Review

CellPress

Special Issue: The Magic of the Sugar Code

Glycan variation and evolution in the eukaryotes

Anthony P. Corfield[1] and Monica Berry[2]

[1] Mucin Research Group, University of Bristol, School of Clinical Sciences, Bristol Royal Infirmary, Bristol, BS2 8HW, UK
[2] University of Bristol, School of Physics, Centre for Nanoscience and Quantum Information, Tyndall Avenue, Bristol, BS8 1FD, UK

In this review, we document the evolution of common glycan structures in the eukaryotes, and illustrate the considerable variety of oligosaccharides existing in these organisms. We focus on the families of *N*- and *O*-glycans, glycosphingolipids, glycosaminoglycans, glycosylphosphatidylinositol (GPI) anchors, sialic acids (Sias), and cytoplasmic and nuclear glycans. We also outline similar and divergent aspects of the glycans during evolution within the groups, which include inter- and intraspecies differences, molecular mimicry, viral glycosylation adaptations, glycosyltransferase specificity relating to function, and the natural dynamism powering these events. Finally, we present an overview of the patterns of glycosylation found within the groups comprising the Eukaryota, namely the Deuterostomia, Fungi, Viridiplantae, Nematoda, and Arthropoda.

Glycan structure: why it matters

It is well known that the surface of cells at all levels of life is coated with an array of glycans, often termed the 'glycocalyx' [1–4]. The presence of this sugar coat is linked with biological functions that are fundamental to the normal behavior of the organisms throughout their life. Such functions range from imparting protein stability and function [5], to mediating the interaction of cells with the extracellular matrix [6–9], to sperm–egg binding processes [10], and host–pathogen interactions [11–16]. The biological choice of glycans as targets for such interactions reflects both their diversity and specificity. It also places demands on glycosylation to maintain and adapt to dynamic biological environments.

Here, we consider glycan structure and glycoconjugate function within the eukaryotes. Details for the individual phyla can be found in supplemental material online. A phylogenetic tree for eukaryotic species that have had their genomes sequenced is shown in Figure 1. The CAZy database (http://www.cazy.org) provides an invaluable source of information regarding all proteins that manipulate glycans in all their forms and is widely used as a primary target for genomic screening of species-specific glycomes [17]. The Consortium for Functional Glycomics website also provides a valuable tool for glycomic research (http://www.functionalglycomics.org). A consequence of the increased biological interest in glycans has been a focus on the chemistry–glycobiology frontier and the need to understand chemical and physical approaches for glycan analysis [3,4,18,19].

Sugars chains face the outside world and can occupy most of the hydrodynamic volume of any molecules that carry them. Together with cognate molecules, they create bonds and assist intercellular recognition. Through spatial organization and concentration, they confer a further degree of selectivity to the formation of these bonds. Glycosylation is the default protein modification across prokaryotes [20] and eukaryotes; it affects fundamental and conserved aspects of the protein life cycle, including folding, intracellular trafficking, and co-translational quality control. The processes that enable protein maturation are homologous and occur in all kingdoms of life. Relevant to this discussion, the assembly of an oligosaccharide on the cytoplasmic side of the plasma membrane by a set of specific glycosyltransferases and the subsequent translocation of the lipid-linked oligosaccharide is a unifying scheme in all systems.

Changes to glycosylation in tumor cells [21] occur to both core and outer carbohydrate structures. Among the first are: (i) overexpression of β1–6 branching of *N*-glycan [22]; (ii) greatly enhanced presence of truncated T, Tn, and sialyl-Tn glycans on mucins [23,24]; and (iii) loss of GPI anchors [25]. Increased Sia content [26,27], and enrichment in hyaluronan in the stroma surrounding a tumor [28–30] are examples of peripheral glycan changes in malignancies.

Here, we focus on eukaryotic glycosylation, which represents only an instance of glycosylation, given that continuous glycan evolution is traceable through to Eubacteria and Archaea. A more comprehensive review of bacterial glycosylation can be found in the article by Tan *et al.* in this special issue of *TiBS* [20]. Several earlier publications have addressed the variation of eukaryotic glycosylation compared with other organisms and their role in evolution (e.g., [31–38]). Furthermore, the scope of eukaryotic glycosylation is too broad to review in full detail; here, we highlight the fundamental patterns and concepts of glycan structure and evolution, illustrated with examples.

Corresponding authors: Corfield, A.P. (mdapc@bristol.ac.uk);
Berry, M. (mon.berry@bristol.ac.uk).
Keywords: glycan; glycosylation; eukaryotes; evolution; glycosyltransferase; molecular mimicry.

0968-0004/
 http://dx.doi.org/10.1016/j.tibs.2015.04.004

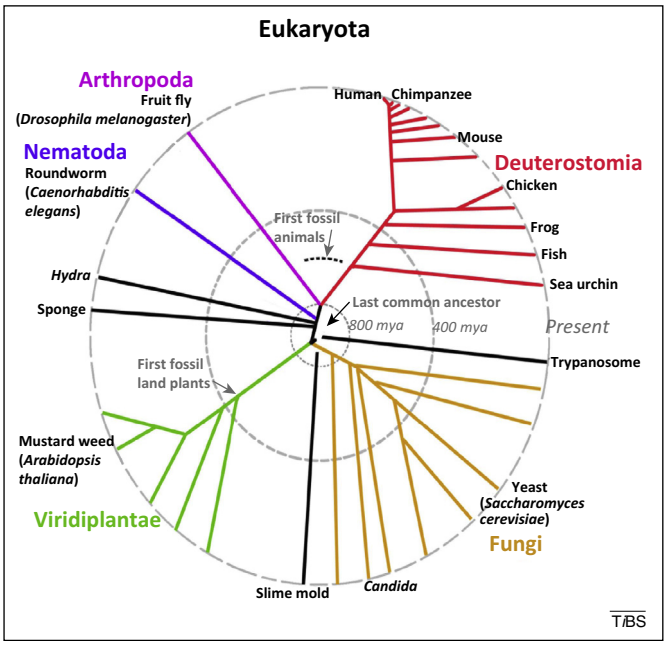

Figure 1. Eukaryote phylogenetic tree. Phylogenetic tree of the multicellular Eukaryota, showing the main groups discussed in this review. Reproduced from Essentials of Glycobiology, Chapter 19 page 282, with permission of Cold Spring Harbor Press, P. Gagneux and TD Pollard. Abbreviation: mya, million years ago.

Glycan variation in the eukaryotes

Examination of glycoconjugate properties immediately uncovers a multiplicity of structures that require appraisal in terms of their function within these molecules. Table 1 shows examples of the major groups of proteins that manipulate glycoconjugates, highlighting the broad range of events that are involved in glycobiological metabolism.

Initial consideration of glycan structures must include the metabolic pathways that lead to their formation. The monosaccharides that form the building blocks are derived from diverse sources. Some are obtained from extracellular origins via salvage pathways from both external and intracellular sources, together with the major intracellular glycolytic pathways that feed monosaccharides into glycan synthesis and degradation.

A series of interlinked metabolic pathways produces the precursors of the nucleotide sugars, while sugar transporters facilitate the transport of glucose across the plasma membrane [34] in either an energy-independent (facilitated diffusion transporters, the GLUT family), or an energy- and sodium-dependent way (SGLT family). Monosaccharides are activated to high-energy nucleotide sugars (UDP-Glc, UDP-Gal, UDP-GlcNAc, UDP-GalNAc, UDP-GlcA, UDP-Xyl, GDP-Man, GDP-Fuc, and CMP-NeuAc), which act as universal donors in glycosyltransferase reactions for each glycosynthetic pathway [39,40]. Glycosulfotransferases utilize 3′-phosphoadenosyl-5′-phosphosulfate (PAPS) as the active sulfate donor. Phosphorylated sugars, such as glucose-6-phosphate, mannose-6-phosphate, and glucosamine-6-phosphate, are similarly formed through the action of kinases with ATP as the phosphate donor [39,40].

Pathways leading to the nucleotide sugars are regulated through feedback inhibition of the key enzymatic steps initiating each route [39]. The entry of glucose into the

Table 1. Major groups of proteins manipulating carbohydrates in eukaryotes[a]

Protein family	Main categories
Glycosyltransferases	Transfer of mono- or oligosaccharides to acceptors, mono- and oligosaccharides, proteins, lipids, and DNA using nucleotide sugars and dolichol phosphate-mono and oligosaccharides
	A large family of enzymes catalyzing the transfer of fucose, mannose, glucose, galactose, xylose, *N*-acetylglucosamine, *N*-acetylgalactosamine, *N*-acetylmuramic acids, glucuronic acid, and Sias
	Also includes transfer of sulfate, phosphate, acetyl, methyl, pyruvate, and ethanolamine to glycans
Glycosidases	Catalyze the hydrolytic cleavage of the glycosidic bond in all glycoconjugates; function in biosynthetic and catabolic cellular processes
	Over 100 families identified
Polysaccharide lyases	Cleave polysaccharides containing uronic acid through a β-elimination mechanism forming a new reducing end at the cleavage site
	Nine families identified
Carbohydrate esterases	Remove acetyl groups from polysaccharides, including pectin, xylan, galactoglucomannan, rhamnogalacturan and glycan linked Sias
	16 or more families identified
Carbohydrate-binding modules	Noncatalytic proteins, that bind soluble and crystalline carbohydrates; usually present in proteins together with catalytic domains in carbohydrate-active enzymes
	Found in glycoside hydrolases, polysaccharide lyases, polysaccharide oxidases, glycosyltransferases, and plant cell wall expansins
	14 families identified
Glycan-binding protein (GBP) molecules	GBPs enable binding of the carbohydrate-active enzyme to its substrate, thus increasing the local concentration of the enzyme and enhancing substrate degradation
	Examples of proteins containing GBPs: C-type lectins, proteoglycan core proteins, Type II membrane receptors, collectins, selectins, dectins, mannose receptors, layilin, galectins, and siglecs
Sugar transporters	Transport of sugars across the plasma membrane into cells
	Hexose transporters
	Energy-independent, facilitated diffusion, glucose transporters (GLUT protein family; *SLC2* gene) in yeast and mammals.
	Energy dependent; e.g., sodium-dependent glucose transporters (SGLT proteins; *SLC5A* gene)

[a]A brief overview of eukaryote proteins that interact with carbohydrates. More detail can be found on the CAZy website (http://www.cazy.org), at http://www.cazypedia.org and on The Consortium for Functional Glycomics website (http://www.fuctionalglycomics.org). All three sites provide further links to other carbohydrate-relevant databases.