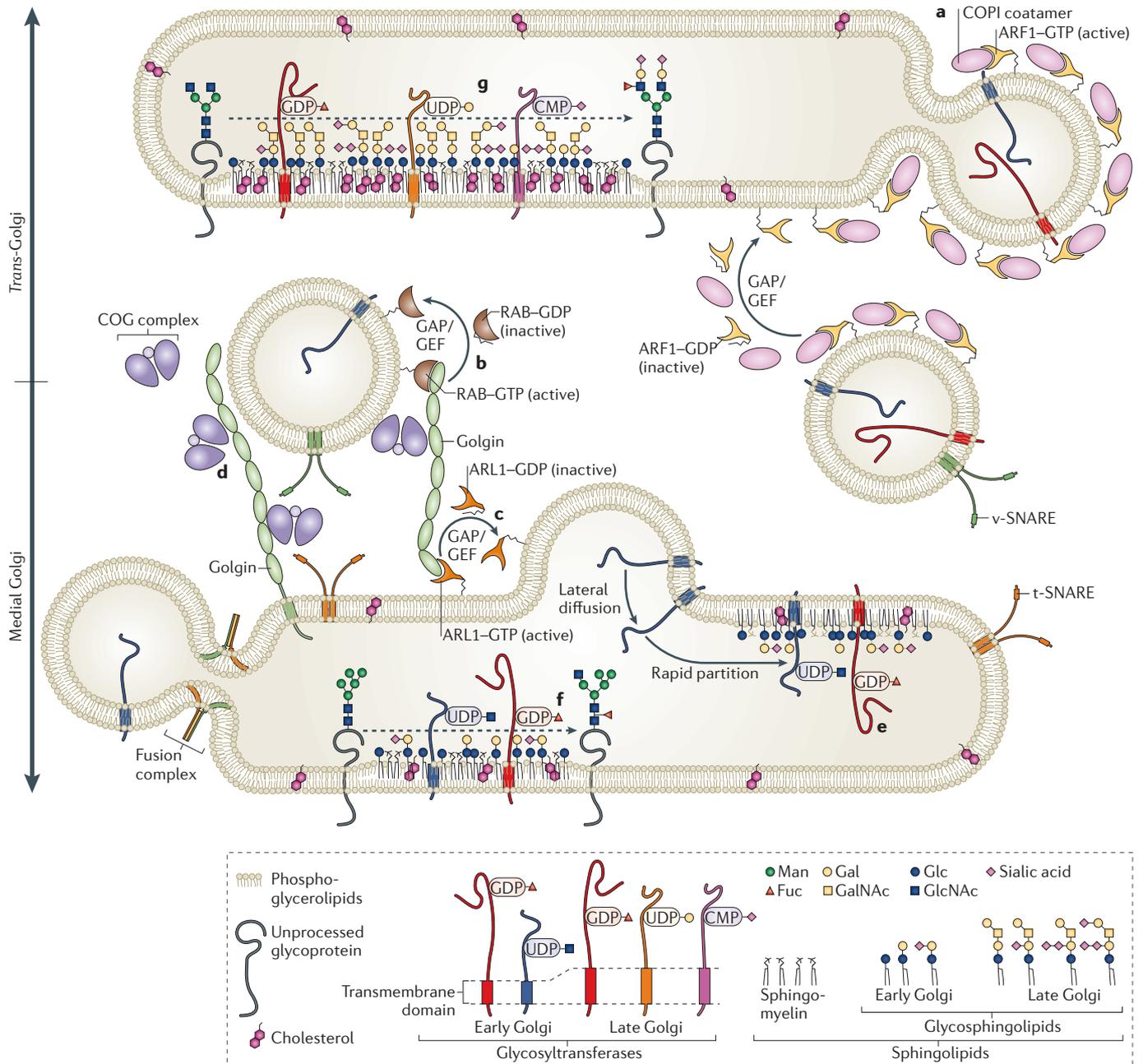
Figure 2 | **Control points for Golgi trafficking and cellular glycomic diversity.** Intra-Golgi transport mechanisms provide regulatory points for glycome modulation. Two idealized Golgi compartments are depicted. Guanine nucleotide exchange cycles acting on small GTPases regulate trafficking steps important for glycosylation. **a** | ADP-ribosylation factor 1 (ARF1) activation recruits coatomer protein complex I (COPI) coats to nascent transport vesicles. **b,c** | Golgin interactions with cisternal and vesicular GTPases influence vesicle capture[150,151]. At least 15 small GTPases of the RAB family and three of the ARF and ARF-like (ARL) family are associated with specific cisternae, vesicles and tethering factors[56,152]. Golgin tethering factors exist as transmembrane or peripheral membrane proteins that are retained at Golgi cisternae through binding interactions with myristoylated Golgi reassembly-stacking protein of 55 kDa (GRASP55; also known as GRS2) or GRASP65 (also known as GRS1) or small GTPases (prenylated RABs, acylated ARF1 or ARL1). **d** | The conserved oligomeric Golgi (COG) complex, a multiprotein-tethering complex, interacts with a subset of proteins called GEARs (including golgins), facilitating the fusion of appropriately tethered transport vesicles and modulating the stability of processing enzymes. **e,f,g** | Sphingolipid and cholesterol content, as well as glycosphingolipid complexity (that is, glycan size and charge) increase from early to late Golgi[148,153]. Glycosyltransferases catalysing sequential reaction sets may congregate into selective Golgi processing domains on the basis of preferences for specific lipid content imposed by the biophysical characteristics of their transmembrane domains. Therefore, altered lipid biosynthesis may affect glycosylation by contracting or expanding facilitative lipid domains within Golgi cisternae. t-SNARE, target soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor; v-SNARE, vesicle SNARE.

and depending on motif locations within the protein[107]. *Poglut1*-knockout mice do not survive past day 9.5, whereas heterozygous *Poglut1*[+/−] mice develop normally, suggesting an essential role for *O*-glucosylation in embryonic development[108].

A third type of novel glycosylation of EGF repeats was recently described in *D. melanogaster*, where a single *O*-GlcNAc residue was found to be added to a Ser/Thr residue following the fifth Cys residue in an EGF domain of the extracellular matrix protein Dumpy[109] and some EGF domains of NOTCH1 (REF. 110) (TABLE 1). Extension of the *O*-GlcNAc residue has not been detected. Previous studies had identified *O*-GlcNAc linkages exclusively on a large collection of cytosolic and nuclear proteins[6], but the detection of EGF domain-specific *O*-GlcNAc modification was the first example of such a linkage on an extracellular protein[109]. *O*-GlcNAc addition to extracellular proteins is catalysed through the action of a recently identified protein called EGF domain-specific *O*-linked GlcNAc transferase (EOGT)[109]. *D. melanogaster Eogt* mutants exhibit aberrant cell–matrix interactions and are embryonic lethal in a Notch-independent manner[109]. Mammalian EOGT orthologues have also been identified. In summary, EGF domains can act as acceptors of at least three unique types of glycosylation that all have crucial roles in metazoan development.

TSR domain-specific glycosylation. Similar to EGF repeats, TSR domains are found in numerous cell surface and secreted proteins as tandem repeats of 50–60 amino acids with disulphide bonds formed by six conserved Cys residues. Two forms of domain-specific glycosylation are found in TSRs that are distinct from those found in EGF domains (TABLE 1). Similar to EGF domains, *O*-linked Fuc can be added to Ser/Thr residues of TSRs by the action of POFUT2 (REFS 111,112), which is a distantly related homologue of the EGF repeat-modifying POFUT1 protein[113], and the *O*-Fuc is extended by the addition of a Glc residue through the enzymatic action of β-3-glycosyltransferase-like (B3GALTL)[114,115]. RNA interference knockdown of *Pofut2* transcripts or mutagenesis of the acceptor sequence in a model TSR-containing protein resulted in reduced secretion, suggesting a functional role for this modification[116]. Recently, mouse knockouts in the *Pofut2* locus revealed several developmental defects in epithelial to mesenchymal transition (EMT), as well as defects in cell localization and patterning[117].

The second type of unique TSR glycosylation is the addition of Man to a Trp residue through a C-C bond (*C*-mannosylation)[111]. The *C*-Man residue is added to the TSR co-translationally, using dolichol-phosphate-Man as the sugar donor, before protein folding[118,119]. Mutagenesis to replace Trp residues within a model TSR-containing protein suggests a role for *C*-mannosylation in the control of protein secretion[120], but the mannosyltransferase that creates this structure has yet to be identified.

Protein-selective glycosylation of dystroglycan. Finally, a distinct class of protein domain-specific glycosylation that contributes to cell–matrix interactions has been found predominantly on the cell surface protein α-dystroglycan (αDG), although other substrates may exist. Dystroglycan consists of two subunits, α and β, which are part of the membrane-spanning dystrophin glycoprotein complex that links the extracellular matrix with the cellular cytoskeleton. The α subunit is comprised of globular N- and C-terminal domains containing *N*-glycosylation sites and an extended central domain rich in classical *O*-GalNAc mucin-like structures, as well as glycans initiated with an *O*-Man linkage[121] that are essential for binding of dystroglycan with its ligands (laminin, agrin, perlecan and neurexin) or to viruses[121]. Defects in *O*-Man addition and extension on αDG lead to several forms of congenital muscular dystrophy collectively known as dystroglycanopathies[95].

The initial synthesis of the *O*-Man linkage is catalysed by protein *O*-mannosyltransferase 1 (POMT1) and POMT2, and the structure is extended by protein *O*-linked-Man β-1,2-*N*-acetylglucosaminyltransferase 1 (POMGNT1), which adds a β1,2-GlcNAc onto the *O*-Man core (TABLE 1). The most predominant *O*-Man structures of rabbit skeletal muscle and mouse brain αDG are linear tetrasaccharides containing terminal Gal and SA residues[122,123], and defects in *Pomt1*, *Pomt2* or *Pomgnt1* cause loss of laminin binding and subsequent dystroglycanopathies. However, the linear tetrasaccharide structure is not necessary for αDG ligand binding[124].

The products of several additional genes, including *LARGE1*, *LARGE2* (also known as *GYLTL1B*), fukutin (*FKTN*) and fukutin-related protein (*FKRP*), contribute to *O*-Man glycan modifications on αDG. Defects in these genes result in underglycosylation of αDG, disrupted ligand interactions, reduced stability of muscle fibres, disrupted basement membrane, altered neuronal migration and characteristic dystroglycanopathies[121,125]. However, the exact functions of these gene products remain unclear. Several lines of evidence indicate that functional *O*-Man structures are comprised of large, branched glycans that have resisted complete structural characterization (BOX 1). Mass spectrometry and NMR have detected a phosphodiester linkage to an unknown substituent at the 6-hydroxyl position of the core *O*-Man residue on αDG[126], but the identity of the responsible phosphotransferase or kinase is currently unknown. In addition, LARGE1 has recently been shown to contain two glycosyltransferase domains that catalyse the extension of a novel (-3-Xyl-α1,3-GlcA-β1-) homopolymer *in vitro*[127]. Although the nature of the bridging moiety that might link the proteoglycan-like homopolymer and the *O*-Man core is still unclear, the size of this product of the glycosyltransferase activity of LARGE1 is consistent with the high molecular weight of the laminin binding forms of αDG isolated from natural sources. Recent work with αDG-null cells and tissues also revealed that there is an approximately equal abundance of *O*-Man oligosaccharides in wild-type and αDG-null animals[122], and overexpression of LARGE1 in these cells was shown to restore laminin binding[128]. Therefore, other proteins besides αDG also carry *O*-Man structures, providing additional targets for understanding the role of LARGE1 in laminin binding.

Dystroglycanopathies
A clinically heterogeneous collection of muscular dystrophies that are associated with aberrant glycosylation of α-dystroglycan, which is a component of the multiprotein dystrophin glycoprotein complex that bridges the extracellular matrix, plasma membrane and cytoskeleton.

Protein domain-specific glycosylation is an emerging field of study with great potential for elucidating important biological recognition functions. As many glycosyltransferase-related coding regions in animal systems remain uncharacterized, we anticipate that additional classes of protein glycosylation will be characterized in the future that have key regulatory roles in biological systems.

Glycan profiling and global regulation

The diversity of glycan structures in animal systems presents a huge analytical challenge for determining detailed glycan structural profiles in complex biological systems[16]. As glycan structures are commonly altered in response to biological signalling pathways, there is a major effort to develop broad analytical strategies that identify glycan signatures for cellular transitions[129] or biomarkers for disease states[130,131]. In addition, detailed snapshots of glycan structural diversity can provide insights into potential biological recognition events and regulatory mechanisms for glycan biosynthesis[12]. Limitations in structural analysis depend on the glycan class, the complexities of the sample workup and the analytical tools that are available for structural inquiry[132].

Numerous approaches are now available for highly sensitive, broad *N*-glycan analysis[16]. These approaches include mass spectrometry analysis directly after enzymatic release of the glycan structures; methods that employ front-end chromatographic separations before mass spectrometry analysis; methods that employ chromatographic separation and *N*-glycan identification on the basis of known elution positions of standards; and the analysis of glycan structures on the basis of lectin or antibody binding. Similar approaches are also available for *O*-glycan and glycolipid analyses, but chemical release rather than enzymatic release methods are generally used before analysis[133]. Recent advances have also provided direct analysis of intact glycolipid structures in complex samples. However, few methods have been developed for low sample consumption and for high-throughput analysis that can tease apart complex mixtures of structural isomers. Moreover, although numerous data sets are now available for glycan structures from various mammalian cell and tissue sources, challenges still exist in linking glycan structures to the corresponding peptide backbones.

A complementary strategy for glycan structural analysis is the use of lectin- and antibody-based microarrays[134]. In this analytical format, target proteins are captured with protein-specific antibodies and then their glycan structures are analysed using lectins or antibodies. This strategy has a particular benefit, as it enables high-throughput diagnostic profiling of glycans on distinct protein backbones. Multiple platforms for capture of target glycoproteins and glycan detection are available[135,136]. However, the benefits of this high-throughput approach are tempered by the limited ability of lectins and antibodies to distinguish fine structural details on glycans.

The availability of high-throughput platforms for glycan profiling has provided unique insights into regulated glycan expression in vertebrate organisms and in human disease. Numerous efforts are underway to identify glycans or glycoproteins as cancer biomarkers[131,137]. Profiling of a broad range of glycan structures in a high-throughput platform is also being performed in combination with human genome-wide association studies to examine the regulation of glycan expression in human populations[138,139]. Correlations between glycan profiles and gene expression data are also emerging as an approach to assess mechanisms of glycan regulation in animal systems[11]; in these studies, complex sets of glycan and transcript data are being mined to examine the relative contribution of transcriptional and post-transcriptional control of glycosyltransferases[12]. Thus, the dynamic changes in glycomic diversity are slowly yielding to structural analysis, providing opportunities to delineate the fine details of glycan structural changes that occur in complex biological systems.

Conclusions

Contemporary studies of protein glycosylation are revealing numerous examples where these post-translational modifications have essential roles in biological recognition events. Glycans have crucial functions throughout the cell, from the cytosol and secretory compartments to the cell surface and the extracellular space. Conserved contributions of *N*-glycan structures in chaperone interactions and protein quality control are initiated co-translationally, and glycan processing continues throughout the dynamic collection of secretory pathway compartments that lead to the cell surface. The complex assortment of glycosylation enzymes comprises an intricate assembly line for glycan maturation from the ER through the Golgi. Notably, the localization, dynamics, interactions and regulation of glycosylation enzymes within these compartments, as well as how glycosylation enzymes compete for the same substrates, remain active areas of study. Of the many roles that glycans have at the cell surface, emerging paradigms have highlighted the importance of protein domain-specific glycosylation in facilitating or modulating biological recognition events.

The diversity of glycan structures clearly provides an additional level of information content in biological systems, but the challenge for the future lies in identifying the critical contexts in which glycan functions contribute to biological regulation within the bewildering array of heterogeneous glycan structures. An essential requirement for future glycosylation studies is the ability to obtain detailed profiles of glycan diversity in various biological contexts. Recent advances in analytical strategies are beginning to provide the necessary breadth, depth and sensitivity of analysis to define the full spectrum of glycan complexity. Continued high-throughput adaptations of these approaches will generate structural data sets that will be essential for revealing the mechanisms that regulate glycan biosynthetic pathways, defining unique glycan signatures for disease states and providing correlations with biological functions, thereby increasing our ability to decode the many functions of glycoprotein glycans in complex biological systems.

REVIEWS

1. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
2. Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73**, 491–537 (2004).
3. Dennis, J. W., Lau, K. S., Demetriou, M. & Nabi, I. R. Adaptive regulation at the cell surface by N-glycosylation. *Traffic* **10**, 1569–1578 (2009).
4. Hoseki, J., Ushioda, R. & Nagata, K. Mechanism and components of endoplasmic reticulum-associated degradation. *J. Biochem.* **147**, 19–25 (2010).
5. Kollmann, K. *et al.* Mannose phosphorylation in health and disease. *Eur. J. Cell Biol.* **89**, 117–123 (2010).
6. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).
7. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104 (2009).
8. Rothman, J. E. & Fine, R. E. Coated vesicles transport newly synthesized membrane glycoproteins from endoplasmic reticulum to plasma membrane in two successive stages. *Proc. Natl Acad. Sci. USA* **77**, 780–784 (1980).
9. Roth, J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **102**, 285–303 (2002).
10. Peanne, R. *et al.* Differential effects of lobe A and lobe B of the Conserved Oligomeric Golgi complex on the stability of β1,4-galactosyltransferase 1 and α2,6-sialyltransferase 1. *Glycobiology* **21**, 864–876 (2011).
11. Nairn, A. V. & Moremen, K. W. in *Handbook of Glycomics* (eds Cummings, R. & Pierce, J.M.) 95–136 (Academic Press, 2009).
12. Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J. Biol. Chem.* **283**, 17298–17313 (2008). **Presents transcript analysis paired with glycan structural data to identify correlations between glycan-related gene expression and glycan composition in mouse tissues.**
13. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–56R (2002).
14. Schachter, H. The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj. J.* **17**, 465–483 (2000).
15. Schachter, H. The 'yellow brick road' to branched complex N-glycans. *Glycobiology* **1**, 453–461 (1991).
16. Zaia, J. Mass spectrometry and glycomics. *Omics: J. Integrative Biol.* **14**, 401–418 (2010).
17. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
18. Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664 (1985).
19. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011). **Reports the first structure of an OST in complex with a peptide substrate and describes a proposed mechanism that is likely to extend to the eukaryotic enzyme.**
20. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R–62R (2006).
21. Zielinska, D. F., Gnad, F., Wisniewski, J. R. & Mann, M. Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907 (2010). **Describes the use of lectin-affinity enrichment of glycopeptides followed by proteomics analysis to identify 6,367 mouse glycosylation sites, including several classes of novel acceptor sequons.**
22. Schreiner, R., Schnabel, E. & Wieland, F. Novel N-glycosylation in eukaryotes: laminin contains the linkage unit β-glucosylasparagine. *J. Cell Biol.* **124**, 1071–1081 (1994).
23. Valliere-Douglass, J. F. *et al.* Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J. Biol. Chem.* **285**, 16012–16022 (2010).
24. Kelleher, D. J., Karaoglu, D., Mandon, E. C. & Gilmore, R. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell* **12**, 101–111 (2003).
25. Imperiali, B. & Hendrickson, T. L. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **3**, 1565–1578 (1995).
26. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50**, 4411–4426 (2011).
27. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21**, 576–582 (2011).
28. Weerapana, E. & Imperiali, B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R–101R (2006).
29. Mohorko, E., Glockshuber, R. & Aebi, M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.* **34**, 869–878 (2011).
30. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
31. Lederkremer, G. Z. Glycoprotein folding, quality control and ER-associated degradation. *Curr. Opin. Struct. Biol.* **19**, 515–523 (2009).
32. D'Alessio, C., Caramelo, J. J. & Parodi, A. J. UDP-GIC:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell Dev. Biol.* **21**, 491–499 (2010).
33. Clerc, S. *et al.* Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *J. Cell Biol.* **184**, 159–172 (2009).
34. Gauss, R., Kanehara, K., Carvalho, P., Ng, D. T. & Aebi, M. A complex of pdi1p and the mannosidase htm1p initiates clearance of unfolded glycoproteins from the endoplasmic reticulum. *Mol. Cell* **42**, 782–793 (2011).
35. Mast, S. W. & Moremen, K. W. Family 47 α-mannosidases in N-glycan processing. *Methods Enzymol.* **415**, 31–46 (2006).
36. Moremen, K. W. & Molinari, M. N-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control. *Curr. Opin. Struct. Biol.* **16**, 592–599 (2006).
37. Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* **35**, 74–82 (2010).
38. Xie, W., Kanehara, K., Sayeed, A. & Ng, D. T. Intrinsic conformational determinants signal protein misfolding to the Hrd1/Htm1 endoplasmic reticulum-associated degradation system. *Mol. Biol. Cell* **20**, 3317–3329 (2009). **Describes the roles of key N-linked glycan sites that mark structural determinants that are sensitive to the overall folding state of the molecule and mediate targeting of misfolded proteins for ER-associated degradation.**
39. Kanehara, K., Xie, W. & Ng, D. T. Modularity of the Hrd1 ERAD complex underlies its diverse client range. *J. Cell Biol.* **188**, 707–716 (2010).
40. An, H. J. *et al.* Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. *Mol. Cell. Proteomics* **11**, M111.010280 (2012).
41. Rabouille, C. *et al.* Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *J. Cell Sci.* **108**, 1617–1627 (1995).
42. Stanley, P. Golgi glycosylation. *Cold Spring Harb. Perspect. Biol.* **3**, a005199 (2011).
43. Mogelsvang, S., Marsh, B. J., Ladinsky, M. S. & Howell, K. E. Predicting function from structure: 3D structure studies of the mammalian Golgi complex. *Traffic* **5**, 338–345 (2004).
44. Papanikou, E. & Glick, B. S. The yeast Golgi apparatus: insights and mysteries. *FEBS Lett.* **583**, 3746–3751 (2009).
45. Ripoche, J., Link, B., Yucel, J. K., Tokuyasu, K. & Malhotra, V. Location of Golgi membranes with reference to dividing nuclei in syncytial *Drosophila* embryos. *Proc. Natl Acad. Sci. USA* **91**, 1878–1882 (1994).
46. Velasco, A. *et al.* Cell type-dependent variations in the subcellular distribution of α-mannosidase I and II. *J. Cell Biol.* **122**, 39–51 (1993).
47. Nilsson, T. *et al.* Overlapping distribution of two glycosyltransferases in the Golgi apparatus of HeLa cells. *J. Cell Biol.* **120**, 5–13 (1993).
48. Chou, C. F. & Omary, M. B. Mitotic arrest with anti-microtubule agents or okadaic acid is associated with increased glycoprotein terminal GlcNAc's. *J. Cell Sci.* **107**, 1833–1843 (1994).
49. Haltiwanger, R. S. & Philipsberg, G. A. Mitotic arrest with nocodazole induces selective changes in the level of O-linked N-acetylglucosamine and accumulation of incompletely processed N-glycans on proteins from HT29 cells. *J. Biol. Chem.* **272**, 8752–8758 (1997).
50. Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. Brefeldin A: insights into the control of membrane traffic and organelle structure. *J. Cell Biol.* **116**, 1071–1080 (1992).
51. Peyroche, A. *et al.* Brefeldin A acts to stabilize an abortive ARF-GDP-Sec7 domain protein complex: involvement of specific residues of the Sec7 domain. *Mol. Cell* **3**, 275–285 (1999).
52. Rogalski, A. A., Bergmann, J. E. & Singer, S. J. Effect of microtubule assembly status on the intracellular processing and surface expression of an integral protein of the plasma membrane. *J. Cell Biol.* **99**, 1101–1109 (1984).
53. Sampath, D., Varki, A. & Freeze, H. H. The spectrum of incomplete N-linked oligosaccharides synthesized by endothelial cells in the presence of brefeldin A. *J. Biol. Chem.* **267**, 4440–4455 (1992).
54. Balch, W. E., Dunphy, W. G., Braell, W. A. & Rothman, J. E. Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. *Cell* **39**, 405–416 (1984).
55. Glick, B. S. & Luini, A. Models for Golgi traffic: a critical assessment. *Cold Spring Harb. Perspect. Biol.* **3**, a005215 (2011).
56. Goud, B. & Gleeson, P. A. TGN golgins, Rabs and cytoskeleton: regulating the Golgi trafficking highways. *Trends Cell Biol.* **20**, 329–336 (2010).
57. Jackson, C. L. Mechanisms of transport through the Golgi complex. *J. Cell Sci.* **122**, 443–452 (2009).
58. Jamieson, J. D. & Palade, G. E. Intracellular transport of secretory proteins in the pancreatic exocrine cell. 3. Dissociation of intracellular transport from protein synthesis. *J. Cell Biol.* **39**, 580–588 (1968).
59. Kartberg, F., Elsner, M., Froderberg, L., Asp, L. & Nilsson, T. Commuting between Golgi cisternae--mind the GAP! *Biochim. Biophys. Acta* **1744**, 351–363 (2005).
60. Marchi, F. & Leblond, C. P. Radioautographic characterization of successive compartments along the rough endoplasmic reticulum-Golgi pathway of collagen precursors in foot pad fibroblasts of [3H] proline-injected rats. *J. Cell Biol.* **98**, 1705–1709 (1984).
61. Reynders, E., Foulquier, F., Annaert, W. & Matthijs, G. How Golgi glycosylation meets and needs trafficking: the case of the COG complex. *Glycobiology* **21**, 853–863 (2011).
62. Munro, S. An investigation of the role of transmembrane domains in Golgi protein retention. *EMBO J.* **14**, 4695–4704 (1995).
63. Patterson, G. H. *et al.* Transport through the Golgi apparatus by rapid partitioning within a two-phase membrane system. *Cell* **133**, 1055–1067 (2008). **Proposes a model for Golgi trafficking in which cargo proteins are not carried through the Golgi as if on a conveyer belt. Rather, cargo proteins partition into processing or transport domains that possess distinctive lipid compositions.**
64. Sharpe, H. J., Stevens, T. J. & Munro, S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169 (2010). **A tour-de-force bioinformatic analysis of the physico-chemical properties of transmembrane domains of organellar proteins from fungi to vertebrates, demonstrating a conserved dichotomy in transmembrane domain properties that distinguish early from late secretory compartments.**
65. Schmitz, K. R. *et al.* Golgi localization of glycosyltransferases requires a Vps74p oligomer. *Dev. Cell* **14**, 523–534 (2008).
66. Tu, L., Tai, W. C., Chen, L. & Banfield, D. K. Signal-mediated dynamic retention of glycosyltransferases in the Golgi. *Science* **321**, 404–407 (2008).
67. Aoki, D., Lee, N., Yamaguchi, N., Dubois, C. & Fukuda, M. N. Golgi retention of a *trans*-Golgi membrane protein, galactosyltransferase, requires cysteine and histidine residues within the membrane-anchoring domain. *Proc. Natl Acad. Sci. USA* **89**, 4319–4323 (1992).

68. Colley, K. J. Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology* **7**, 1–13 (1997).

69. Colley, K. J., Lee, E. U. & Paulson, J. C. The signal anchor and stem regions of the β-galactoside α2,6-sialyltransferase may each act to localize the enzyme to the Golgi apparatus. *J. Biol. Chem.* **267**, 7784–7793 (1992).

70. Fenteany, F. H. & Colley, K. J. Multiple signals are required for α2,6-sialyltransferase (ST6Gal I) oligomerization and Golgi localization. *J. Biol. Chem.* **280**, 5423–5429 (2005).

71. Nilsson, T., Lucocq, J. M., Mackay, D. & Warren, G. The membrane spanning domain of β-1,4-galactosyltransferase specifies *trans* Golgi localization. *EMBO J.* **10**, 3567–3575 (1991).

72. Nilsson, T., Rabouille, C., Hui, N., Watson, R. & Warren, G. The role of the membrane-spanning domain and stalk region of *N*-acetylglucosaminyltransferase I in retention, kin recognition and structural maintenance of the Golgi apparatus in HeLa cells. *J. Cell Sci.* **109**, 1975–1989 (1996).

73. Banfield, D. K. Mechanisms of protein retention in the Golgi. *Cold Spring Harb. Perspect. Biol.* **3**, a005264 (2011).

74. Tu, L. & Banfield, D. K. Localization of Golgi-resident glycosyltransferases. *Cell. Mol. Life Sci.* **67**, 29–41 (2010).

75. Nilsson, T., Slusarewicz, P., Hoe, M. H. & Warren, G. Kin recognition. A model for the retention of Golgi enzymes. *FEBS Lett.* **330**, 1–4 (1993).

76. **Hassinen, A., Rivinoja, A., Kauppila, A. & Kellokumpu, S. Golgi N-glycosyltransferases form both homo- and heterodimeric enzyme complexes in live cells. *J. Biol. Chem.* 285, 17771–17777 (2010). Bimolecular fluorescence complementation reveals that key glycosyltransferases form homocomplexes, but also form distinct heterocomplexes with functionally related glycosyl-transferases that are localized to either medial or *trans*-Golgi compartments.**

77. Daniotti, J. L., Martina, J. A., Giraudo, C. G., Zurita, A. R. & Maccioni, H. J. GM3 α2,8-sialyltransferase (GD3 synthase): protein characterization and sub-golgi location in CHO-K1 cells. *J. Neurochem.* **74**, 1711–1720 (2000).

78. Giraudo, C. G. & Maccioni, H. J. Ganglioside glycosyltransferases organize in distinct multienzyme complexes in CHO-K1 cells. *J. Biol. Chem.* **278**, 40262–40271 (2003).

79. Yano, H. *et al.* Distinct functional units of the Golgi complex in *Drosophila* cells. *Proc. Natl Acad. Sci. USA* **102**, 13467–13472 (2005).

80. Freeze, H. H. Congenital disorders of glycosylation: CDG-I, CDG-II, and beyond. *Curr. Mol. Med.* **7**, 389–396 (2007).

81. Foulquier, F. COG defects, birth and rise! *Biochim. Biophys. Acta* **1792**, 896–902 (2009).

82. Pokrovskaya, I. D. *et al.* Conserved oligomeric Golgi complex specifically regulates the maintenance of Golgi glycosylation machinery. *Glycobiology* **21**, 1554–1569 (2011).

83. Smith, R. D. & Lupashin, V. V. Role of the conserved oligomeric Golgi (COG) complex in protein glycosylation. *Carbohydr. Res.* **343**, 2024–2031 (2008).

84. Wu, X. *et al.* Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nature Med.* **10**, 518–523 (2004). **First demonstration of a human disorder associated with altered Golgi transport. Patient fibroblasts were shown to be deficient in intra-Golgi transport.**

85. Oka, T., Ungar, D., Hughson, F. M. & Krieger, M. The COG and COPI complexes interact to control the abundance of GEARs, a subset of Golgi integral membrane proteins. *Mol. Biol. Cell* **15**, 2423–2435 (2004).

86. Gill, D. J., Chia, J., Senewiratne, J. & Bard, F. Regulation of *O*-glycosylation through Golgi-to-ER relocation of initiation enzymes. *J. Cell Biol.* **189**, 843–858 (2010). **SRC-dependent growth factor stimulation drives the relocation of polypeptide GalNAc transferases to earlier secretory compartments, enhancing the production of GalNAc initiated *O*-linked glycans.**

87. Baas, S. *et al.* Sugar-free frosting, a homolog of SAD kinase, drives neural-specific glycan expression in the *Drosophila* embryo. *Development* **138**, 553–563 (2011).

A partial loss-of-function mutation in a neural specific kinase decreases expression of a set of neural-specific glycans, demonstrating a direct link between protein phosphorylation, Golgi architecture and glycan end-products.

88. Farhan, H. *et al.* MAPK signaling to the early secretory pathway revealed by kinase/phosphatase functional screening. *J. Cell Biol.* **189**, 997–1011 (2010). **An RNA interference screen for kinases and phosphatases that alter ER and Golgi trafficking identified 122 such proteins, indicating a broad role for cell signaling in sculpting flux through the secretory pathway with subsequent, but not yet demonstrated, glycomic consequences.**

89. Lowe, M. *et al.* Cdc2 kinase directly phosphorylates the *cis*-Golgi matrix protein GM130 and is required for Golgi fragmentation in mitosis. *Cell* **94**, 783–793 (1998).

90. Muniz, M., Alonso, M., Hidalgo, J. & Velasco, A. A regulatory role for cAMP-dependent protein kinase in protein traffic along the exocytic route. *J. Biol. Chem.* **271**, 30935–30941 (1996).

91. Colley, K. J. Structural basis for the polysialylation of the neural cell adhesion molecule. *Adv. Exp. Med. Biol.* **663**, 111–126 (2010).

92. Rana, N. A. & Haltiwanger, R. S. Fringe benefits: functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors. *Curr. Opin. Struct. Biol.* **21**, 583–589 (2011).

93. Hanisch, F. G. & Breloy, I. Protein-specific glycosylation: signal patches and *cis*-controlling peptidic elements. *Biol. Chem.* **390**, 619–626 (2009).

94. Luther, K. B. & Haltiwanger, R. S. Role of unusual O-glycans in intercellular signaling. *Int. J. Biochem. Cell Biol.* **41**, 1011–1024 (2009).

95. Muntoni, F., Torelli, S., Wells, D. J. & Brown, S. C. Muscular dystrophies due to glycosylation defects: diagnosis and therapeutic strategies. *Curr. Opin. Neurol.* **24**, 437–442 (2011).

96. Stanley, P. & Okajima, T. Roles of glycosylation in Notch signaling. *Curr. Top. Dev. Biol.* **92**, 131–164 (2010).

97. Wopereis, S., Lefeber, D. J., Morava, E. & Wevers, R. A. Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin. Chem.* **52**, 574–600 (2006).

98. Sakaidani, Y. *et al.* O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1. *Biochem. Biophys. Res. Commun.* **419**, 14–19 (2012). **The authors identified a unique O-GlcNAc transferase, EOGT, involved in the O-GlcNAcylation of extracellular proteins.**

99. Shao, L., Moloney, D. J. & Haltiwanger, R. Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the Abruptex region. *J. Biol. Chem.* **278**, 7775–7782 (2003).

100. Wang, Y. *et al.* Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein *O*-fucosyltransferase. *J. Biol. Chem.* **276**, 40338–40345 (2001).

101. Okajima, T. & Irvine, K. D. Regulation of Notch signaling by O-linked fucose. *Cell* **111**, 893–904 (2002).

102. Shi, S. & Stanley, P. Protein *O*-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc. Natl Acad. Sci. USA* **100**, 5234–5239 (2003).

103. Yao, D. *et al.* Protein O-fucosyltransferase 1 (Pofut1) regulates lymphoid and myeloid homeostasis through modulation of Notch receptor ligand interactions. *Blood* **117**, 5652–5662 (2011).

104. Acar, M. *et al.* Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **132**, 247–258 (2008).

105. Sethi, M. K. *et al.* Identification of glycosyltransferase 8 family members as xylosyltransferases acting on *O*-glucosylated notch epidermal growth factor repeats. *J. Biol. Chem.* **285**, 1582–1586 (2010).

106. Sethi, M. K. *et al.* Molecular cloning of a xylosyltransferase that transfers the second xylose to *O*-glucosylated epidermal growth factor repeats of notch. *J. Biol. Chem.* **287**, 2739–2748 (2012).

107. Rana, N. A. *et al.* O-Glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. *J. Biol. Chem.* **286**, 31623–31637 (2011).

108. Fernandez-Valdivia, R. *et al.* Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. *Development* **138**, 1925–1934 (2011).

109. Sakaidani, Y. *et al.* O-linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell-matrix interactions. *Nature Commun.* **2**, 583 (2011).

110. Matsuura, A. *et al.* O-linked *N*-acetylglucosamine is present on the extracellular domain of notch receptors. *J. Biol. Chem.* **283**, 35486–35495 (2008).

111. Hofsteenge, J. *et al.* C-mannosylation and O-fucosylation of the thrombospondin type 1 module. *J. Biol. Chem.* **276**, 6485–6498 (2001).

112. Luo, Y., Koles, K., Vorndam, W., Haltiwanger, R. S. & Panin, V. M. Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *J. Biol. Chem.* **281**, 9393–9399 (2006).

113. Martinez-Duncker, I., Mollicone, R., Candelier, J. J., Breton, C. & Oriol, R. A new superfamily of protein-*O*-fucosyltransferases, α2-fucosyltransferases, and α6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. *Glycobiology* **13**, 1C–5C (2003).

114. Kozma, K. *et al.* Identification and characterization of a β1,3-glucosyltransferase that synthesizes the Glc-β1,3-Fuc disaccharide on thrombospondin type 1 repeats. *J. Biol. Chem.* **281**, 36742–36751 (2006).

115. Sato, T. *et al.* Molecular cloning and characterization of a novel human β1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain. *Glycobiology* **16**, 1194–1206 (2006).

116. Ricketts, L. M., Dlugosz, M., Luther, K. B., Haltiwanger, R. S. & Majerus, E. M. *O*-fucosylation is required for ADAMTS13 secretion. *J. Biol. Chem.* **282**, 17014–17023 (2007).

117. Du, J. *et al.* O-fucosylation of thrombospondin type 1 repeats restricts epithelial to mesenchymal transition (EMT) and maintains epiblast pluripotency during mouse gastrulation. *Dev. Biol.* **346**, 25–38 (2010).

118. Doucey, M. A., Hess, D., Cacan, R. & Hofsteenge, J. Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor. *Mol. Biol. Cell* **9**, 291–300 (1998).

119. Krieg, J. *et al.* Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. *Mol. Biol. Cell* **9**, 301–309 (1998).

120. Wang, L. W., Leonhard-Melief, C., Haltiwanger, R. S. & Apte, S. S. Post-translational modification of thrombospondin type-1 repeats in ADAMTS-like 1/punctin-1 by *C*-mannosylation of tryptophan. *J. Biol. Chem.* **284**, 30004–30015 (2009).

121. Barresi, R. & Campbell, K. P. Dystroglycan: from biosynthesis to pathogenesis of human disease. *J. Cell Sci.* **119**, 199–207 (2006).

122. Stalnaker, S. H. *et al.* Glycomic analyses of mouse models of congenital muscular dystrophy. *J. Biol. Chem.* **286**, 21180–21190 (2011). **Analysis of O-linked glycans from brains of wild-type and knockout mouse models using mass spectrophotometry to determine that POMGNT1 is required for addition of β1,2-linked GlcNAc on O-Man glycans.**

123. Stalnaker, S. H. *et al.* Site mapping and characterization of O-glycan structures on α-dystroglycan isolated from rabbit skeletal muscle. *J. Biol. Chem.* **285**, 24882–24891 (2010).

124. Combs, A. C. & Ervasti, J. M. Enhanced laminin binding by α-dystroglycan after enzymatic deglycosylation. *Biochem. J.* **390**, 303–309 (2005).

125. Hara, Y. *et al.* A dystroglycan mutation associated with limb-girdle muscular dystrophy. *N. Engl. J. Med.* **364**, 939–946 (2011).

126. Yoshida-Moriguchi, T. *et al.* O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science* **327**, 88–92 (2010). **First description of a novel GAG-like structure containing Xyl and GlcA in unknown linkage to an essential extracellular protein that interacts with the extracellular matrix.**

127. Inamori, K. *et al.* Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science* **335**, 93–96 (2012). **Reports the bifunctional glycosyltransferase activity of LARGE1 on O-Man-containing glycans of αDG.**

128. Zhang, Z., Zhang, P. & Hu, H. LARGE expression augments the glycosylation of glycoproteins in addition to α-dystroglycan conferring laminin binding. *PLoS ONE* **6**, e19080 (2011).

129. Nairn, A. V., dela Rosa, M. & Moremen, K. W. Transcript analysis of stem cells. *Methods Enzymol.* **479**, 73–91 (2010).

130. Hua, S. *et al.* Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst* **136**, 3663–3671 (2011).

131. An, H. J., Kronewitter, S. R., de Leoz, M. L. & Lebrilla, C. B. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **13**, 601–607 (2009).

132. Bielik, A. M. & Zaia, J. Historical overview of glycoanalysis. *Methods Mol. Biol.* **600**, 9–30 (2010).

133. Aoki, K. *et al.* The diversity of *O*-linked glycans expressed during *Drosophila melanogaster* development reflects stage- and tissue-specific requirements for cell signaling. *J. Biol. Chem.* **283**, 30385–30400 (2008).

134. Haab, B. B. Antibody-lectin sandwich arrays for biomarker and glycobiology studies. *Expert Rev. Proteom.* **7**, 9–11 (2010).

135. Yue, T. *et al.* The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol. Cell. Proteomics* **8**, 1697–1707 (2009).

136. Li, C. *et al.* Pancreatic cancer serum detection using a lectin/glyco-antibody array method. *J. Proteome Res.* **8**, 483–492 (2009).

137. Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Biochim. Biophys. Acta* 9 Dec 2011 (doi:10.1016/j.bbagen.2011.12.001).

138. Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-glycome identifies HNF1α as a master regulator of plasma protein fucosylation. *PLoS Genet.* **6**, e1001256 (2010).
Describes the first study that linked high-throughput glycan structural analysis and genome-wide association to identify determinants that control population-specific expression of carbohydrate epitopes.

139. Huffman, J. E. *et al.* Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum. Mol. Genet.* **20**, 5000–5011 (2011).

140. Esko, J. D. & Selleck, S. B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435–471 (2002).

141. Wang, A., de la Motte, C., Lauer, M. & Hascall, V. Hyaluronan matrices in pathobiological processes. *FEBS J.* **278**, 1412–1418 (2011).

142. Tian, E. & Ten Hagen, K. G. Recent insights into the biological roles of mucin-type *O*-glycosylation. *Glycoconj. J.* **26**, 325–334 (2009).

143. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287**, 1955–1961 (2012).

144. Schegg, B., Hulsmeier, A. J., Rutschmann, C., Maag, C. & Hennet, T. Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases. *Mol. Cell. Biol.* **29**, 943–952 (2009).

145. Pontier, S. M. & Schweisguth, F. Glycosphingolipids in signaling and development: from liposomes to model organisms. *Dev. Dyn.* **241**, 92–106 (2012).

146. Farquhar, M. G. & Palade, G. E. The Golgi apparatus: 100 years of progress and controversy. *Trends Cell Biol.* **8**, 2–10 (1998).

147. Perry, R. J. & Ridgway, N. D. Molecular mechanisms and regulation of ceramide transport. *Biochim. Biophys. Acta* **1734**, 220–234 (2005).

148. van Meer, G. & de Kroon, A. I. Lipid map of the mammalian cell. *J. Cell Sci.* **124**, 5–8 (2011).

149. Smith, M. H., Ploegh, H. L. & Weissman, J. S. Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. *Science* **334**, 1086–1090 (2011).

150. Seemann, J., Jokitalo, E., Pypaert, M. & Warren, G. Matrix proteins can generate the higher order architecture of the Golgi apparatus. *Nature* **407**, 1022–1026 (2000).

151. Slusarewicz, P., Nilsson, T., Hui, N., Watson, R. & Warren, G. Isolation of a matrix that binds medial Golgi enzymes. *J. Cell Biol.* **124**, 405–413 (1994).

152. Ramirez, I. B. & Lowe, M. Golgins and GRASPs: holding the Golgi together. *Semin. Cell Dev. Biol.* **20**, 770–779 (2009).

153. Maccioni, H. J., Quiroga, R. & Spessott, W. Organization of the synthesis of glycolipid oligosaccharides in the Golgi complex. *FEBS Lett.* **585**, 1691–1698 (2011).

154. Gerken, T. A. *et al.* Emerging paradigms for the initiation of mucin-type protein *O*-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* **286**, 14493–14507 (2011).

155. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, 310–322 (2002).

156. Roch, C., Kuhn, J., Kleesiek, K. & Gotting, C. Differences in gene expression of human xylosyltransferases and determination of acceptor specificities for various proteoglycans. *Biochem. Biophys. Res. Commun.* **391**, 685–691 (2010).

157. Manya, H. *et al.* Regulation of mammalian protein *O*-mannosylation: preferential amino acid sequence for *O*-mannose modification. *J. Biol. Chem.* **282**, 20200–20206 (2007).

158. Pierleoni, A., Martelli, P. L. & Casadio, R. PredGPI: a GPI-anchor predictor. *BMC Bioinformat.* **9**, 392 (2008).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Complex Carbohydrate Research Center at the University of Georgia: http://www.ccrc.uga.edu
Carbohydrate-Active enZYmes Database: http://www.cazy.org
CFG Functional Glycomics Gateway: http://www.functionalglycomics.org/static/consortium/consortium.shtml
Essentials of Glycobiology, 2nd edn: http://www.ncbi.nlm.nih.gov/books/NBK1908/
Glycobiology journal: http://glycob.oxfordjournals.org
Integrated Technology Resource for Biomedical Glycomics: http://www.ccrc.uga.edu/~moremen/glycomics/
Society for Glycobiology: http://www.glycobiology.org
YinOYang WWW server (neural network predictions for *O*-β-GlcNAc attachment sites): http://www.cbs.dtu.dk/services/YinOYang

ALL LINKS ARE ACTIVE IN THE ONLINE PDF