

THE COMPLEXITY OF GLYCAN STRUCTURES, FUNCTIONS, AND ORIGINS

Russell W. Carlson^{*}

Professor Emeritus, Complex Carbohydrate Research Center, Department of Biochemistry & Molecular Biology, University of Georgia, Athens, GA 30602, USA

Abstract

Sugars have more reactive groups than either amino acids or nucleotides. This allows them to form branched oligo- and polysaccharides. Therefore, the number of possible structures is far greater than comparably sized peptides or nucleic acids. The result is that carbohydrates have far greater information-carrying potential than either proteins or nucleic acids. Carbohydrates are present in, and essential for, all three of life's kingdoms; archaea, bacteria, and eukaryotes. This article illustrates the structural, functional, and evolutionary complexity of animal carbohydrates, i.e. glycans. These glycans include glycosaminoglycans, glycolipids, and glycoproteins. Attachment of glycans to proteins increases their functional range and allows their functions to be fine-tuned in response to the environment. Different copies of a single protein have different glycans attached. Each copy is a glycoform and the combination of a protein's glycoforms is called its glycotype. Protein glycotypes are cell-, tissue- and developmentally-specific. The combination glycoproteins, glycosaminoglycans, and glycolipids form glycopatterns that are essential for organization of cells into tissues during development. The synthesis of a single glycan structure requires numerous enzymes, and the production of the necessary functional protein glycotypes and cellular glycopatterns requires integrated regulated expression of these enzymes by (glyco)gene regulatory networks (gGRNs) and micro RNA (miRNA) regulatory networks. Explanations for the origin of glycans must include mechanisms for the origin of glycosyl transferases, hydrolases, metabolic pathways for the activated sugar precursors, and all the components and information necessary to form the aforementioned regulatory networks. The challenges that this poses for a standard neo-Darwinian explanation are discussed.

Keywords

glycan • complexity • structures • function • evolution

Introduction

Glycans (carbohydrates) are one of the four major classes of molecules, together with nucleic acids, proteins, and lipids, that are necessary for life. This article is not intended to give a comprehensive review of carbohydrate chemistry or glycobiology, but, instead, provide examples that illustrate the complexity of their chemistry, function, and origins. For an excellent comprehensive book on glycobiology, the readers are referred to 'Essentials in Glycobiology' which is available online (<https://ncbi.nlm.nih.gov/books/NBK579918/>), as well as an excellent description of glycan diversity by Richard Cummings in this journal (1).

Glycans are comprised of oligomers and polymers of sugars (monosaccharides). They are ubiquitous in nature and necessary for life (2, 3). A living organism contains many different types of glycans. For example, in animals they are present as polysaccharides such as glycosaminoglycans (GAGs; these include heparin and hyaluronic acid which are not linked to proteins, and heparin sulfate, keratin sulfate and

dermatan sulfate which are linked to proteins and known as proteoglycans) and oligosaccharides attached to lipids and proteins, and they all have essential functions for the cell and for development. Unlike a protein, a glycan does not have a direct DNA template that determines its structure or from which one can predict its structure. In fact, as we will see, the synthesis of a single glycan structure requires many proteins, and, therefore, many genes. Its final structure depends of the expression of these genes (called glycogenes) as directed by regulatory networks at both the transcriptional and translational level, and can also be affected by epigenetic factors.

The importance of glycans for life has been described in numerous publications (2, 4–6). It has been noted that 'every cell [bacteria, archaea, and eukaryotes] on the planet is covered with a dense and complex array of glycans' [(2) (p. 13)] and that most proteins in multicellular organisms are glycosylated (7, 8). In fact, multicellular organisms would not exist without glycans. The elimination of any class of glycans

* Corresponding author e-mail: rwc@uga.edu

is lethal, and defects in glycosylation result in numerous and serious diseases [(2, 8–10) (chapter 3 of Ref. (2))]. That glycans are essential for life is reflected in this statement by the Committee on Assessing the Importance and Impact of Glycomics and Glycosciences: ‘... cell surface glycosylation is as important to understanding life as is the genetic code, yet our understanding of the information contained in glycosylation is rudimentary at best’ [(2) (p. 15)].

It was not that long ago that the essential nature of glycans for the functioning of a cell was unknown. In fact, as pointed out by Arnold et al. (11), it was not until 1988 that the term ‘glycobiology’ was first used in the *Annual Review of Biochemistry* (12) and ‘glycobiology’ was not included as word in the *Oxford English Dictionary* until 1992. Thus, the discovery that glycans are essential for organismal development and cellular function is, relative to proteins and nucleic acids, rather recent.

The structural complexity of glycans is far greater than that of nucleic acids or proteins

Why are carbohydrates structurally complex? The number of possible oligosaccharide structures is far greater than

the number of possible oligopeptide or oligonucleotide structures. Another way of stating this is that the ‘diversity [structural] space’ for oligosaccharides is much larger than for oligopeptides or oligonucleotides (13). This is due to the difference in chemistry between monosaccharides, amino acids, and nucleotides. As shown in Figure 1, L-amino acids form peptides via the formation of amide bonds between their carboxyl and α -amino groups, while nucleotides form oligo- or polynucleotides via phosphate diester bonds between their 5'- and 3'-hydroxyl groups. The result is that polypeptides and nucleic acids are linear oligo- or polymers of L-amino acids, or nucleotides and deoxynucleotides, respectively. In comparison, the glycosidic hydroxyl group (i.e. the C1-hydroxyl) of one sugar can react with one of several hydroxyls of another sugar (e.g. the C2, C3, C4, or C6-hydroxyls for a six-membered pyranose residue and C2, C3, C5 and C6 for a five-membered furanose residue). Thus, sugars can form numerous oligo- or polysaccharides that are not necessarily linear. In fact, they have a wide array of branching possibilities. The mathematically possible number of hexasaccharides that can be formed from D-hexose residues is $>10^{12}$ (14), while the number of possible hexanucleotides is 4096 (4^6) and the number of hexapeptides is 64,000,000 (20^6). It is for this reason that carbohydrates have a ‘... potential information content several orders of magnitude higher in a short

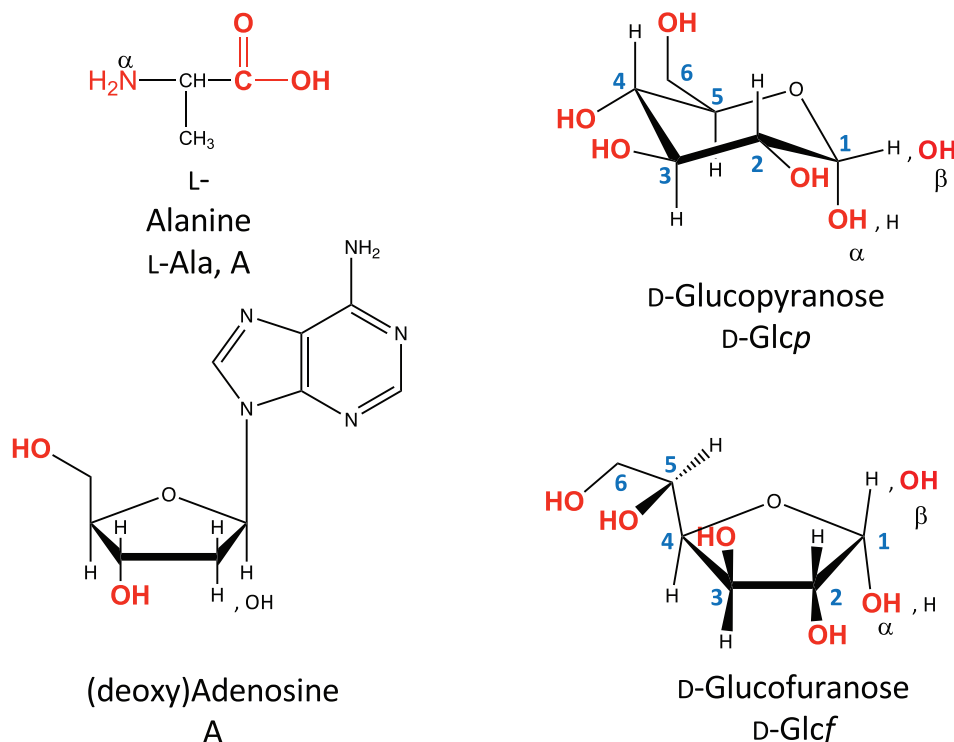


Figure 1. A structural comparison of an amino acid (alanine), a nucleotide (adenosine), and a hexose (D-glucopyranose and D-glucofuranose). The reactive functional groups are indicated in red.

sentence than in any other biological oligomer' (14), and that this level of complexity represents an 'Isomer Barrier' to '... the development of a single analytical method for the absolute characterization of carbohydrates' (14). It is this 'barrier' that makes structural characterization of carbohydrates, as well as their synthesis, much more complicated than proteins or nucleic acids.

Ten common monosaccharides found in mammalian cells (13) are shown in Figure 2 together with their abbreviations and symbolic representations. These monosaccharides are largely present as 6-membered pyranose ring structures and, when restricted to pyranose forms, the mathematical number of possible hexasaccharides is around 193 billion (13). However, most cellular oligosaccharides, whether linked to proteins or lipids, are larger than hexasaccharides, e.g. they are comprised of 9–15 rather than 6 saccharides, and, therefore the number of possible cellular oligosaccharides is orders of magnitude >193 billion. Of course, living organisms do not form all the mathematically possible number of any

biological polymer or oligomer. In fact, the actual number of oligosaccharide structures attached to proteins and lipids in a mammalian cell is probably 3000–4000 (this does not include the GAG structures). The cell accomplishes the amazing task of synthesizing these specific structures, out of the trillions of possibilities, at just the right time and just the right place for the proper development and functioning of an organism, even for the functioning of a single protein.

The attachment of glycans to proteins dramatically increases protein complexity

Glycans increase the structural and functional complexity of proteins. A very large portion of protein posttranslational modification is due to the attachment of glycans. In the '-omics' world, the proteome of an organism consists of all the proteins produced by that organism, the glycome consists of all the glycans produced, and the glycoproteome would be all the glycoproteins produced (e.g. see the Human Glycome Project at <https://human-glycome.org>). There are 3000 or so

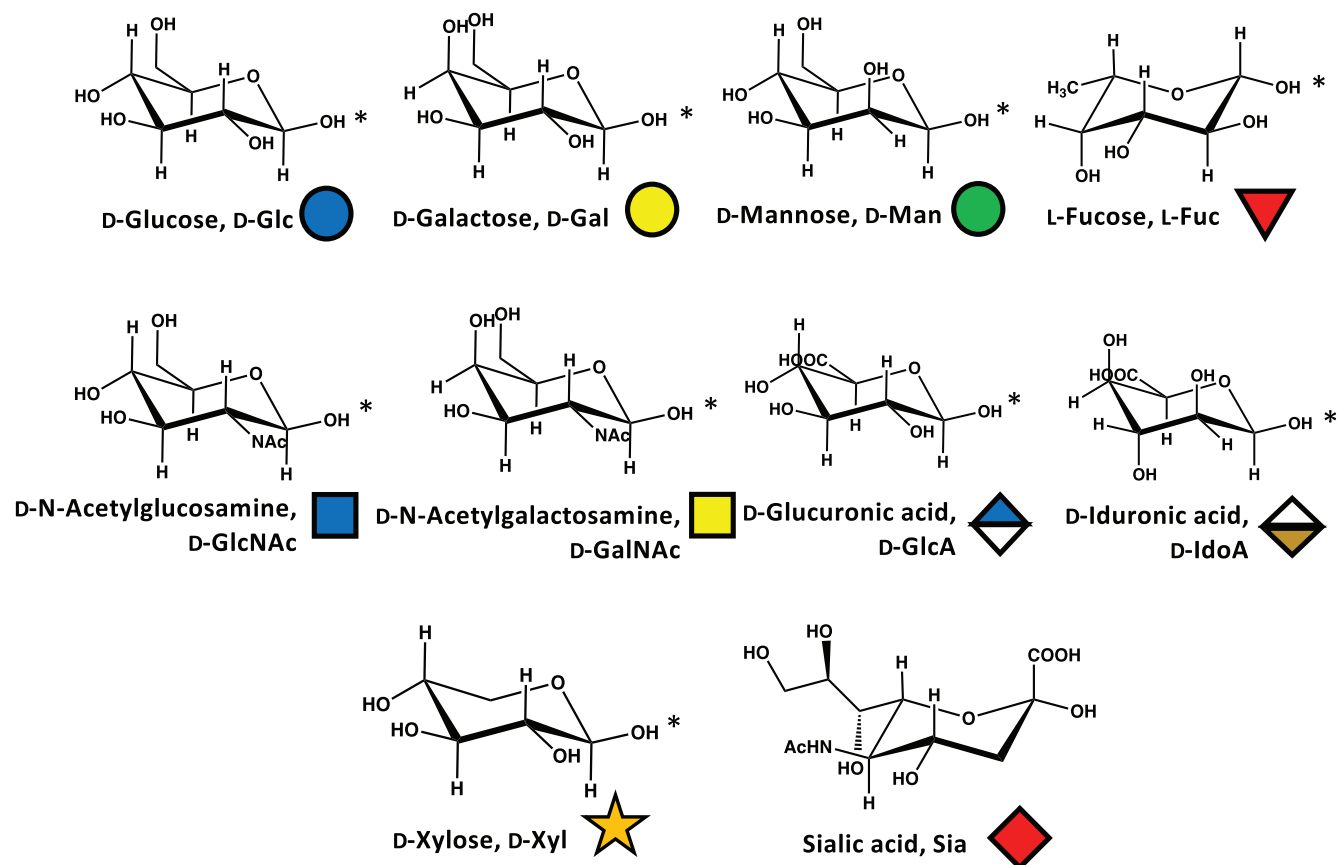


Figure 2. The ten most common saccharides found in mammalian cells and their symbols. The '*' indicates the anomeric hydroxyl which can be present in either the α - or β -configuration; the β -configurations are shown here.

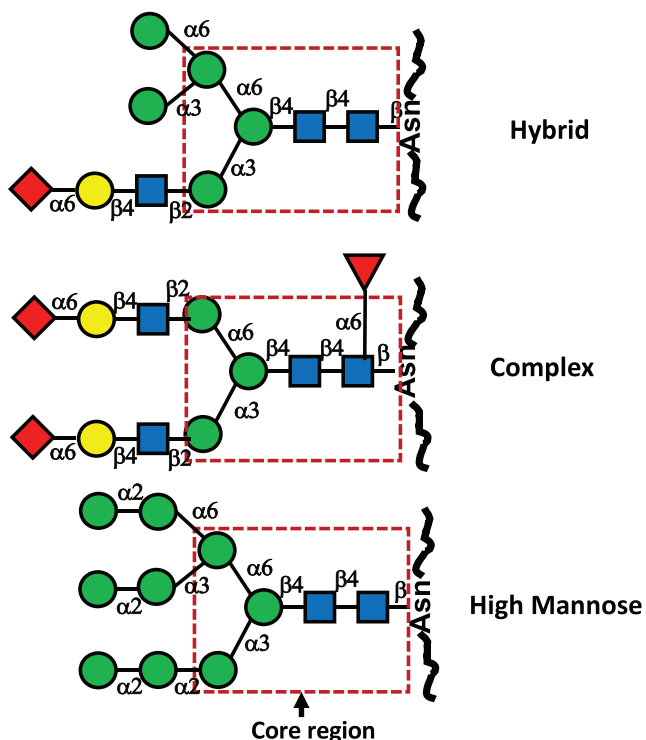


Figure 3. A symbolic representation of the different classes of *N*-linked glycan structures attached to proteins. The definitions of the symbols are shown in Figure 2. The core region is indicated by the red dashed box. Various saccharides are attached to a core region to form a variety of high mannose, complex, or hybrid glycans. High mannose glycans consist mannose residues attached to the core region. Complex glycans can have a variety of different monosaccharides attached to the core region, and, while two branches are shown here, hybrid glycans exist with up to six branches. The branches on these complex structures all begin with a GlcNAc residue, and each branch can be extended with repeats of a Galβ1-4GlcNAc (i.e. LacNAc, *N*-acetyllactosamine) disaccharide. GlcNAc, *N*-acetylglucosamine.

glycans that can be attached to proteins via *N*-glycosylation of asparagine residues located at consensus -Asn-X-Ser/Thr- sites (where X can be any amino acid except Pro). They vary in size from around 10 to 15 monosaccharides (some are larger) and are grouped into three general classes [see Figure 3 and Stanley et al. (15)]; *high mannose* which consist largely of mannose-comprised structures attached to a core oligosaccharide region, *complex* which consist of a variety of monosaccharides attached to the core region, and *hybrid* which have a high-mannose region as well as a region which can have a variety of different monosaccharides. The *complex* and *hybrid* structures can be greatly expanded due to the addition of branching *N*-acetylglucosaminosyl residues and further addition of monosaccharides to these branching residues. Other glycans are linked to proteins via *O*-glycosylation to the hydroxyl group of -Ser- or -Thr- residues which greatly increases the complexity of the glycoproteome. The ability of protein glycosylation to increase its functional

range is also discussed in this journal by Cummings (1). Here, the examples chosen to illustrate the functional complexity of carbohydrates will largely involve *N*-linked glycans, but the reader should know that this complexity extends to the other cellular glycans; i.e. protein *O*-linked glycans, glycolipids, GAGs, etc. (16).

Each protein can have multiple *N*-glycosylation sites, each of which can hypothetically be glycosylated by any of the 3000 different glycan structures so that a protein exists as a population consisting of a combination of many different glycoforms. The result is that the glycoproteome is orders of magnitude greater than the proteome. The number of possible glycoforms (i.e. the number of variations of a single protein that differ in the pattern of glycans attached) that can be produced by glycosylation of one protein can be described by the equation,

$$N = n^m + m \left[\sum_{x=1}^{x=m-1} n^x \right]$$

where N is the number of glycoforms, n = number of glycan structures, and m = number of glycosylation sites. As a very simple illustration consider a protein that has three glycosylation sites ($m = 3$) in which each site can be glycosylated by any one of three different glycan structures ($n = 3$), and the protein can have one, two, or all three sites occupied by a glycan. The possible number of glycoforms, N , of this protein would be $3^3 + 3[(3^1) + (3^2)] = 63$. If instead of three glycans any of the 3000 glycan structures could occupy the three glycosylation sites, then the number of possible glycoforms for this one protein would be $3000^3 + 3[(3000^1) + (3000^2)] = 27,027,009,000$. Many proteins have many more than three glycosylation sites, which greatly increases this potential complexity.

Of course, in the cell all possible glycoforms do not occur and those that do occur are not all equally probable. Rather, select combinations of glycans are attached to an individual protein forming a defined functional combination of glycoforms called a glycotype. As an example, consider immunoglobulin G, IgG. Immunoglobulins initiate the immune response toward a foreign molecule. They are 'Y'-shaped proteins; a schematic of the IgG molecule is shown in Figure 4. The arms of the 'Y' form the antigen binding portion of the protein, and, via a combinatorial genetic process, the cell can produce many variations of this structure in response to innumerable different foreign antigens. It is known as the **fragment antigen binding** (Fab) region of IgG. The stem portion of the 'Y' is, in terms of its amino acid sequence, relatively conserved compared to the Fab region and interacts with various receptors on cells,

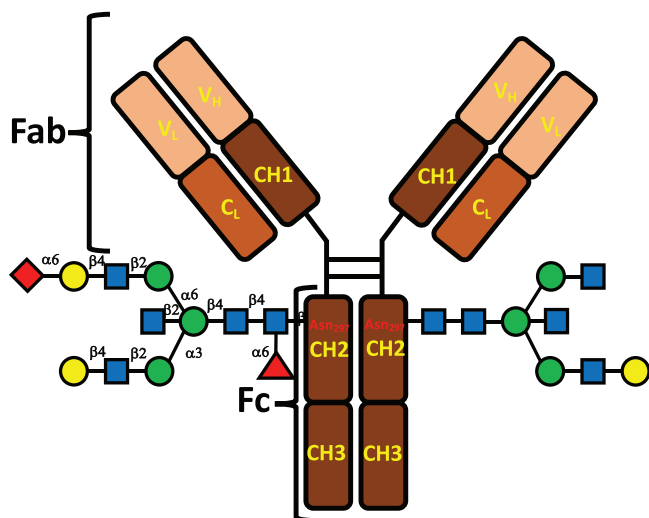


Figure 4. A schematic diagram of IgG is shown with oligosaccharides attached to the Fc portion of the heavy chain. Fc contains two glycosylation sites, and each site can remain unglycosylated or glycosylated with any one of 32 different oligosaccharide. As described in the text, it is the combination of glycoforms, the glycotype, of Fc which determines its interaction with FcRs and the resulting strength of the immune response. FcRs, Fc receptors.

e.g. leukocytes, which initiates cellular immune responses. It is known as the Fc region because when cleaved and separated from the Fab portion this stem fragment is easily crystallized. Both the Fab and Fc regions have attached carbohydrates. The Fc region has two conserved sites of glycosylation as shown in Figure 4 and is required for binding of IgG to various classes of the **Fc** receptors (FcRs) present on leukocytes, macrophages, etc. It is the Fc glycotype of IgG that determines the strength of the immune response (11). Each of the two IgG Fc glycosylation sites can be substituted by around 32 different glycan structures (11) [also see review by Lauc and Zoldoš (17)]. So, in the case of IgG, around 32 of the possible 3000 *N*-glycans are found on the Fc heavy chains to modulate and optimize (i.e. fine-tune) binding of IgG to its receptors. Thus, for IgG, the more realistic possible number of different Fc glycoforms (if all 32 glycans are equally probable) would be $32^2 + 2[32^1] = 1088$. In the case of Fc, symmetry may reduce this number, and, again, not all of the 32 selected glycan structures are equally probable, and, also, the fraction of Fc molecules which have both sites, or one site occupied may differ. In fact, about 11 glycan structures comprise 90% of those found on the IgG Fc protein. These 11 glycan structures form reproducibly different Fc glycotypes depending on the physiological state of the cell; e.g. age, or disease such as rheumatoid arthritis (11, 18, 19). The glycosylation of immunoglobulins such as IgG is only one example of the more general phenomenon that addition of glycans has profound effects on the complexity and functions of proteins, which is discussed in more detail below with regard to IgG.

How are glycan structures 'written' and 'read'?

Sugars are the alphabet of carbohydrate structures ('words'), but, as has been stated, 'compared to amino acids and nucleotides their versatility (see above description) for isomer formation (code words) is unsurpassed' (20, 21). However, as previously stated, unlike proteins, carbohydrate structures do not have a direct template in the DNA. So, how are specific carbohydrate structures synthesized, i.e. 'written', and then translated, i.e. 'read', into cellular functions?

The synthesis of a glycan structure requires numerous proteins called glycosyl transferases (GTs) and glycosyl hydrolases (GHs) as well as regulation of their expression. Additionally, just as synthesis of proteins and nucleic acids requires activated forms of the amino acids and bases, the synthesis of glycans require activated forms of sugars. This activation occurs through the formation of **nucleotide diphosphate sugars**, NDP-Sug. Different sugars are activated by different NDPs. For example, glucose (Glc), *N*-acetylglucosamine (GlcNAc), galactose (Gal), *N*-acetylgalactosamine (GalNAc), glucuronic acid (GlcA) and a number of others are activated by attachment to UDP (uridine diphosphate), mannose (Man) and fucose (Fuc) to GDP (guanosine diphosphate), and sialic acid (Sia) to CMP (cytidine monophosphate). In some instances, a sugar is also activated by attachment to a lipid, e.g. dolichol phosphate in animal cells and various forms of bactoprenyl phosphate in bacteria and archaea (22, 23). The synthesis of various NDP-Sug also requires specific metabolic pathways for each Sug. These pathways are not discussed in this paper, but add another level of complexity required for the synthesis of carbohydrate structures. The assembly of a specific carbohydrate structure then occurs through the transfer of specific sugars from their activated forms to an acceptor and is carried out by GTs. Each GT has specificity for the NDP-Sug donor as well as for the structure of the glycan (i.e. acceptor) to which the new sugar is added. All of the different GTs for a specific glycan structure must be expressed at the right time and in the correct order for a specific structure to be synthesized. In addition to GTs, the formation of the *N*-linked protein oligosaccharides also involves processing of a precursor glycan by specific GHs that also have to be expressed in the correct order, at the right time, and at the right location. In bacteria, archaea, and eukaryotes there are currently 116 and 184 families of GTs and GHs, respectively, with each family containing numerous enzymes (see www.CAZy.org) (24). The number of families and enzymes in these families is constantly increasing as research progresses. In addition to the GTs and GHs, there are numerous enzymes that modify carbohydrate structures via the addition or removal of substituent groups such as esters, phosphate, sulfate, etc. (see www.CAZy.org) (24).

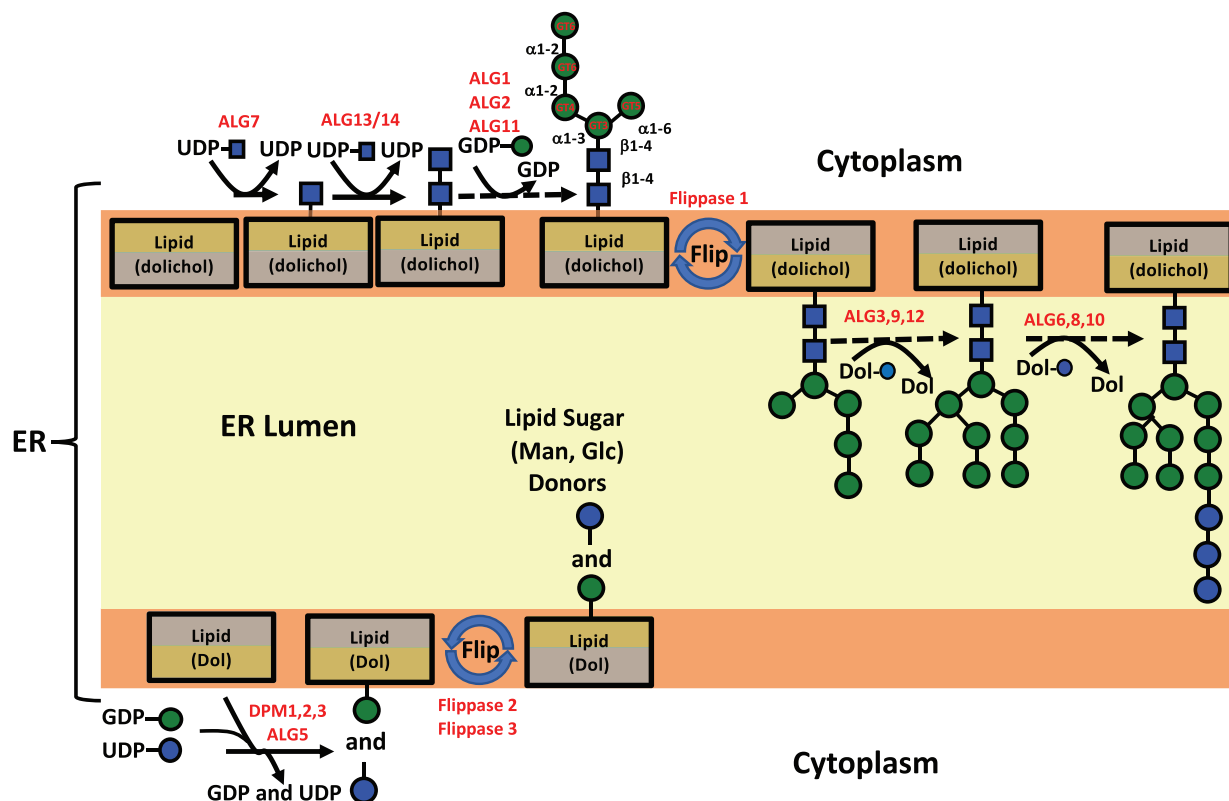


Figure 5. The synthesis of the glycoprotein glycan precursor, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$. GTs are known as Alg (asparagine-linked glycan) enzymes. The initial $\text{Man}_5\text{GlcNAc}_2\text{-Dol}$ is made on the cytoplasmic face of the ER using UDP-GlcNAc and GDP-Man as the activated sugars. This is then flipped to the lumen side of the ER membrane. Dol-PP-Man and Dol-PP-Glc are made from UDP-Glc and GDP-Man on the cytoplasmic side of the ER membrane and then flipped to the lumen side of the ER membrane where they are used as substrates to add the remaining 4 Man residues and 3 Glc residues to the glycan giving the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-Dol}$ structure. Stanley et al. (15). give a more detailed description in their chapter and in their Figure 9.3. ER, endoplasmic reticulum; Glc, glucose; GlcNAc, N-acetylglucosamine; GTs, glycosyl transferases; Man, mannose.

The synthesis of an *N*-linked oligosaccharide is illustrated in Figures 5–7. For a typical *N*-linked glycan, e.g. in the case of the *complex* structure shown in Figure 3 (without the Fuc residue), a total of 34 proteins is required: 21 GTs, 7 GHs, 3 flippases, and 3 quality control proteins. The GTs for *N*-glycan synthesis are known as **A**sn-**L**inked **g**lycosylation (Alg) enzymes. The description that follows is intentionally abbreviated, but is described in detail in numerous research publications and books, e.g. see Stanley et al. (15). Initially a common high Man structure containing 9 mannosyl, 3 glucosyl and 2 *N*-acetylglucosaminosyl residues ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) is synthesized on a dolichol carrier (Figure 5) and transferred to the -Asn-X-Ser/Thr (where X is not Pro) site of the growing peptide chain in the endoplasmic reticulum (ER) (Figure 5) by an oligosaccharide transferase OST. This structure is trimmed in the ER by glucosidases (GI and GII) to a $\text{Glc}_2\text{Man}_8\text{GlcNAc}_2$ structure in a process that also provides a quality control mechanism to efficiently produce a correctly folded glycoprotein (Figure 6). The resulting $\text{Glc}_2\text{Man}_8\text{GlcNAc}_2$ -protein is folded into the correct three dimensional form by lectins calnexin or

calreticulin complexed with a thiol-disulfide oxidoreductase called ERp57. Once a correctly folded glycoprotein is completed in the ER GII removes the single Glc and it is released from the folding complex. An additional Man residue is removed and it is transferred to the Golgi apparatus for further structural modification to produce the mature glycoprotein. If the protein is not folded correctly a GT (UGGT in Figure 6) adds a Glc residue back to the oligosaccharide and it is recycled through the folding complex. UGGT is a glucosyltransferase that specifically recognizes a mis-folded but not a correctly folded protein. If, after several tries the protein is still not correctly folded two Man residues are removed and it is exported out of the ER for degradation. The location and composition of the glycan on the protein act as a glyco-code that ensures both correct and efficient folding of the protein (25, 26). This quality control system is essential for the life of the cell, and further details are described Suzuki et al. (27) in *Essentials in Glycobiology*. Once the $\text{Man}_8\text{GlcNAc}_2$ -protein is in the Golgi it is processed to a $\text{Man}_5\text{GlcNAc}_2$ -protein structure in the *cis*-Golgi and then to a $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ -protein structure in

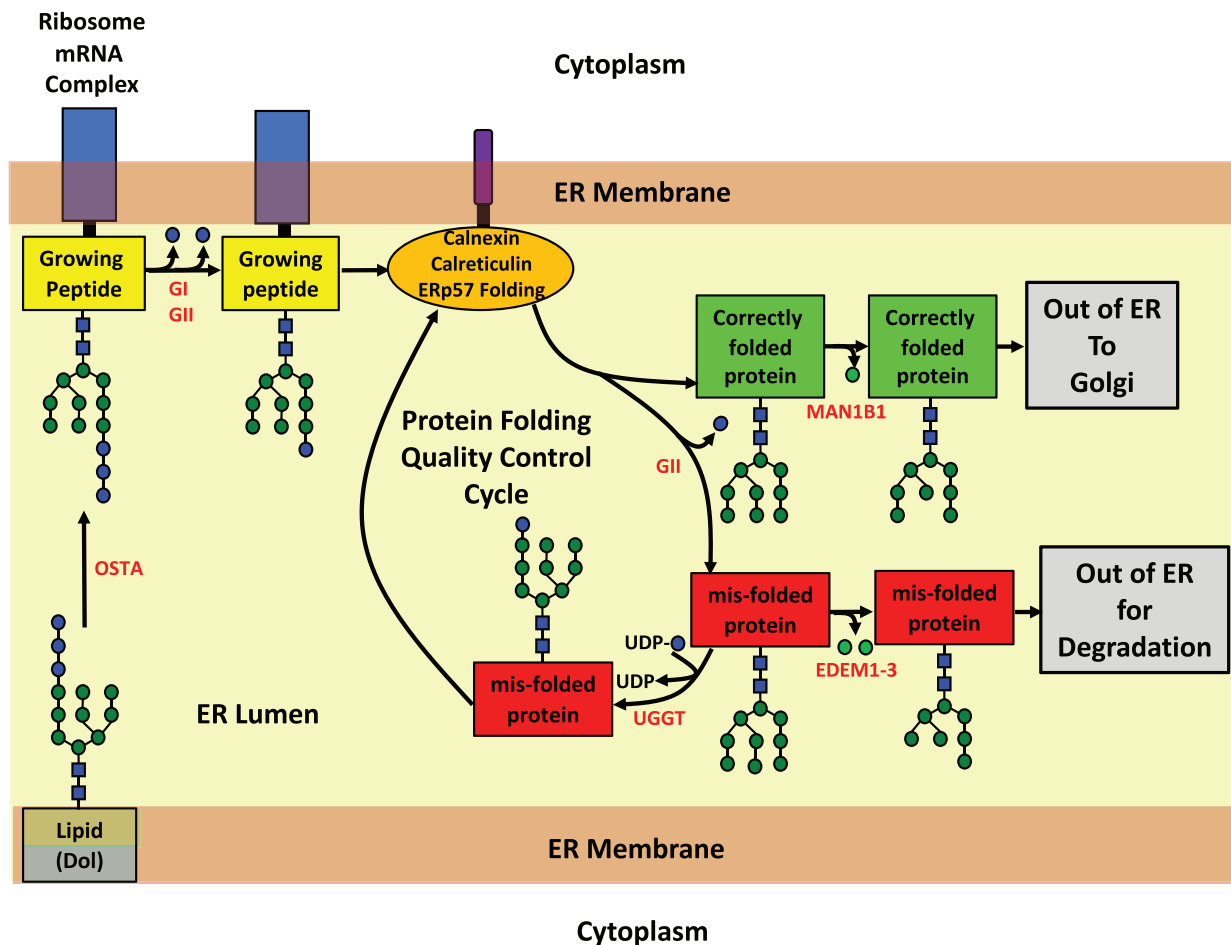


Figure 6. Attachment of the precursor glycan to a -Asn-X-Ser (where X is not Pro) site of the growing polypeptide, completion of the *N*-glycan precursor structure, and correct folding of the glycoprotein. GI and GII are glucosyl hydrolases, OSTA is the oligosaccharyl transferase, UGGT is a glucosyl transferase, MAN1B1 and EDEM1-3 are mannosyl hydrolases. Details are as described in the text. A more detailed description of this process is given by Suzuki et al. (27) and shown in their Figure 39.2. ER, endoplasmic reticulum; GI, glucosidases.

the medial-Golgi, see Figure 7. It is then processed to its' final mature structure in the *trans*-Golgi and then secreted or moved to its final location (Figure 7) (15). In the Golgi numerous glycan structures can be made via the variable addition of Gal, SA, branching GlcNAc, and Fuc residues through the regulated expression of the required various GTs.

Mechanistically this process initially involves the transport of the precursor *N*-glycoprotein from the ER to the *cis*-Golgi from ER exit sites (ERES). It has been shown that there are networks of 'tubular vessels' that extend from these sites toward the Golgi and direct the glycoprotein to the Golgi facilitated by coat protein I and II (COPI and COPII) complexes (28). The various GTs and GHs then produce the mature *N*-glycoprotein as it passes through the various Golgi cisternae. There are two models proposed for how the glycoproteins pass through the Golgi and are processed by the GT and GHs into their mature structures; the vesicular

transport model and the cisternal maturation model (29–33). In the vesicular transport model, the GTs and GHs within the various Golgi cisternae (*cis*, *medial*, and *trans*) are stable and the glycoproteins are transported to these various cisternae via COPII vesicles. In the cisternal maturation model, which is the favored model, a new *cis* cisterna is formed and as it becomes more mature it accumulates the *medial* and, then, *trans* GT and GH enzymes by retrograde passage of these enzymes from *trans* to *medial* to *cis* via COPI vesicles which then produce the mature glycoprotein; i.e. as the Golgi cisternae undergoes the *cis* to *medial* to *trans* maturation the changing GTs and GHs synthesize the mature glycoprotein within the maturing vesicle. This process also involves numerous other Golgi components, such as the conserved oligomeric Golgi complex (COG), and is both complex and precise, as described in and excellent review by Blackburn, et al. (32). The essential nature of these Golgi components for *N*-glycan synthesis is shown by the severe defects resulting

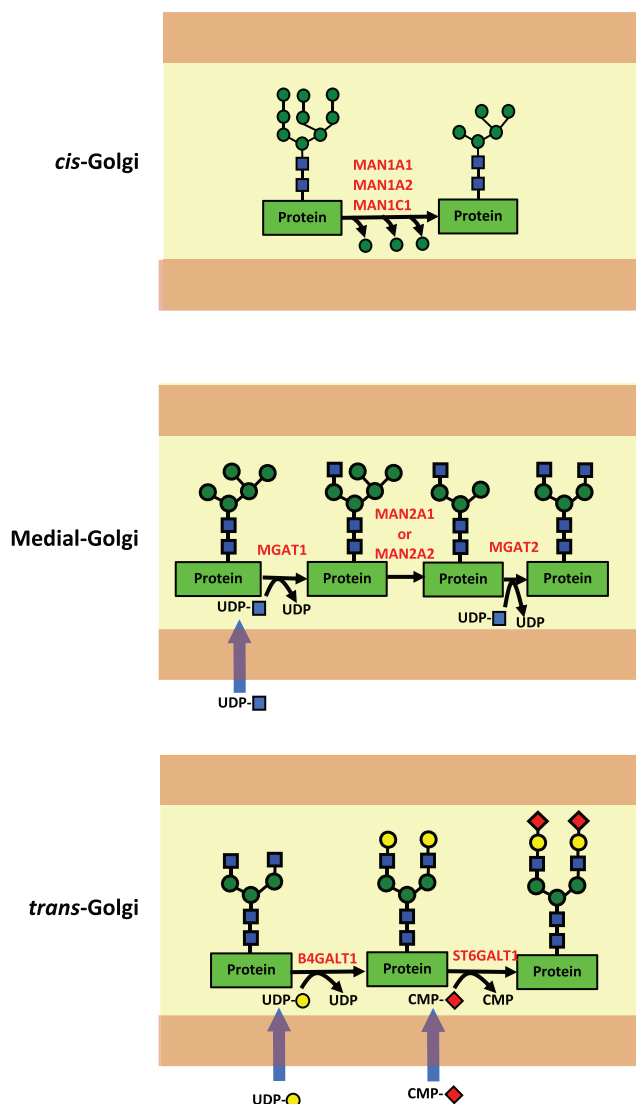


Figure 7. Processing to the mature glycan structure in the Golgi. In this figure the addition of two Gal and two SA residues is shown, but other structures can be formed by the expression of various GTs that can add other branching GlcNAc residues, Fuc residues, as well as additional combinations of Gal, LacNAc disaccharides, and SA residues. Stanley et al. (15) describe this in more detail in their chapter and in their Figure 9.4. Fuc, fucose; Gal, galactose; GlcNAc, *N*-acetylglucosamine; GTs, glycosyl transferases.

from their disruption (31, 32); e.g. COG defects are correlated with alterations to the Golgi structure resulting in defective protein glycosylation and severe developmental defects (31, 32). Thus, the synthesis of functional *N*-glycans not only requires the 34 proteins mentioned above, but also the proteins in the COPI and COPII complexes as well as many other ER and Golgi components.

In addition to *N*-linked glycans, the ER and Golgi are also the location for the synthesis of GAGs, glycolipids, O-linked

glycans, etc. It is the timing, location, and expression of the glycosyltransferases and their products which determine a functional glycan structure. Thus, a functional glycan structure is ultimately determined by the regulatory networks controlling glycosyltransferase expression. This is discussed further below.

The 'translation' of glycan structures into their cellular functions, i.e. the 'reading' of their structures, can occur by a number of mechanisms that have been categorized into several classes (34). One class is that of modulating or fine-tuning the activity of a protein. A second class is molecular mimicry, which is important in host-pathogen interactions. Finally, a third class is molecular recognition that is involved in cellular differentiation and developmental processes. Molecular recognition involves receptors for the glycoproteins. Some receptors are specific for the protein portion of the glycoprotein, while others bind the glycan; i.e. they are glycan-binding proteins (GBPs). The GBPs can be divided into two general types (35). One type, lectins, contains carbohydrate recognition domains (CRDs) which have a defined binding pocket into which a specific glycan structure fits; e.g. the calnexin and calreticulin proteins involved in the quality control process described above are considered lectins. The second type are GBPs that bind to charged glycans, e.g. GAGs, via ionic interactions. As summarized by Taylor et al. (35) in chapter 28 of '*Essentials in Glycobiology*', the interactions between the glycan structures and GBPs are required for numerous cellular functions which include the trafficking and targeting of glycoproteins both intra- and intercellularly, the clearance of glycoproteins, and cellular recognition and adhesion required for such things as the interaction between lymphocytes and endothelial cells and interactions needed for the organization of cells with each other and the cellular matrix. Lectins and GBPs are also involved in microbe-host cell interactions that determine the pathogenicity of a microbe, as well as host resistance to a microbial pathogen involving both the adaptive and innate immune responses. An example of the role of glycans in fine-tuning a protein's activity is the interaction action of the IgG protein with its receptor is described in the next section. The reader is referred to chapters 28–38 of '*Essentials in Glycobiology*' (<https://ncbi.nlm.nih.gov/books/NBK579918/>) for much more detail regarding the many necessary functions of lectins and GBPs for cellular life.

The role of glycans in fine-tuning a protein's function

As stated in the previous section, glycans can modulate or fine-tune a protein's function. This happens by forming

a specific combination of its glycoforms, known as its glycotype. The ability of a protein to exist as a glycotype allows its' functionality to be flexible to meet the changing needs of the cell that occur during development, aging, and in response to the environment. A protein's function cannot be simply on or off; it must be just-right. Directed, specific, and slight changes to glycans attached to the protein allow the formation of the correct glycotype for the just-right level of activity needed. As stated by Rademacher et al. (12) '... a glycoprotein has composite activity [of all its glycoforms], which a cell can control by varying the relative incidence of each glycoform'.

Glycosylation of the Fc portion of IgG is a very good illustration of how the function of a protein is fine-tuned. It is important that the immune system has a 'just-right' response to a potential

pathogen as too little is not sufficient to fight the pathogen, and too much can adversely affect the host. It is also necessary that, under normal physiological conditions, there is a proper balance between the pro- and anti-inflammatory responses, and that these responses are adjustable to meet changing conditions. The glycosylation of the Fc portion enables this fine-tuning by forming the appropriate combination of Fc glycoforms, i.e. its glycotype, which interact with the various FcRs on effector cells (36, 37).

The types of glycans attached to Fc and how they affect the pro- and anti-inflammatory responses is illustrated in Figures 8 and 9. As previously stated, there can be 32 or so different glycans attached to the Fc portion of the IgG protein resulting in a complex glycotype. Slight changes in the glycans can shift the response toward a pro- or anti-inflammatory

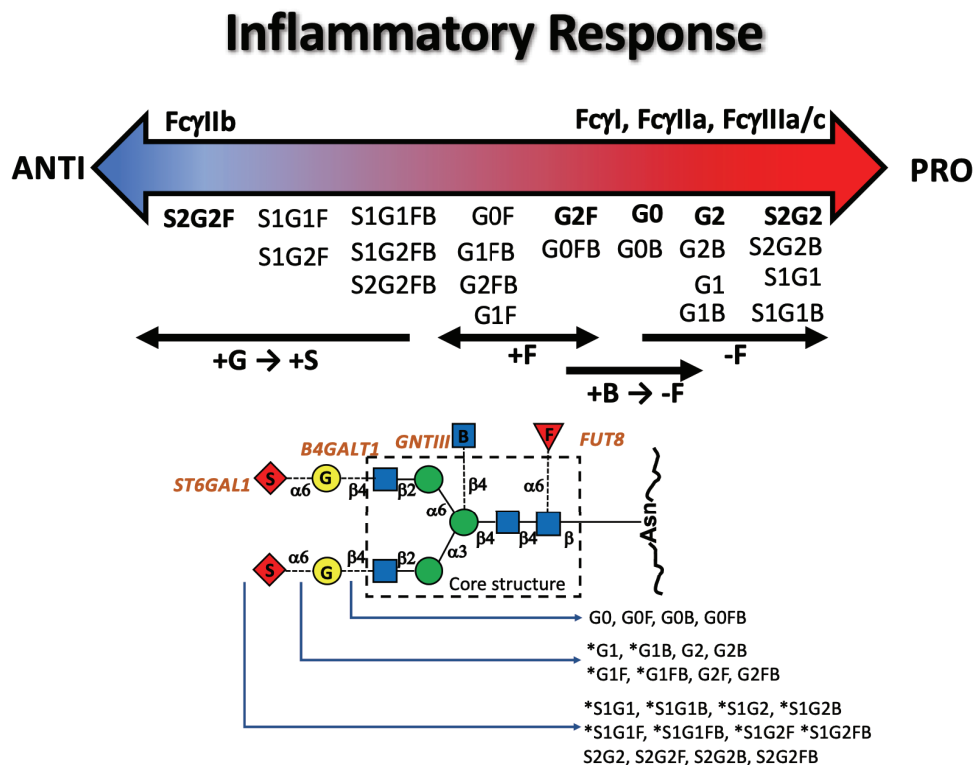


Figure 8. Fine-tuning of the inflammatory response via glycosylation of the Fc portion of IgG. The combination of Fc glycoforms interact with FcRs on the surface of various types of immune system cells to produce an inflammatory response. The strength of the response is determined by the composite interaction of these glycoforms with their receptors and this is determined by the structures of the glycans that are attached to each Fc glycoform. In this Figure, G0 stands for the GlcNAc₂Man₃GlcNAc₂-Fc structure shown in the boxed region that lacks any Gal (G) residues, G0F for that structure with the added Fuc residue (F) through the action of the GT, FUT8. B stands for the addition of a branching GlcNAc residue by GNTIII. G1 stands for a Gal₁GlcNAc₂Man₃GlcNAc₂-Protein in which one Gal (G) residue has been added by B4GALT1. The addition of B and F as well as an additional G to these structures in various combinations can also occur. Similarly, the addition of sialic acid, SA (S), to the various combinations of galactosylated structures is done by ST6GAL1. These four GTs, FUT8, GNTIII, B4GALT1, and ST6GAL1, account for the specific combinations of Fc glycoforms that can subtly shift the immune response toward a pro- or anti-inflammatory response. FcRs, Fc receptors; Fuc, fucose; Gal, galactose; GlcNAc, N-acetylglucosamine; GNTIII, N-acetylglucosaminosyl transferase III; GTs, glycosyl transferases.

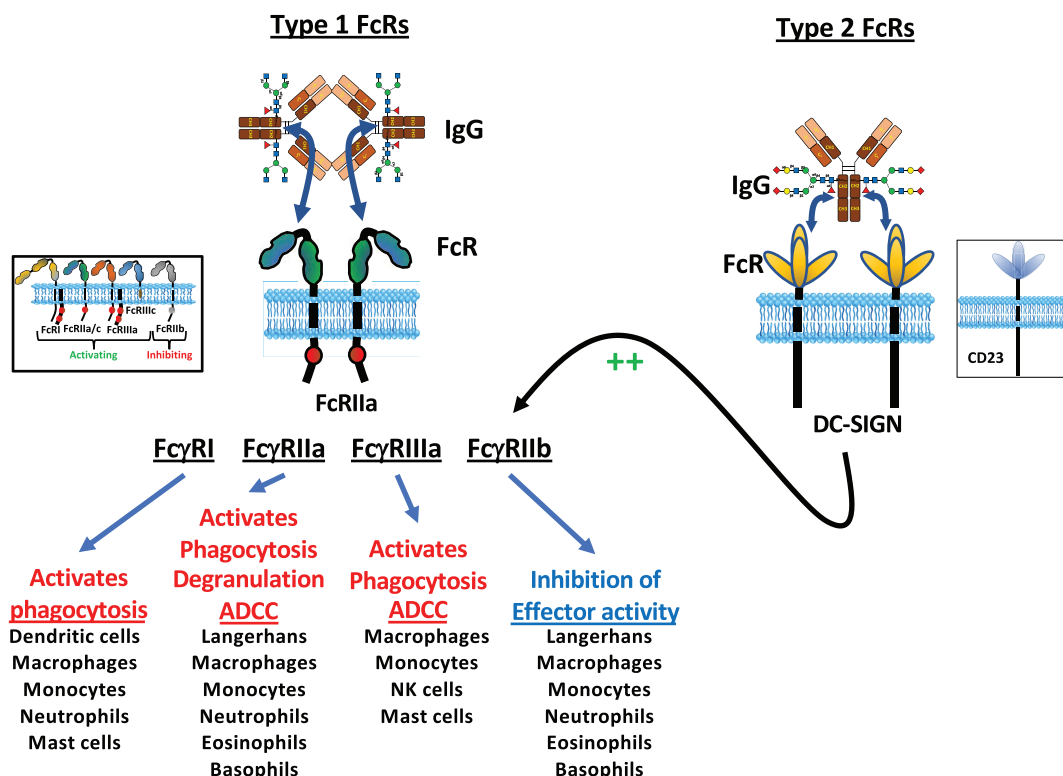


Figure 9. The Type 1 and Type 2 FcRs. Recognition of Fc by these receptors either activate or inhibit the immune response. The Type I receptor, FcγRIIb inhibits the response and its expression is increased by interaction with the Type II receptor, DC-SIGN. The Fc-FcR interactions are modulated by the pattern of Fc glycoforms (see Figure 9) as well as the pattern of glycoforms for each FcR which are also glycosylated. These interactions determine the overall activation/inhibition ratio as well as the fine-tuning of the various activation pathways (see text for details). ADCC, antibody-dependent cell cytotoxicity; FcRs, Fc receptors.

response. As indicated in Figure 8, these changes largely involve galactose (Gal, or G), sialic acid (SA, or S), branching *N*-acetylglucosamine (GlcNAc or B for branching), and fucose (Fuc or F) residues. Decreasing Fuc transferase gene (FUT8) expression shifts the glycotype toward glycoforms without Fuc resulting in increased binding of pro-inflammatory FcRs (38–42), e.g. FcγRIII, as shown in Figure 9. Increasing *N*-acetylglucosaminosyl transferase III (GNTIII) expression results in addition of a branching GlcNAc (B) prior to addition of Fuc and, thereby, reduces the effectiveness of FUT8 (18) and lowers Fuc addition, which, shifts the glycotype toward glycoforms without Fuc and toward a pro-inflammatory response (18, 37) but in a less intense manner than reduction of FUT8 activity directly. Addition of Gal (G) residues to the glycans by expression of the Gal transferase gene (B4GALT1) allows subsequent addition of SA (S) by ST6GAL1 (38, 42). This results in shifting the IgG glycotype toward sialylated glycoforms which induce expression of FcγRIIb via, possibly, interaction with the type II receptors DC-SIGN and CD23 (43), see Figure 9. The result is a shift toward an anti-inflammatory response. Thus, the pro-/anti-inflammatory balance can be subtly modulated by changes in the differential expression

of FUT8, GNTIII, B4GALT1 and ST6GAL1 that results in changes to the IgG glycotype presented to FcRs.

The importance of being able to maintain and fine-tune the pro- and anti-inflammatory response via the glycoforms that make up the IgG Fc glycotype is revealed by examination of various diseases and aging (19). For example, patients with rheumatoid arthritis, juvenile arthritis, Crohns disease, lupus, and mouse lines with autoimmune disease show increased levels of IgG Fc glycoforms without Fuc which results in an increased inflammatory response (see Figure 8) (37, 43, 44). There are also age and gender dependent differences in the IgG Fc glycotypes (44). Glycoforms without Fuc increase with age, which would imply a shift in the balance toward pro-inflammatory response. Pregnant females show increases in Gal- and SA-containing fucosylated glycoforms indicating a shift toward anti-inflammatory responses. Therapeutic antibody applications are based on shifting the anti- and pro-inflammatory balance by altering the antibody's Ig Fc glycans (44). So, the fine-tuning of IgG glycoforms is necessary to maintain the proper pro- and anti-inflammatory response balance and disruption of this balance is evident in numerous diseases, as well as aging. This brief

description of the role of glycans in the fine-tuning of IgG is just one example of many modulatory roles of glycans for proper cell functioning.

Glycoforms to glycopatterns: me or not me, that is the question

In addition to their role in fine-tuning the activity of individual proteins, glycans on the surface of cells serve as markers of tissue identity; as determinates of self vs. non-self, i.e. 'me or not me' (45). This is, of course, important in order for a kidney cell to be and function as a kidney cell, a nerve cell to function as a nerve cell and so forth. It is also important in order for the immune system to recognize potential pathogens, i.e. non-self or 'not me'; and mount an appropriate defense response. This process is accomplished by cell surface glycans forming tissue-specific patterns for individual proteins (tissue-specific glycotypes), as well as a 'global' cellular cell-type pattern (a glycopattern) comprised of the combination of glycotypes of cell surface glycoproteins, glycolipids, and GAGs. Early work by Parekh et al. (46) in 1987 showed that the *N*-linked oligosaccharides from Thy-1, a surface glycoprotein, are different in rat brain compared to thymus tissue. Over the years a very large amount of information has been gathered on tissue-, and species-specific glycan structures. These data were obtained by a variety of techniques, including mass spectrometry techniques such as MALDI-TOF MS, LC-MS, etc. in which glycopeptides are produced and analyzed, as well as by lectin microarrays in which the variation of glycan structures on a protein, or on the cell surface, can be partially characterized and visualized. Bojar et al. (47) describe the annotation, via machine-learning algorithms, of the structural affinity of lectins. These techniques allow glycan analysis of numerous specific glycoproteins as well as the determination of specific glycan structures and their locations on the protein. The results of glycoproteome studies using these techniques are deposited at the Consortium for Functional Glycomics web site, www.functionalglycomics.org. This site also lists many other glyco-databases as well as links to various analytical tools. Also, a list of the various glyco-databases can be found at the Human Glycome Project (<https://human-glycome.org/glyco-databases>), and Glyspace Alliance (<http://www.glyspace.org>). An article by Zeng et al. (48) illustrates how the authors used data deposited at the Human Glycome Project to show how the glycopattern at individual *N*-glycosylation sites varies from protein to protein, and also from the same protein isolated from different tissue types. For example, lysosomal-associated membrane protein 1 (LAMP-1), sodium/potassium-transporting ATPase subunit beta-3 (ATPB-3) and aminopeptidase each have a tissue-specific glycotype. Zeng et al. further examined the overall variation of glycan structures in glycoproteins from various tissues and

showed that different tissues express unique combinations of glycoprotein patterns, i.e. there are 'global' tissue-specific glycopatterns. The work by Zeng et al., is just one illustration that the variety of glycoforms of a protein, or proteins, is cell-, tissue-, and species-specific. It is the detailed knowledge of these glycosylation patterns (i.e. the glycoproteome) that provides functional information and also enables the development of diagnostic and therapeutic applications (49, 50).

The importance of a protein's glycotype or a cell's or tissue's global glycopattern for function is clearly revealed when examining different diseases. The Fc protein of IgG illustrates this for an individual protein. As has been already mentioned the glycosylation of Fc varies depending on age, and disease, such as rheumatoid arthritis, etc. A recent review details how the IgG glycosylation pattern (it's glycotype) changes as a function of age and age-related diseases such as Alzheimer's, type 2 diabetes, vascular disease and various cancers (19). Cellular 'global' glycopatterns are also altered in many types of cancer. Alterations to GAGs, glycolipids, O-linked as well as *N*-linked glycoprotein glycans, result in changes to the 'global' cellular glycopatterns that affect cell adhesion and promote metastasis [see review by Rodrigues et al. (51)]. For example, disruption of the T-cell glycopattern can lead to immunosuppression and affect the interaction between T-cells and cancer cells thereby promoting cancer since T-cell recognition of transformed cells is involved in the elimination of those cells [for a review, see Fernandes et al. (52)]. The Fernandes et al. (52) review describes the molecular details of the role of T-cell glycans in cancer as well as the use of this knowledge to develop anti-tumor therapies. Additionally, the alteration of glycan patterns for many types of cancers is being studied in order to understand the biology of each disease, and also to develop diagnostic tools and possible therapeutic applications; examples of these studies include those for liver (53), lung (54, 55), bladder (56), breast (57), gastric (58), kidney (59), endometrial (60), pancreatic (61, 62), esophageal (63), etc., cancers. These examples are only a few recent review articles and reflects a very small portion of a voluminous amount of work on the altered glycopatterns observed in various diseases. The readers are also referred to chapters 44–47 of '*Essentials in Glycobiology*' for more discussion regarding glycans and disease (64). It is very clear from these studies that proper glycopatterns are extremely important for the correct functioning of a cell and for the development and life of the organism.

Regulatory networks determine glycan structural/functional patterns

The cellular glycopatterns described above are due to the combination of glycoproteins (*N*- and O-linked) and other

glycans (e.g. GAGs, and glycolipids) that are present at the cell surface, i.e. the glycocalyx and extracellular matrix (ECM) of a cell. The specific patterns are determined by the expression of various glycosyltransferases (GTs and GHs, etc.), that ultimately dictate the glycopatterns. As described by Neelamegham and Mahal (65) and Dall'Olio and Trinchera (66), regulation occurs both during transcription and translation of glycosyltransferases, see Figure 10.

Transcriptional regulation of glycosyltransferase expression is accomplished by gene regulatory networks (GRNs). These

networks are composed of transcription factors (TFs) that work together with non-protein coding regions of the DNA called *cis*-regulatory modules (CRMs), and these networks can interact with one another to form interactomes that dictate the dynamic and precise glycan structures necessary for development. Developmental gene regulatory networks (dGRNs) are those that determine an organism's body plan (67–70), and since glycans are essential for tissue formation and function, it is likely that glycosyltransferase gene regulatory networks (gGRNs) are part of the dGRN family. As depicted in Figure 10, gGRNs involve TF factors interacting with CRMs of various glycosyltransferases. Combinations

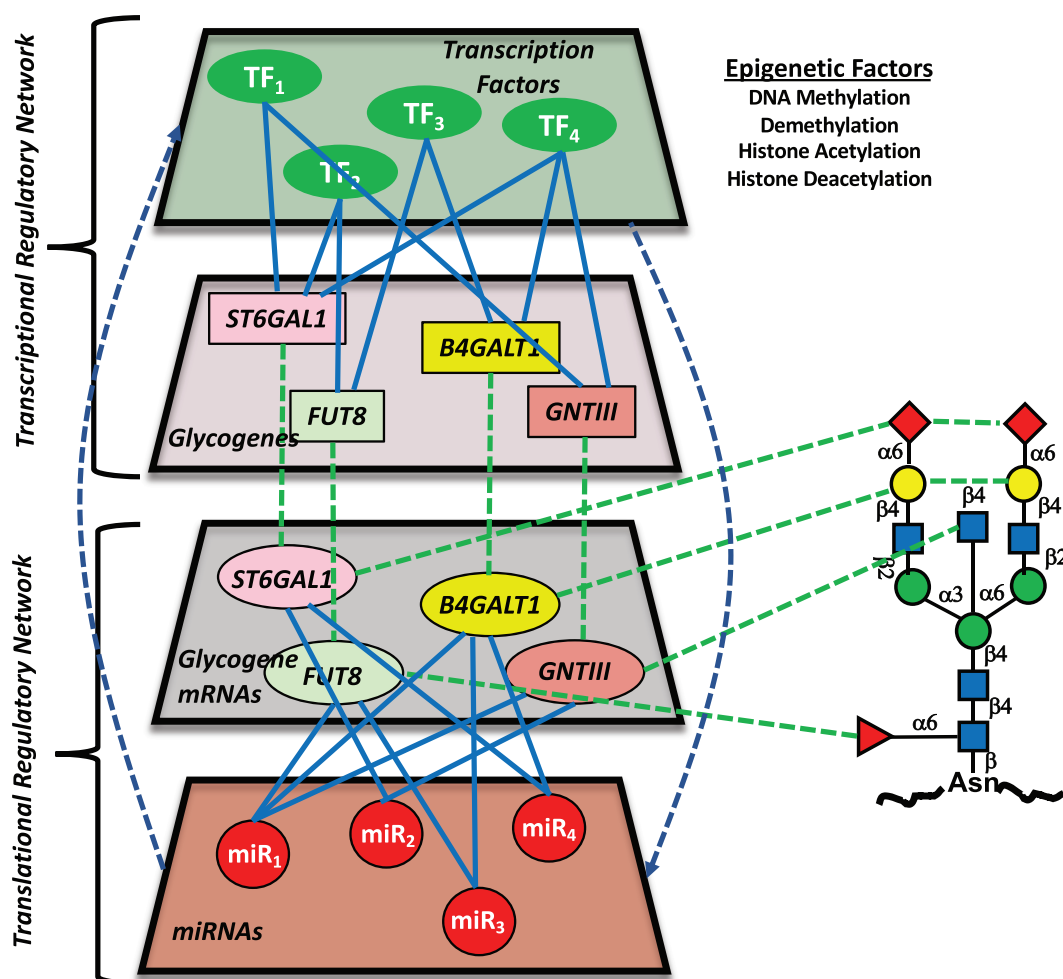


Figure 10. A schematic illustrating the transcriptional and translational regulatory networks that determine glycan structures. The actual complexity and intricacy of these networks is described in more detail by Groth et al. (67) and Thu and Mahal (68). The top portion of this figure illustrates transcriptional regulation via TFs interacting with CRMs of various glycosyltransferases and, thereby, activating their expression and resulting production of corresponding mRNAs (shown in the middle). Each TF can activate multiple genes and, therefore, different TFs may activate overlapping sets of glycosyltransferases creating various GRNs. Translational regulation occurs by the action of miRNAs (shown in the bottom portion of this figure). As with TFs, a single miRNA can affect multiple mRNAs and different miRNAs can affect overlapping sets of glycosyltransferase mRNAs. It is also likely that the miRNA and TF networks interact to produce the correct glycan structures at the right time and place. Expression of glycosyltransferases can also be affected by epigenetic factors. CRMs, *cis*-regulatory modules; GRNs, gene regulatory networks; miRNA, micro RNA; TFs, transcription factors.

of TFs form a network which results in a specific glycogene expression pattern. Figure 10 is a simple illustration, and the actual transcriptional networks (71) are far more intricate and complex.

Glycan regulation also occurs at the translational level via the action of micro RNA (miRNA) networks (65, 66, 72) as illustrated in Figure 10. As with gGRNs, the actual miRNA networks (73) are far more complex and intricate than the simple illustration shown in Figure 10. MiRNAs largely repress translation of mRNA via promoting mRNA degradation or by binding and preventing translation; however, recent work in the case of two sialyltransferases also showed that miRNA can also upregulate their activity (74). One miRNA can affect the mRNA transcripts of several glycogenes, while another miRNA can affect transcripts of a second set of glycogenes

and these two sets can overlap so that several, but not all, glycogene transcripts may be common to both. Additionally, miRNAs can vary the level at which different glycogene transcripts are affected, thereby not only determining the presence or absence of a GT, but also the level of its activity. MiRNAs could also impact regulation at the transcriptional level by affecting the translation of glycogene TF mRNAs.

Epigenetic factors can also influence these networks by altering, for example, the expression of glycogene TF genes, or directly affecting the expression of genes encoding the various GTs or GHs. It is known that epigenetic factors affect dGRNs (69), and since glycans play an essential role in tissue differentiation during development, it is reasonable to assume that gGRNs would also be affected by epigenetic factors.

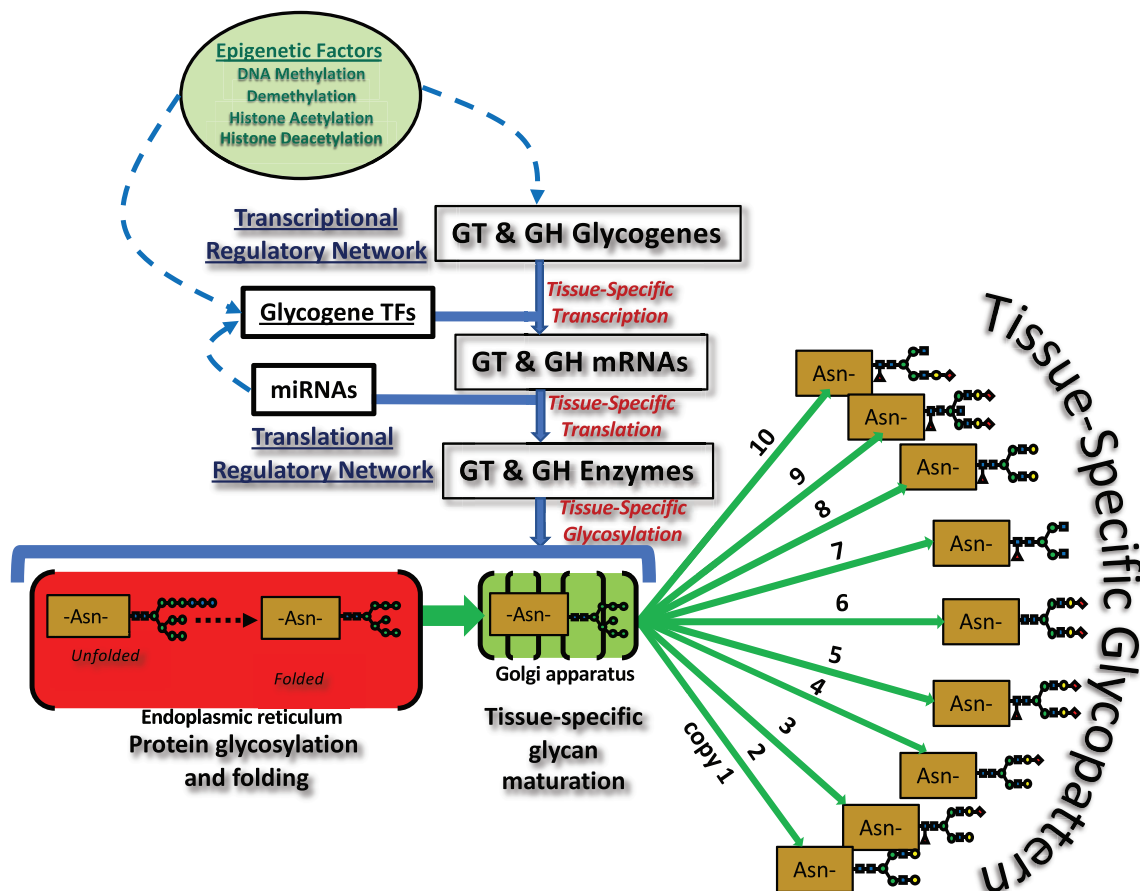


Figure 11. Production of a tissue-specific protein glycotype, and a global cellular tissue-specific glycopattern. During the synthesis of multiple copies of a single protein, each copy can have a unique glycosylation pattern called a glycoform resulting in a multiple glycoforms of that protein which is called a glycotype. The protein's glycotype is tissue-specific. This specificity is determined by regulation of the various GTs and GHs via transcriptional and translational regulatory networks as depicted in Figure 10 and described in the text. The major structural diversity of glycans is produced during their maturation in the Golgi apparatus. The combination of various protein glycotypes as well as glycolipid glycotypes, and GAGs produce a globular cellular tissue-specific glycopattern. GAGs, glycosaminoglycans; GHs, glycosyl hydrolases; GTs, glycosyl transferases; miRNA, micro RNA; TFs, transcription factors.

Figure 11 illustrates how a protein's glycosylation pattern, i.e. its glycotype, is produced. When a protein is being synthesized in the ER it is also being glycosylated with the initial proto-glycans, which also play a role in assuring the correct folding of the protein (as described above and shown in Figure 6). The glycosylated and correctly folded glycoprotein is then transferred to the Golgi apparatus, where the glycans are processed into their mature structures via the expression of appropriate GTs and GHs as determined by the gGRNs and by miRNA regulatory networks. It is in the Golgi that the large variability of glycan structures is produced. The result is a unique glycosylated protein called a protein glycoform. When another copy of this same protein is made the result is a second glycoform, etc. Thus, numerous copies of the protein result in a combination of different glycoforms of that protein which is referred to as its glycotype. It is not certain if each individual cell produces all the tissue-specific glycoforms or if each cell produces a single cell-specific glycoform and it is the combination of these different cell-specific glycoforms that result in the tissue-specific glycotype. While Figure 11 is a simple illustration for an *N*-linked glycoprotein, these regulatory networks also affect *O*-linked glycoproteins, GAGs, glycolipids, etc. resulting in global cellular and tissue-specific glycopatterns.

In order to understand the complexity of these glycome regulatory networks, extensive glyco-data bases and bioinformatic tools for investigation have been established. Several reviews outline various informatic tools used for rapid high-throughput glycome analysis (71, 75–82). A recent review by Lisacek et al. (78) 'Worldwide Glycoscience Informatics Infrastructure: The Glyspace Alliance' describes the three major glycoscience bioinformatic databases called Glyco@ExPASy, GlyCosmos, and GlyGen, the types of information they contain, and access to and analytical bioinformatic tools. The 'Glyspace Alliance' is directed toward informing and enabling users to identify and efficiently use the data bases (mentioned above) for their research. The glyco-data bases are then used with other genomic data bases to determine glycome expression/regulatory networks. For example, TF-glycome relationships have been determined in the study of various types of cancer by integrating the glycome data with **chromatin immunoprecipitation sequencing** (ChIP-Seq) and **RNA sequence** (RNA-Seq) data found in the Cistrome Cancer database. This work was able to identify the gGRNs involved in these cancers (71). Similarly, regulation of glycomes at the translational level can be evaluated by using the glyco data bases together with the **microRNA data base** (miRBase, mirbase.org) to identify miRNAs that regulate translation of glycome mRNAs (83).

Using these tools, a 2021 bioinformatic investigation by Groth et al. (71) identified 23,000 'potentially significant TF-glycome

relationships'. By examining different types of cancers, they found that these interactions occurred as 'communities' of TF-glycome relationships (71); i.e. varied and extensive networks of interacting gGRNs. This work further suggested that such gGRNs affect the synthesis of all classes of glycans; *O*- and *N*-linked glycans, GAGs, and glycolipids (71). As stated above these gGRNs, could also be affected by epigenetic factors such as histone acetylation (activates)/deacetylation (represses), histone methylation (represses)/demethylation (activates), and DNA methylation (represses)/demethylation (activates) which explains the effect that various environmental factors can have on glycan structural modification that occurs in various cancers (71, 84, 85).

It is apparent that a functional cell-specific glycopattern is the result of a hierarchy of regulatory networks acting on glycomes. The gGRNs determine the combination of GT and GH glycomes that are expressed, which, in turn result in a pattern of glycome mRNAs whose translation is subsequently determined by miRNAs networks. The result is the expression of specific GTs and GHs that give rise to the final protein glycotype or cellular glycopattern. As mentioned above gGRNs are also involved in the production of functional tissue-specific glycolipids, *O*-linked glycoproteins, hyaluronic acid (HA), and other GAGs; see the review by Basu et al. (86) describing the spatiotemporal regulation of GAGs. The final result of the gGRN and miRNA networks is a tissue-specific functional pattern of glycan structures consisting of a combination of protein glycoforms, glycolipids, and GAGs that are necessary for the development and function the organism.

The complexity of glycan evolution

Lauc and coworkers have stated that the 'simple Darwinian model of evolution fails to explain adequately ... the difference in the rate of reproduction between prokaryotic microorganisms and higher eukaryotes, plants and animals in particular' and that in order for these organisms to keep up with the 'evolutionary arms race' they need to have another mechanism to adapt that does not involve their 'reproductive capacity' (5). This mechanism is the ability to make adaptive functional changes that do not require a change in protein structure via changes to the genome by Darwinian random mutation and natural selection. One major mechanism by which this is accomplished is the post-translational modification of proteins via glycosylation (1, 5, 87) which is controlled by gGRNs and miRNA networks. Lauc and coworkers (5) further state that 'Darwinian evolution does not explain ... the shaping of development and functional integration of trillions of cells in a multicellular organism'. As has been described above, glycans are involved in both functional adaptation to disease and the

environment and also in the organization of cells into tissues and organs during development. It has been proposed, in contrast to a Darwinian model, that heritable epigenetic changes allow plant and animal cells to keep up with prokaryotes in the evolutionary arms race and that glycosylation of proteins is part of this mechanism (4, 5, 17, 87, 88). The role of epigenetic regulation of glycan structures is currently a growing area of research and has many potential applications to the treatment of numerous diseases (88).

Eric Davidson also rejects neo-Darwinism as a mechanism of animal evolution and, instead, proposes that 'the evolution of animal body plans is a system level property of the dGRNs which control the ontogeny of the body plan' (70). As mentioned above due to the role that glycans play in organizing cells into various tissues, these dGRNs would most certainly include gGRNs. Thus, understanding the origin of glycans requires understanding the origins of dGRNs as well as understanding the origins of the various GT and GH 'tools' that are directed by these networks to build the glycan structural patterns required for development and function.

In spite of these reservations, generally it has been assumed, for example, that GTs involved in *N*-glycosylation originated via the Darwinian model of random mutation, gene duplication and natural selection from ancient bacterial and/or archaeal ancestors. It is known that the *N*-glycosylation of proteins occurs in all three kingdoms of life; bacteria, archaea, and eukaryotes via the addition of mono- or oligosaccharides to the asparagine residue of proteins having the -Asn-X-Ser/Thr- consensus sequence (89–92). With regard to the initial steps of *N*-glycosylation in the ER (see Figures 5 and 6), it has been proposed that the eukaryotic *N*-glycosylation system had either bacterial or archaeal origins, with the latter as the more favored hypothesis, or that some components were of archaeal origin, more specifically proteoarchaeal, and others of bacterial origin (93–95). The presence of homologs for certain eukaryotic *N*-glycosylation GTs are cited as evidence in support of this hypothesis. For example, Alg7 (see Figure 5) has homologs in archaea (AglH) and bacteria (WecA, MraY, and TagO) (93). All of these enzymes have very similar functions; namely they are dolichol or undecaprenylphosphate GlcNAc transferases. Similarly Alg13, and 14 have homologs in archaea and bacteria and all are UDP-GlcNAc transferases (93). Other enzymes, such as Alg1, 2, 11 are in the GT1 superfamily and archaea may have GTs that are also members of this superfamily; however, their prokaryotic homologs are either unknown or not clear and have been referred to as possible eukaryotic innovations (93). Alg5 has no clear prokaryotic homolog nor does Alg3, 9 and 12 (all are mannosyl transferases) or Alg6, 8 and 10 (glucosyl transferases), all of which have been suggested

to be unique to eukaryotes (93). Finally, the OST complex which transfers the glycan from the lipid carrier to the protein, consists of eight subunits (15, 96). The catalytic subunit, Stt3, has homologs in archaea (AglB) and in one bacterial species (PglB) (93). However, the prokaryotes do not have any of the remaining 7 subunits, all of which are necessary for activity in eukaryotes (93, 96). In addition, the activity of the eukaryotic OST complex requires the precise arrangement of eight phospholipids (96). As described above most of the diversity of *N*-glycan structures, as well as GAGs and glycolipids, occurs in the Golgi and, similar to the ER enzymes described above, the evolution of the various GTs and GHs in the Golgi was examined by comparing the sequences of functional domains of the various enzymes (97). Of course, in comparing these functional domains, one is comparing sequences that catalyze similar if not identical reactions. It is proposed that these Golgi enzymes arose via duplication and divergence of these domains, gene fusion, as well as *de novo* evolution from non-coding DNA (97). All of this work raises a number of questions regarding a neo-Darwinian origin of the *N*-glycosylation machinery. First, the finding that some GT proteins have homologs in archaea and bacteria that carry out very similar, if not identical functions is not surprising, but is it evidence of a neo-Darwinian common ancestry? Is it possible that proteins could be functionally but not ancestrally related, or both functionally and ancestrally related? No details are given how one might distinguish between these possibilities. Also, evolution of these enzymes via the transfer and/or merging of functional domains and does not explain the origins of these domains. Second, quite a few of the eukaryotic GTs do not, as yet, have any clear prokaryotic homologs and have been referred to as being unique to eukaryotes or as eukaryotic 'innovations'. Can random mutation, gene duplication and natural selection gradually 'innovate' a new functional protein? In fact, it has been stated that such genes did not arise via gene duplication and rapid diversion but rather by *de novo* evolution from non-coding DNA or perhaps from alternative coding of an existing gene, and that it has happened frequently and often (98). How this might have occurred is under investigation, but a gradual process of random mutation and natural selection over a long period of time is unlikely. Finally, the transfer of the oligosaccharide from its lipid carrier to the protein requires the OST complex which consists of eight subunits and involves a specific arrangement of 8 phospholipid molecules. The Stt3 catalytic subunit of this complex has homologs in archaea and bacteria but there are no homologs of the other subunits. Since all eight subunits and a specific arrangement of eight lipids are required for activity, it seems unlikely that this complex originated via a gradual process of random mutation and natural selection. In addition, the GTs require activated NDP-Sug or lipid-Sug substrates and the metabolic pathways for their synthesis (not described in this paper) need to be

present prior to or concurrent with the relevant GTs. In the case of lipids the concept of an 'anteome' has been given to describe these 'prior to' pathways, and how they pose a problem for the neo-Darwinian explanation of sphingolipid biosynthesis (99). Finally, protein glycosylation, as well as the synthesis of GAGs and glycolipids, occurs in the ER and the Golgi requiring a mechanistic explanation for the origin of these intracellular membrane organelles as well as for how the GTs and GHs became functionally located in these intracellular structures.

Understanding the origin of glycans also requires understanding the origins of the gGRNs and miRNA regulatory networks as well as the mechanisms by which they interact since the synthesis of a functional glycan results from the spatiotemporal regulation of the various GTs and GHs which determines the functional glyco-type of numerous glycoproteins (as described above for IgG), as well as the functional patterns of the other cellular glycans [GAGs, etc (86)]. Much has been published regarding the evolution of dGRNs by Eric Davidson and others [for reviews see Refs (67–70, 100–104)] and, given the importance of glycans in development, dGRNs will likely include gGRNs. Davidson argues that dGRNs are hierarchal in that those involved in the formation of a basic body plan, 'deep' or 'kernel' dGRNs, are very resistant to change, and that any mutation is catastrophic (70, 100). Erwin and Davidson state that these 'kernel' dGRNs are responsible for the basic body plans that appeared in the Cambrian explosion (100). At this point, it is should be noted that each plant and animal cell is surrounded by glycans attached to the cell called the glycocalyx as well as glycans not attached to the cell called the ECM which consist of a complex functionally defined mixture of *N*- and *O*-linked glycoproteins, glycolipids, GAGs, etc. The glycocalyx and ECM are essential for the formation of bone and connective tissues and the different cell types during the development of an organism as well as for maintaining homeostasis of each cell type (86). Producing all the functional structures for each glycan requires the coordinated interaction of gGRN and miRNA regulatory networks, i.e. interactomes, of the various glycogenes. Multicellular organisms would not exist without the ECM and, in particular, vertebrates would not exist (87, 105). Lauc et al. (87) state that the formation of the ECM is 'one of the critical steps in the evolution of multicellularity' allowing cells to 'function as a coordinated unit and, thereby, enabling intercellular communication, cell differentiation, and development. ECM is characteristic of metazoan life forms and, therefore, would be required for the rapid appearance of the body plans that occurred in the Cambrian explosion, as well as being essential for the diversity of life stemming from that event. Thus, it is highly likely that understanding the origin of dGRNs includes understanding the origin gGRNs.

Davidson proposes that evolutionary changes after the Cambrian period occur via regulatory subcircuits that 'plug-in' to dGRNs and modify but do not change the fundamental aspects of an already existing body-plan dGRN (100). However, there is still a great mystery surrounding the origin of the 'kernel' dGRNs, which would likely include gGRNs. As stated by Stephen Meyer in his book 'Darwin's Doubt' it is unlikely that the neo-Darwinian mechanism of gradual change by random mutation and natural selection over a long period of time gave rise to new integrated dGRNs, and therefore new gGRNs, that resulted in new body plans (106). With regard to the origin of glycans in plants and animals, therefore, one could reasonably expect that there are 'kernel' gGRNs that are highly resistant to change, as well as gGRN 'subcircuits' that may be more flexible and are responsible for functional modification and fine-tuning of existing structures.

Perhaps an even more fundamental question is the identity and origin of the informational structures that guide the operation of these dGRNs during the development of an organism. It is known that cross-species cloning will not lead to successful development of an adult organism (107, 108) due to the fact that, as Denis Noble has stated, DNA sequences that 'may be very successful in one organism or environment might be lethal in another' (107), e.g. a frog genome developmentally functions in a frog egg cell and a fish genome in a fish egg cell but not vice versa. That is, there is non-genomic (i.e. epigenetic) structural information in the membrane structure and components of the egg cell that initiates and guides the dGRN-controlled expression of the genome during development. Evidence for the presence of epigenetic structural information in cellular membranes and its implications regarding neo-Darwinian evolution has been reviewed by Jonathan Wells in his article 'Membrane patterns carry ontogenetic information, that is, specified independently of DNA' (109). Noble (107) and Davidson (70) propose that the evolutionary diversity of animal species occurs at the system level, i.e. via dGRNs, and that the neo-Darwinian gradual process of random gene mutation and natural selection is an inadequate mechanism. Likewise, neo-Darwinism cannot be a mechanism for the origin of this epigenetic, i.e. non-genomic, information present in the membrane architecture of the cell that guides the initiation and operation of the dGRNs, gGRNs, and miRNA networks that direct the developmental process of an organism.

Summary

This paper has described, by giving illustrative examples, the structural, functional, and evolutionary complexity of carbohydrates.

First it was shown that carbohydrate structural complexity is due to the chemistry of sugars which allows many branching possibilities and, therefore, an oligosaccharide can form many more possible structures than an oligopeptide or oligonucleotide. This means that carbohydrates have much more potential information than either proteins or nucleic acids. This increased informational content is reflected in the amount of information that the cell devotes to the synthesis of one functional carbohydrate structure. A protein is encoded by the information in the sequence of the deoxynucleotides in a gene; i.e. the gene serves as a direct template for the protein's amino acid sequence. This is not the case for carbohydrates. The structure of an oligosaccharide that is attached to a protein or a lipid requires numerous enzymes (e.g. more than 35 GTs and GHs, etc.) to assemble a single structure in addition to all the TFs and CRMs that comprise the gGRNs, and the miRNA networks to ensure the synthesis of the correct functional glycopattern for the development and maintenance of functional tissues.

Second, the immense potential structural complexity of carbohydrates gives the cell a great deal of flexibility to both fine-tune and increase the functional range of a protein's activity, as described above for IgG, and form the pattern of the cell's glycocalyx required for development of an organism, the function and maintenance of a cell type, as well as enabling the cell to adapt to changing environments. Such changes are functionally specific and directed by the regulatory networks mentioned in the previous paragraph.

Third, the evolution of glycans is also complex. Sequence similarity between the functional domains of some animal GTs with those found in bacteria and archaea have led some to hypothesize that glycosylation mechanisms (e.g. protein *N*-glycosylation) originated via a neo-Darwinian mechanism from a last common eukaryotic ancestor (LECA) which arose from a combination of the last archaeal common ancestor (LACA) and the last bacterial common ancestor (LBCA), which, in turn, came from the cenancestor (93). However, numerous GTs involved in protein *N*-glycosylation are thought to be unique eukaryotic innovations, while others have multiple components in which only the functional domain of one component has similarity to a domain of similar function in archaea or bacteria. While it is known that proteins with similar sequences are functionally related, this does not necessarily mean they are ancestrally related. Perhaps similar sequences can be functionally but not ancestrally related, or both functionally and ancestrally related. Others, such as Lauc, Davidson, and their coworkers, have stated that animal diversity arising from the Cambrian era cannot be explained by neo-Darwinism. For example, Lauc and coworkers propose heritable epigenetic variation in which glycans would play a

major role. Others, due to the fact that development involves dGRNs suggest that animal diversity is better explained by evolutionary changes at the system level involving subGRNs that add to 'kernel' dGRNs which modify but do not make dramatic changes to existing body plans since any mutation to 'kernel' dGRNs is catastrophic to the organism. Since glycans are essential to development, and since functional glycan structures and patterns are directed by gGRNs as well as interacting miRNA networks, it seems reasonable to suggest that this system level evolution would also apply to cellular glycans. What is not explained, and what remains a mystery, is the origin of these dGRNs and gGRNs. Finally, it is not understood how the dGRNs, which would include gGRNs, are initiated and guided during development, but it has been proposed that there is non-genomic (i.e. epigenetic) informational structures in the cell, e.g. its membrane, that provide this function (109), and that the origin of this structural information cannot be explained by the gene-centric neo-Darwinian hypothesis.

The formation of glycan structures and patterns required for cellular differentiation into different tissues during development as well as for maintaining the function of each cell type involves a process that is guided by gGRNs interaction with miRNA networks; a process that appears purposeful and goal oriented and did not likely originate via a neo-Darwinian mechanism. That is, glycans are formed via a 'teleo' process. A recent special issue of the *Biological Journal of the Linnean Society* entitled 'Teleonomy in Living Systems' (vol. 139, issue 4) is dedicated to the idea of purpose in biological evolution. In an overview of that issue Vane-Wright and Corning state 'The evolved purposefulness of living systems has been both a major outcome and *an important causal factor* in the history of life on Earth' (110). It is reasonable, therefore, to suggest that the origin of glycans is also best explained as a purposeful process; i.e. a 'teleo' process. A 'teleo' process can be described as *teleologic* or *teleonomic*. The former description suggests the possibility that this purposeful process involves an 'agency' of mind or consciousness as suggested by intelligent design advocates, while the latter states that the purposefulness is self-caused and that it is inherent in the cell itself. The articles included in this special issue (vol. 139, issue 4) of the *Biological Journal of the Linnean Society* attempt to resolve the 'teleological dilemma' that the data evinces (111), and by making a case for a self-causing teleonomic process (112). Heylighen (113) in his paper on 'the meaning and origin of goal directedness' in the evolutionary process states that 'goal-directedness *does not presuppose* any mysterious force, such as intelligent design, vitalism, conscious intention, or backward causation'. However, I would suggest science should follow the evidence to the most reasonable conclusion. That the evidence of

dGRNs, which include those involved in producing functional glycan structures, supports a systems level purposeful, goal-oriented evolutionary mechanism and that, given experience shows that such processes invariably are a product of mind, intelligent design is a reasonable explanation of their origin. It is not a matter of presupposition, but of following the evidence where it leads.

References

- Cummings RD. Evolution and diversity of glycomolecules from unicellular organisms to humans. *BioCosmos*. 2024;1: 1–35. doi: 10.2478/biocosmos-2024-0001
- Walt D, Aoki-Kinoshita KF, Bertozzi CR, Boons G-J, Darvill A, Hart G, et al. *Transforming glycoscience: a roadmap for the future*. Vol. 2012. Washington, DC: The National Academy Press; 2012. p.191.
- Varki A, Kornfeld S. Historical background and overview. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.1–20.
- Zoldos V, Horvat T, Lauc G. Glycomics meets genomics, epigenomics and other high throughput omics for system biology studies. *Current Opinion in Chemical Biology*. 2013;17(1): 34–40. doi: 10.1016/j.cbpa.2012.12.007
- Lauc G, Vojta A, Zoldos V. Epigenetic regulation of glycosylation is the quantum mechanics of biology. *Biochimica et Biophysica Acta*. 2014;1840(1): 65–70. doi: 10.1016/j.bbagen.2013.08.017
- Sackstein R, Hoffmeister KM, Stowell SR, Kinoshita T, Varki A, Freeze HH. Glycans in acquired human diseases. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.615–630.
- Kristic J, Lauc G. Ubiquitous importance of protein glycosylation. *Methods in Molecular Biology (Clifton, NJ)*. 2017;1503: 1–12. doi: 10.1007/978-1-4939-6493-2_1
- Lauc G, Zoldos V. Epigenetic regulation of glycosylation could be a mechanism used by complex organisms to compete with microbes on an evolutionary scale. *Medical Hypotheses*. 2009;73: 510–512. doi: 10.1016/j.mehy.2009.03.059
- Freeze HH, Kinoshita T, Varki A. Chapter 46 glycans in acquired human diseases. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Press; 2017. p. 521–526.
- Varki A, Kannagi R, Toole B, Stanley P. Chapter 47 glycosylation changes in cancer. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Press; 2017. p. 597–609.
- Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual Review of Immunology*. 2007;25: 21–50. doi: 10.1146/annurev.immunol.25.022106.141702
- Rademacher TW, Parekh RB, Dwek RA. Glycobiology. *Annual Review of Biochemistry*. 1988;57: 785–838. doi: 10.1146/annurev.bi.57.070188.004033
- Werz DB, Ranzinger R, Herget S, Adibekian A, von der Lieth CW, Seeberger PH. Exploring the structural diversity of mammalian carbohydrates ('glycospace') by statistical databank analysis. *ACS Chemical Biology*. 2007;2(10): 685–691. doi: 10.1021/cb700178s
- Laine RA. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*. 1994;4(6): 759–767. doi: 10.1093/glycob/4.6.759
- Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.103–116.
- Stanley P, Wuhrer M, Lauc G, Stowell SR, Cummings RD. Structures common to different glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.165–184.
- Lauc G, Zoldos V. Protein glycosylation – an evolutionary cross-road between genes and environment. *Molecular bioSystems*. 2010;6(12): 2373–2379. doi: 10.1039/c0mb00067a
- Wang TT. IgG Fc glycosylation in human immunity. *Current Topics in Microbiology and Immunology*. 2019;423: 63–75. doi: 10.1007/82_2019_152
- Kristic J, Lauc G, Pezer M. Immunoglobulin G glycans – biomarkers and molecular effectors of aging. *Clinica Chimica Acta*. 2022;535: 30–45. doi: 10.1016/j.ccca.2022.08.006
- André S, Kaltner H, Manning JC, Murphy PV, Gabius HJ. Lectins: getting familiar with translators of the sugar code. *Molecules (Basel, Switzerland)*. 2015;20(2): 1788–1823. doi: 10.3390/molecules20021788
- Rudiger H, Gabius HJ. The biochemical basis and coding capacity of the sugar code. In: Gabius HJ. (ed.) *The sugar code: fundamentals of glycosciences*. Weinheim: Wiley-Blackwell; 2009. p. 3–14.
- Lombard J. Early evolution of polyisoprenol biosynthesis and the origin of cell walls. *PeerJ*. 2016;4: e2626. doi: 10.7717/peerj.2626
- Eichler J, Guan Z. Lipid sugar carriers at the extremes: the phosphodolichols archaea use in N-glycosylation. *Biochimica et biophysica acta. Molecular and Cell Biology of Lipids*. 2017;1862(6): 589–599. doi: 10.1016/j.bbalip.2017.03.005
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database

- (CAZy) in 2013. *Nucleic Acids Research*. 2014;42(Database issue): D490–D495. doi: 10.1093/nar/gkt1178
25. Guay KP, Ke H, Canniff NP, George GT, Eyles SJ, Mariappan M, et al. ER chaperones use a protein folding and quality control glyco-code. *Molecular Cell*. 2023;83(24): 4524–4537.e5. doi: 10.1016/j.molcel.2023.11.006
26. Caramelo JJ, Parodi AJ. A sweet code for glycoprotein folding. *FEBS Letters*. 2015;589(22): 3379–3387. doi: 10.1016/j.febslet.2015.07.021
27. Suzuki T, Cummings RD, Aebi M, Parodi A. Glycans in glycoprotein quality control. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.529–538.
28. Weigel AV, Chang CL, Shtengel G, Xu CS, Hoffman DP, Freeman M, et al. ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER. *Cell*. 2021;184(9): 2412–2429.e16. doi: 10.1016/j.cell.2021.03.035
29. Colley KJ, Varki A, Haltiwanger RS, Kinoshita T. Cellular organization of glycosylation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.43–52.
30. Connerly PL. How do proteins move through the Golgi apparatus? *Nature Education*. 2010;3(9): 60–66. doi: www.nature.com/scitable/topicpage/how-do-proteins-move-through-the-golgi-14397318/
31. D'Souza Z, Blackburn JB, Kudlyk T, Pokrovskaya ID, Lupashin VV. Defects in COG-mediated Golgi trafficking alter endo-lysosomal system in human cells. *Frontiers in Cell and Developmental Biology*. 2019;7: 118. doi: 10.3389/fcell.2019.00118
32. Blackburn JB, D'Souza Z, Lupashin VV. Maintaining order: COG complex controls Golgi trafficking, processing, and sorting. *FEBS Letters*. 2019;593(17): 2466–2487. doi: 10.1002/1873-3468.13570
33. Toustou C, Walet-Balieu ML, Kiefer-Meyer MC, Houdou M, Lerouge P, Foulquier F, et al. Towards understanding the extensive diversity of protein N-glycan structures in eukaryotes. *Biological Reviews of the Cambridge Philosophical Society*. 2022;97(2): 732–748. doi: 10.1111/brv.12820
34. Gagneux P, Hennot T, Varki A. Biological functions of glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.79–92.
35. Taylor ME, Drickamer K, Imberty A, van Kooyk Y, Schnaar RL, Etzler ME, et al. Discovery and classification of glycan-binding proteins. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds.) *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.375–386.
36. Bournazos S, Ravetch JV. Fcγ receptor pathways during active and passive immunization. *Immunological Reviews*. 2015;268(1): 88–103. doi: 10.1111/imr.12343
37. Lux A, Nimmerjahn F. Impact of differential glycosylation on IgG activity. *Advances in Experimental Medicine and Biology*. 2011;780: 113–124. doi: 10.1007/978-1-4419-5632-3_10
38. Bournazos S, Ravetch JV. Diversification of IgG effector functions. *International Immunology*. 2017;29(7): 303–310. doi: 10.1093/intimm/dxx025
39. Bournazos S, Ravetch JV. Fcγ receptor function and the design of vaccination strategies. *Immunity*. 2017;47(2): 224–233. doi: 10.1016/j.immuni.2017.07.009
40. Yamaguchi Y, Nishimura M, Nagano M, Yagi H, Sasakawa H, Uchida K, et al. Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. *Biochimica et Biophysica Acta*. 2006;1760(4): 693–700. doi: 10.1016/j.bbagen.2005.10.002
41. Yamaguchi Y, Barb AW. A synopsis of recent developments defining how N-glycosylation impacts immunoglobulin G structure and function. *Glycobiology*. 2020;30(4): 214–225. doi: 10.1093/glycob/cwz068
42. Sondermann P, Pincetic A, Maamary J, Lammens K, Ravetch JV. General mechanism for modulating immunoglobulin effector function. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(24): 9868–9872. doi: 10.1073/pnas.1307864110
43. Kronimus Y, Dodel R, Galuska SP, Neumann S. IgG Fc N-glycosylation: alterations in neurologic diseases and potential therapeutic target? *Journal of Autoimmunity*. 2019;96: 14–23. doi: 10.1016/j.jaut.2018.10.006
44. Maverakis E, Kim K, Shimoda M, Gershwin ME, Patel F, Wilken R, et al. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *Journal of Autoimmunity*. 2015;57: 1–13. doi: 10.1016/j.jaut.2014.12.002
45. Medzhitov R, Janeway CA Jr. Decoding the patterns of self and nonself by the innate immune system. *Science (New York, NY)*. 2002;296(5566): 298–300. doi: 10.1126/science.1068883
46. Parekh RB, Tse AGD, Dwek RA, Williams AF, Rademacher TW. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *The EMBO Journal*. 1987;6: 1233–1244. doi: 10.1002/j.1460-2075.1987.tb02359.x
47. Bojar D, Meche L, Meng G, Eng W, Smith DF, Cummings RD, et al. A useful guide to lectin binding: machine-learning directed annotation of 57 unique lectin specificities. *ACS Chemical Biology*. 2022;17(11): 2993–3012. doi: 10.1021/acscchembio.1c00689
48. Zeng X, Novotny MV, Clemmer DE, Trinidad JC. A graphical representation of glycan heterogeneity. *Glycobiology*. 2022;32(3): 201–207. doi: 10.1093/glycob/cwab116
49. Lis H, Sharon N. Protein glycosylation. Structural and functional aspects. *European Journal of Biochemistry/FEBS*. 1993;218(1): 1–27. doi: 10.1111/j.1432-1033.1993.tb18347.x

50. Sharon N, Lis H. Carbohydrates in cell recognition. *Scientific American*. 1993;268(1): 82–89. doi: 10.1038/scientificamerican0193-82
51. Rodrigues JG, Balmaña M, Macedo JA, Poças J, Fernandes Â, de-Freitas-Junior JCM, et al. Glycosylation in cancer: selected roles in tumour progression, immune modulation and metastasis. *Cellular Immunology*. 2018;333: 46–57. doi: 10.1016/j.cellimm.2018.03.007
52. Fernandes Â, Azevedo CM, Silva MC, Faria G, Dantas CS, Vicente MM, et al. Glycans as shapers of tumour microenvironment: a sweet driver of T-cell-mediated anti-tumour immune response. *Immunology*. 2023;168(2): 217–232. doi: 10.1111/imm.13494
53. Kong Y, Chen H, Chen M, Li Y, Li J, Liu Q, et al. Abnormal ECA-binding membrane glycans and galactosylated CAT and P4HB in lesion tissues as potential biomarkers for hepatocellular carcinoma diagnosis. *Frontiers in Oncology*. 2022;12: 855952. doi: 10.3389/fonc.2022.855952
54. Ma T, Wang Y, Jia L, Shu J, Yu H, Du H, et al. Increased expression of core-fucosylated glycans in human lung squamous cell carcinoma. *RSC Advances*. 2019;9(38): 22064–22073. doi: 10.1039/C9RA04341A
55. Liu L, Li D, Shu J, Wang L, Zhang F, Zhang C, et al. Protein glycopatterns in bronchoalveolar lavage fluid as novel potential biomarkers for diagnosis of lung cancer. *Frontiers in Oncology*. 2020;10: 568433. doi: 10.3389/fonc.2020.568433
56. Yang G, Tan Z, Lu W, Guo J, Yu H, Yu J, et al. Quantitative glycome analysis of N-glycan patterns in bladder cancer vs normal bladder cells using an integrated strategy. *Journal of Proteome Research*. 2015;14(2): 639–653. doi: 10.1021/pr5006026
57. Yang J, Liu X, Shu J, Hou Y, Chen M, Yu H, et al. Abnormal Galactosylated-Glycans recognized by Bandeiraea Simplicifolia Lectin I in saliva of patients with breast Cancer. *Glycoconjugate Journal*. 2020;37(3): 373–394. doi: 10.1007/s10719-020-09910-6
58. Yu H, Li X, Chen M, Zhang F, Liu X, Yu J, et al. Integrated glycome strategy for characterization of aberrant LacNAc contained N-glycans associated with gastric carcinoma. *Frontiers in Oncology*. 2019;9: 636. doi: 10.3389/fonc.2019.00636
59. Zhu H, Liu M, Yu H, Liu X, Zhong Y, Shu J, et al. Glycopatterns of urinary protein as new potential diagnosis indicators for diabetic nephropathy. *Journal of Diabetes Research*. 2017;2017: 5728087. doi: 10.1155/2017/5728087
60. Pu C, Biyuan, Xu K, Zhao Y. Glycosylation and its research progress in endometrial cancer. *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*. 2022;24(10): 1865–1880. doi: 10.1007/s12094-022-02858-z
61. Marciel MP, Haldar B, Hwang J, Bhalerao N, Bellis SL. Role of tumor cell sialylation in pancreatic cancer progression. *Advances in Cancer Research*. 2023;157: 123–155. doi: 10.1016/bs.acr.2022.07.003
62. Lumibao JC, Tremblay JR, Hsu J, Engle DD. Altered glycosylation in pancreatic cancer and beyond. *The Journal of Experimental Medicine*. 2022;219(6): e20211505. doi: 10.1084/jem.20211505
63. Godefda TM, Derks S, Thijssen V. Galectins in esophageal cancer: current knowledge and future perspectives. *Cancers (Basel)*. 2022;14(23): 5790. doi: 10.3390/cancers14235790
64. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
65. Neelamegham S, Mahal LK. Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. *Current Opinion in Structural Biology*. 2016;40: 145–152. doi: 10.1016/j.sbi.2016.09.013
66. Dall'Olio F, Trinchera M. Epigenetic bases of aberrant glycosylation in cancer. *International Journal of Molecular Sciences*. 2017;18(5): 998. doi: 10.3390/ijms18050998
67. Groth T, Gunawan R, Neelamegham S. A systems-based framework to computationally describe putative transcription factors and signaling pathways regulating glycan biosynthesis. *Beilstein Journal of Organic Chemistry*. 2021;17: 1712–1724. doi: 10.3762/bjoc.17.119
68. Thu CT, Mahal LK. Sweet control: microRNA regulation of the glycome. *Biochemistry*. 2020;59(34): 3098–3110. doi: 10.1021/acs.biochem.9b00784
69. Halfon MS. Perspectives on gene regulatory network evolution. *Trends in Genetics: TIG*. 2017;33(7): 436–447. doi: 10.1016/j.tig.2017.04.005
70. Schember I, Halfon MS. Common themes and future challenges in understanding gene regulatory network evolution. *Cells*. 2022;11(3): 510. doi: 10.3390/cells11030510
71. Frankel N. Multiple layers of complexity in cis-regulatory regions of developmental genes. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*. 2012;241(12): 1857–1866. doi: 10.1002/dvdy.23871
72. Davidson EH. Evolutionary bioscience as regulatory systems biology. *Developmental Biology*. 2011;357(1): 35–40. doi: 10.1016/j.ydbio.2011.02.004
73. Agrawal P, Kurcon T, Pilobello KT, Rakus JF, Koppolu S, Liu Z, et al. Mapping posttranscriptional regulation of the human glycome uncovers microRNA defining the glycode. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(11): 4338–4343. doi: 10.1073/pnas.1321524111
74. Jame-Chenarboo F, Ng HH, Macdonald D, Mahal LK. High-throughput analysis reveals miRNA upregulating alpha-2,6-sialic acid through direct miRNA-mRNA interactions. *ACS Central Science*. 2022;8(11): 1527–1536. doi: 10.1021/acscentsci.2c00748
75. Groth T, Diehl AD, Gunawan R, Neelamegham S. GlycoEnzOnto: a GlycoEnzyme pathway and molecular function ontology. *Bioinformatics*. 2022;38(24): 5413–5420. doi: 10.1093/bioinformatics/btac704

76. Kelkar A, Groth T, Neelamegham S. Forward genetic screens of human glycosylation pathways using the GlycoGene CRISPR library. *Current Protocols*. 2022;2(4): e402. doi: 10.1002/cpz1.402
77. Stewart N, Wisnovsky S. Bridging glycomics and genomics: new uses of functional genetics in the study of cellular glycosylation. *Frontiers in Molecular Biosciences*. 2022;9: 934584. doi: 10.3389/fmolb.2022.934584
78. Lisacek F, Tiemeyer M, Mazumder R, Aoki-Kinoshita KF. Worldwide glycoscience informatics infrastructure: the GlySpace alliance. *Journal of the American Chemical Society Au*. 2023;3(1): 4–12. doi: 10.1021/jacsau.2c00477
79. Bojar D, Lisacek F. Glycoinformatics in the artificial intelligence era. *Chemical Reviews*. 2022;122(20): 15971–15988. doi: 10.1021/acs.chemrev.2c00110
80. Hayes C, Daponte V, Mariethoz J, Lisacek F. This is GlycoQL. *Bioinformatics (Oxford, England)*. 2022;38(Suppl_2): ii162–ii167. doi: 10.1093/bioinformatics/btac500
81. Mariethoz J, Alocci D, Karlsson NG, Packer NH, Lisacek F. An interactive view of glycosylation. *Methods in Molecular Biology (Clifton, NJ)*. 2022;2370: 41–65. doi: 10.1007/978-1-0716-1685-7_3
82. Aoki-Kinoshita KF, Campbell MP, Lisacek F, Neelamegham S, York WS, Packer NH. Glycoinformatics. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p. 705–718.
83. Kasper BT, Koppolu S, Mahal LK. Insights into miRNA regulation of the human glycome. *Biochemical and Biophysical Research Communications*. 2014;445(4): 774–779. doi: 10.1016/j.bbrc.2014.01.034
84. Indelicato R, Trinchera M. Epigenetic regulation of glycosylation in cancer and other diseases. *International Journal of Molecular Sciences*. 2021;22(6): 2980. doi: 10.3390/ijms22062980
85. Indelicato R, Trinchera M. Epigenetic regulation of glycosylation. *Advances in Experimental Medicine and Biology*. 2021;1325: 173–186. doi: 10.1007/978-3-030-70115-4_8
86. Basu A, Patel NG, Nicholson ED, Weiss RJ. Spatiotemporal diversity and regulation of glycosaminoglycans in cell homeostasis and human disease. *American Journal of Physiology. Cell Physiology*. 2022;322(5): C849–C864. doi: 10.1152/ajpcell.00085.2022
87. Lauc G, Kristic J, Zoldos V. Glycans – the third revolution in evolution. *Frontiers in Genetics*. 2014;5: 145. doi: 10.3389/fgene.2014.00145
88. Stambuk T, Klasic M, Zoldos V, Lauc G. N-glycans as functional effectors of genetic and epigenetic disease risk. *Molecular Aspects of Medicine*. 2021;79: 100891. doi: 10.1016/j.mam.2020.100891
89. Nothaft H, Szymanski CM. New discoveries in bacterial N-glycosylation to expand the synthetic biology toolbox. *Current Opinion in Chemical Biology*. 2019;53: 16–24. doi: 10.1016/j.cbpa.2019.05.032
90. Valguarnera E, Kinsella RL, Feldman MF. Sugar and spice make bacteria not nice: protein glycosylation and its influence in pathogenesis. *Journal of Molecular Biology*. 2016;428(16): 3206–3220. doi: 10.1016/j.jmb.2016.04.013
91. Nothaft H, Szymanski CM. Bacterial protein N-glycosylation: new perspectives and applications. *Journal of Biological Chemistry*. 2013;288(10): 6912–6920. doi: 10.1074/jbc.R112.417857
92. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, et al. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science (New York, NY)*. 2002;298(5599): 1790–1793. doi: 10.1126/science.298.5599.1790
93. Lombard J. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biology Direct*. 2016;11: 36. doi: 10.1186/s13062-016-0137-2
94. Meyer BH, Albers SV, Eichler J, Aebi M. Archaea. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Press; 2022. p.297–306.
95. Eichler J. N-glycosylation in Archaea-New roles for an ancient posttranslational modification. *Molecular Microbiology*. 2020;114(5): 735–741. doi: 10.1111/mmi.14569
96. Bai L, Wang T, Zhao G, Kovach A, Li H. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature*. 2018;555(7696): 328–333. doi: 10.1038/nature25755
97. Wang P, Wang H, Gai J, Tian X, Zhang X, Lv Y, et al. Evolution of protein N-glycosylation process in Golgi apparatus which shapes diversity of protein N-glycan structures in plants, animals and fungi. *Scientific Reports*. 2017;7: 40301. doi: 10.1038/srep40301
98. Neme R, Tautz D. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics*. 2013;14: 117. doi: 10.1186/1471-2164-14-117
99. Santos TCB, Dingjan T, Futerman AH. The sphingolipid anteome: implications for evolution of the sphingolipid metabolic pathway. *FEBS Letters*. 2022;596(18): 2345–2363. doi: 10.1002/1873-3468.14457
100. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. *Science (New York, NY)*. 2006;311(5762): 796–800. doi: 10.1126/science.1113832
101. Davidson EH, Erwin DH. Evolutionary innovation and stability in animal gene networks. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*. 2010;314(3): 182–186. doi: 10.1002/jez.b.21329
102. Peter IS, Davidson EH. Evolution of gene regulatory networks controlling body plan development. *Cell*. 2011;144(6): 970–985. doi: 10.1016/j.cell.2011.02.017
103. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics*. 2009;10(2): 141–148. doi: 10.1038/nrg2499
104. Frankel N, Wang S, Stern DL. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic

- evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(51): 20975–20979. doi: 10.1073/pnas.1207715109
105. Hynes RO. The evolution of metazoan extracellular matrix. *The Journal of Cell Biology*. 2012;196(6): 671–679. doi: 10.1083/jcb.201109041
106. Meyer SM. Darwin's Doubt: the explosive origin of animal life and the case for intelligent design. New York, NY: Harper Collins; 2013. p.498.
107. Noble D. Neo-Darwinism, the modern synthesis and selfish genes: are they of use in physiology? *The Journal of Physiology*. 2011;589(Pt 5): 1007–1015. doi: 10.1113/jphysiol.2010.201384
108. Noble D, Noble R. Origins and demise of selfish gene theory. *Theoretical Biology Forum*. 2022;115(1–2): 29–43. doi: 10.19272/202211402003
109. Wells J. Membrane patterns carry ontogenetic information that is specified independently of DNA. *BIO-Complexity*. 2014;2014(2): 1–28. doi: 10.5048/BIO-C.2014.2
110. Vane-Wright RI, Corning PA. Teleonomy in living systems: an overview. *Biological Journal of the Linnean Society*. 2023;139: 341–356. doi: 10.1093/biolinnean/blad037
111. Babcock G, McShea DW. Resolving teleology's false dilemma. *Biological Journal of the Linnean Society*. 2023;139(4): 415–432. doi: 10.1093/biolinnean/blac058
112. Gontier N. Teleonomy as a problem of self-causation. *Biological Journal of the Linnean Society*. 2023;139(4): 388–414. doi: 10.1093/biolinnean/blac111
113. Heylighen F. The meaning and origin of goal-directedness: a dynamical systems perspective. *Biological Journal of the Linnean Society*. 2023;139: 370–389. doi: 10.1093/biolinnean/blac060