BioMol Concepts, Vol. 1 (2010), pp. 253–261 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BMC.2010.027

Review

Epigenetic regulation of protein glycosylation

Vlatka Zoldoš[1], Srđana Grgurević[2] and Gordan Lauc[2,3,*]

[1] Department of Molecular Biology, Faculty of Science, University of Zagreb, Horvatovac 102, Zagreb, Croatia
[2] Glycobiology Division, Genos Ltd., Planinska 1, 10000 Zagreb, Croatia
[3] Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovačića 1, 10000 Zagreb, Croatia

* Corresponding author
e-mail: glauc@pharma.hr

Abstract

Protein N-glycosylation is an ancient metabolic pathway that still exists in all three domains of life (Archaea, Bacteria and Eukarya). The covalent addition of one or more complex oligosaccharides (glycans) to protein backbones greatly diversifies their structures and makes the glycoproteome several orders of magnitude more complex than the proteome itself. Contrary to polypeptides, which are defined by a sequence of nucleotides in the corresponding genes, the glycan part of glycoproteins are encoded in a complex dynamic network of hundreds of proteins, whereby activity is defined by both genetic sequence and the regulation of gene expression. Owing to the complex nature of their biosynthesis, glycans are particularly versatile and apparently a large part of human variation derives from differences in protein glycosylation. Composition of the individual glycome appears to be rather stable, and thus differences in the pattern of glycan synthesis between individuals could originate either from genetic polymorphisms or from stable epigenetic regulation of gene expression in different individuals. Studies of epigenetic modification of genes involved in protein glycosylation are still scarce, but their results indicate that this process might be very important for the regulation of protein glycosylation.

Keywords: epigenetics; glycome; glycosyltransferases; protein glycosylation.

Introduction

Protein N-glycosylation is an ancient metabolic pathway developed before the diversification of Archaea, Bacteria and Eukarya, which still exists in all the three domains of life (1, 2). Over a half of all known eukaryotic proteins are N-glycosylated by covalent addition of branched oligosaccharides (glycans) to the asparagine residues within a sequence Asn-X-Ser/Thr (3). Many proteins are also O-glycosylated by the addition of glycans to serine or threonine. In unicellular organisms glycans generally function only as structural components of the cell membrane, whereas in multicellular organisms they acquired various complex functions needed to integrate numerous cells into a single functional unit (4, 5). At least 2000 different glycan determinants have been found to exist in mammalian glycoproteins (6). For example, over 30 different glycans (Figure 1) can be attached to the conserved Asn_{297} of the IgG heavy chain (7, 8). Because there are two heavy chains in each IgG molecule, over 900 different IgG isomers can be generated from this single glycosylation site. Between two and five glycans are attached to an average glycoprotein, resulting in an exceedingly complex glycoproteome (defined as the complete set of all glycoproteins in an organism), estimated to be at least several orders of magnitude more complex than the proteome itself (9).

N-glycosylation is essential for multicellular life and its complete absence is embryonically lethal (10). Variations in glycosylation are of great physiological significance because it has been demonstrated that changes in glycans significantly change the structure and function of polypeptide parts of glycoproteins (11). Proper glycosylation of membrane receptors is particularly important as it modulates adaptive properties of the cell membrane and affects communication between the cell and its environment (12). Dysregulation of glycosylation is associated with a wide range of diseases, including cancer, diabetes, cardiovascular, congenital, immunological and infectious disorders (13–15). Nevertheless, knowledge about the structure and function of glycans is still significantly lagging behind the knowledge of other macromolecules (16) due to experimental difficulties in analyzing complex glycan structures. However, recent technological advances have allowed reliable, high-throughput quantification of N-glycans (17–19), which now permits large-scale studies aimed at understanding the complexity of protein glycosylation.

The great functional diversity of glycans is a reflection of the great diversity in glycan structures and the diversity of proteins to which glycans are attached: they are important for proper folding of proteins; they can modify and/or regulate the function of protein backbones (20), contributing to the adaptive nature of the cell membrane (12); their strategic placement provides protease protection without interfering with the function of the protein; they serve as recognition motifs for specific carbohydrate binding proteins (lectins) and mediate cell–cell interactions (21); they enable proteins and lipids to 'jump' from one cell to another (22); and they can also have many other functions, some of which are still poorly understood (5). Carbohydrate recognition is an inte-

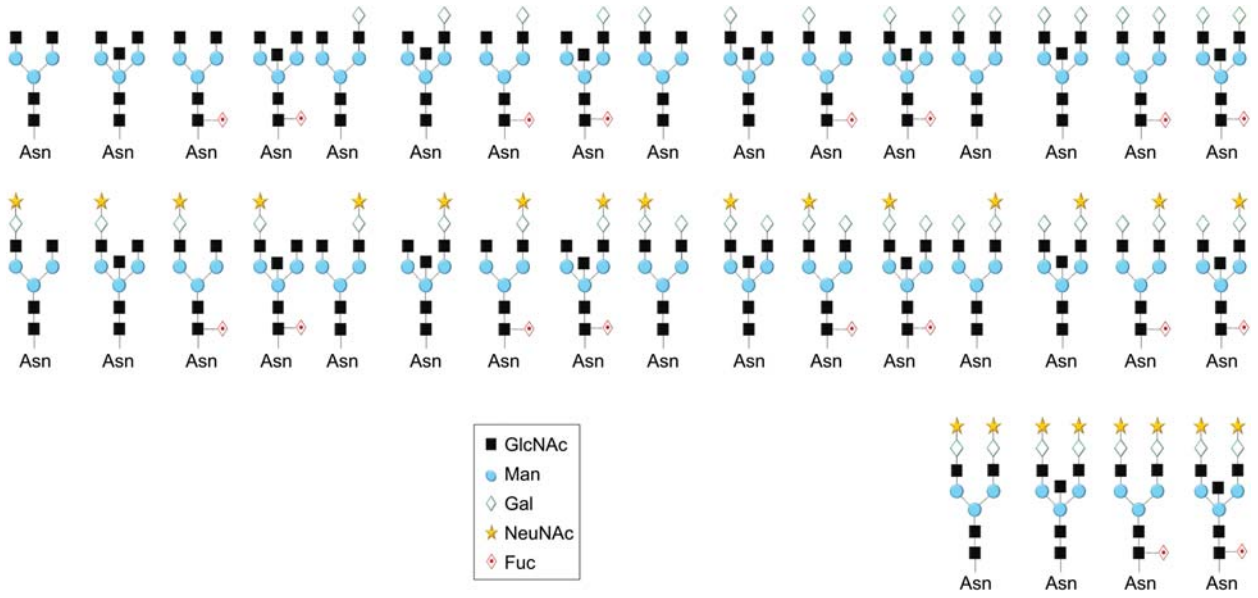

Figure 1 Structures of glycans on IgG heavy chains.
Structures of the main neutral, monosialylated and disialylated glycans which can be attached to conserved Asn_{297} of the IgG heavy chain are shown. These glycans have a common core composed of two core N-acetylglucosamines (squares), three core mannoses (circles) and two antennary N-acetylgalactosamines, but they differ in the number of galactoses (diamonds) and sialic acids (stars) on the antenna, and the presence or absence of N-acetylglucosamine attached to the first core mannose and the presence or absence of fucose (diamond with a dot) attached to the first core N-acetylglucosamine. Glycan structures significantly affect binding of IgG to its receptors and modulate functions of IgG in the immune response.

gral part of normal biological development (23) and the immune defense against pathogens is mediated through the identification of exogenous carbohydrates (14). Glycocalyx, a dense layer of glycoconjugates on the surface of all cells of higher eukaryotes, enables their communication with the outside world. Many bacterial and viral pathogens initially adhere to host tissues by binding specifically to carbohydrates on the host cell surfaces (24). Therefore, the versatility of glycan structures in higher eukaryotes helps them to compete with the fast evolution of pathogens and contributes to the evolutionary success of higher organisms despite their much longer generation times.

Glycans are very complex and highly versatile biomolecules

A typical glycan is a complex molecule containing between 10 and 15 monosaccharides. Contrary to proteins and DNA which are linear molecules, glycans are nonlinear branched structures and the particular glycan structure is characterized not only by the sequence of monomeric units but also by the exact position of the glycosidic bond, its anomeric configuration (α or β), the number of branches and the position of branching. In addition, although genes unequivocally determine the structure of each polypeptide (by encoding the sequence of amino acid residues), there is no genetic template for the glycan part. Instead, a complex dynamic network of hundreds of genes code for enzymes involved in the synthesis of glycans (25); over 600 'glyco-genes' (genes coding for various glycosyltransferases, glycosidases, enzymes for sugar nucleotide biosynthesis, transporters, etc.) are currently known (26, 27). The expression and activity of the 'glyco-genes' are affected by various transcription factors, Golgi organizers, proton pumps, etc., what additionally increases the number of genes and protein products that are directly or indirectly involved in the synthesis of glycans. Taking this into consideration, protein glycosylation can represent by far the most complex and the most expensive metabolic pathway in eukaryotes.

Glycosylation is the only post-translational modification that can produce significant structural changes to proteins. Contrary to the core N-glycan structure, which existence is essential for multicellular life (10), mutations in genes involved in modifications of glycan antennas are common and apparently cause a large part of individual phenotypic variations that exist in humans and other higher organisms. The most prominent example is the ABO system of blood group determination, which arises from the existence of three allelic variants of a single glycosyltransferase gene. However, the majority of human variability originates from single nucleotide polymorphisms (SNPs) that individually do not have visible phenotypes, but if present in specific combinations within the same individual can have significant phenotypic effects (28–30). Owing to hundreds of genes that interact to generate glycans, glycosylation is particularly prone to this type of variability. Some combinations of individual SNPs can be manifested as specific glyco-phenotypes, which might represent potential evolutionary advantages or disadvantages. The most prominent examples are various forms of congenital disorders of glycosylation (31), which

are usually caused by a combination of several individual mutations (32). Most of these mutations are actually leaky mutations that result in varying residual enzyme activity. However, when combined in the same individual they result in a complex phenotype associated with significant mortality and extensive motoric, immunological, digestive and neurological symptoms (31, 33).

Recently, we performed the first large-scale analysis of human plasma glycome (defined as the complete set of glycan structures in an organism), which revealed unusually high variability in level of individual glycans in human plasma (34). The median difference between minimal and maximal levels of individual plasma glycans was found to be over six-fold, much more than that recorded for any other classes of macromolecules. This variability appears to be genetically predetermined, because environmental factors were found to explain <5% of variance for the majority of glycans analyzed (35). Further support in favor of genetic predetermination of the individual glycome composition comes from a demonstration that the individual plasma glycome changes very little even after a prolonged period of time (36). Several clearly identifiable glyco-phenotypes which significantly differed from the normal glyco-phenotype in levels of one or more glycans were observed to exist in both European and Chinese populations. Some of these phenotypes were associated with specific pathological conditions, whereas others apparently did not have any identifiable adverse consequences for health (37). However, these analyses were performed in steady-state (normal) conditions, and it is unknown whether a specific aberrant glyco-phenotype would represent a disadvantage or advantage when an organism is challenged. For example, a decrease in the level of bi-galactosylated biantennary glycans with age was found to be associated with the increase in their sialylation (35). This indicates that cellular machinery for protein sialylation is operating at its near-maximal capacity and that any additional restriction (e.g., caused by a benign polymorphisms in the pathway) could limit the ability of the organism to respond to environmental challenges, what might be of particular importance when the organism is challenged with infection or system inflammation and needs to produce a large amount of (sialylated) acute phase proteins or antibodies.

Epigenetic mechanisms regulate gene expression

Epigenetics can be defined as mitotically heritable changes in gene expression that are not encoded in a DNA sequence but are affected by modifications of DNA and histone proteins, i.e., the chromatin template. The long histone tails of the nucleosomes, the primary chromatin subunits, rich in highly conserved Lys, Arg and Ser residues provide sites for post-translational modifications. Combinatorial and consecutive histone modifications, the 'histone code', are recognized by different enzymes to establish specific chromatin structures – open *versus* closed – and/or transcriptional states – expression *versus* silencing (38) of a particular chromosome/nuclear region. Synergistic and/or antagonistic action of specific histone modifying enzymes, such as histone acetyltransferases and deacetylases or histone methyltransferases, are responsible for correct structural and regulatory states of the particular nuclear region. The functional importance of these enzymes is highlighted by their fundamental regulatory roles in developmental processes and by the fact that their deregulation has been linked to the development and progression of many human diseases, most notably cancer (39–41). Indeed, the coordinated action of histone modifying enzymes with some other important enzymes (such as DNA methyltransferase I) as well as with chromatin-assembly factors and chromatin remodelers enable reliable inheritance of the epigenetic information (42). Hence, any cell lineage will show its own epigenetic memory.

'Histone code' alone does not provide the complete epigenetic information to a cell but is tightly connected with the most common epigenetic modification, the CpG methylation. There is a complex crosstalk between DNA and histone modifications forming a complex network of epigenetic information. CpG methylation, histone deacetylation and methylation of histone H3 at lysine 9 are associated with repressed chromatin and it has been shown that these three events are interrelated, self-reinforcing and self-perpetuating (43). Relationships between DNA methylation and histone modifications have implications for normal development, cell reprogramming and tumorigenesis. Both modifications are extensively altered in human cancers (44). Promoters of most human genes ($\sim$60%) are located in CpG rich sequences 0.3–2 kb in length, the so-called CpG islands, which are normally unmethylated. In cancer, hypermethylation within promoter region causes silencing of tumor suppressor genes and the loss of function (45). It has recently been shown that aberrant histone modifications cooperate with hypermethylation in this process (46–48); nevertheless, the whole picture of how these two epigenetic mechanisms are interrelated is still poorly understood (49). Two different groups of drugs, DNA methyltransferase inhibitors (such as 5-azacytidine, 5-aza-C and 5-aza-2′-deoxycytidine, 5-aza-2′-dC) and histone deacetylase inhibitors (such as butyrate), have been widely used in cancer research, and it has been shown that their synergistic effects can restore gene function silenced due to aberrant epigenetic changes in cancers (50).

Cancer-associated aberrant epigenetic modifications are not restricted to tumor suppressor genes and many other genes are affected. 'Glyco-genes' are one of the groups of genes which epigenetic status was found to be changed in tumors; aberrant methylation and histone deacetylation of several 'glyco-genes' have been detected in different tumors such as colorectal, gastric, pancreatic, breast, lung and blood cancers (51–56).

Epigenetic regulation affects glycan biosynthesis

Glycans were found to differ significantly between individuals (34), but within an individual glycome composition is

rather stable and changes very little with time (36). Stable individual differences in the glycan biosynthesis could be explained either by genetic polymorphisms or by stable epigenetic differences in transcriptional status of 'glyco-genes' (Figure 2). Altered glycan structures are often produced as a consequence of changes in gene expression during different physiological or pathophysiological processes. This is particularly exemplified in different cancers which are associated with various glycosylation changes (57). Studies on epigenetic modifications involved in protein glycosylation are still limited but their results indicate that this process might be very important for the regulation of protein glycosylation. These observations, organized by the affected gene families, are summarized below.

N-acetylglucosaminyltransferases

GnT-IVa (N-acetylglucosaminyltransferase IVa, *MGAT4A*) encodes a key glycosyltransferase that regulates the formation of highly branched glycan structures in the Golgi apparatus. GnT-IV is the only GnT for N-glycosylation that forms a gene family in higher vertebrates (25). It catalyzes the transfer of N-acetylglucosamine (GlcNAc) from UDP-GlcNac to the Man-α-1,3-Man-β-1,4-GlcNAc arm of complex glycans. It was reported to have a role in the regulation of the availability of serum glycoproteins, in oncogenesis and in differentiation.

The induction of GnT-IVa (isoenzyme N-acetylglucosaminyltransferase IVa) was observed after the treatment of human pancreatic cancer cells with both butyrate and 5-aza-C, suggesting that the downregulation of *MGAT4A* in cancer is due to an epigenetic abnormality. Because many cancer suppressor genes are regulated by epigenetic mechanisms, the downregulation of *MGAT4A* could be related to tumor suppressor functions (52).

N-acetylgalactosaminyltransferases

***B4GALNT2* (β-1,4-N-acetyl-galactosaminyltransferase 2)** catalyzes the last step in the biosynthesis of the human Sd^a antigen through the addition of an N-acetylgalactosamine residue via a β-1,4 linkage to a subterminal galactose residue substituted with an α-2,3-linked sialic acid. Sd^a antigen is reported as a human blood group antigen and is found in more than 90% of human red blood cells and in several tissue types: stomach, colon, kidney and oocytes, as well as in various body fluids such as saliva, milk, serum and urine. It has recently been shown that cytosine methylation plays a crucial

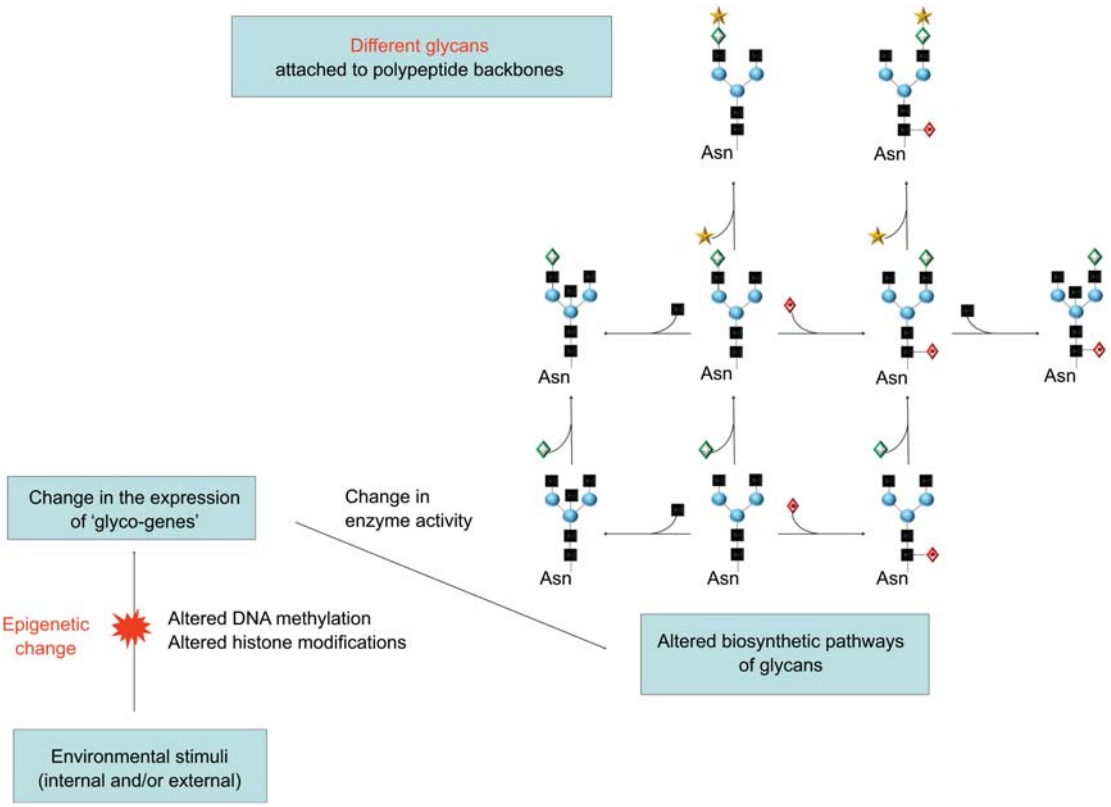

Figure 2 Epigenetic regulation of protein glycosylation.
Various signals from the external and/or internal environment of an individual can be a trigger for epigenetic changes resulting in a change of the expression of 'glyco-genes'. Depending on the activity of numerous glycosyltransferases (and other proteins involved in protein glycosylation), biosynthesis of glycans can proceed through several alternative biosynthetic pathways and result in the synthesis of different glycan structures.

role in regulation of the Sd^a β4GalNAcT-II gene and Sd^a antigen expression (58). When colon cancer cells, possessing a hypermethylated Sd^a gene promoter, were treated with 5-aza-2′-dC the expression of the Sd^a gene was induced as a result of demethylation of the promoter region. Therefore, downregulation of Sd^a gene expression in oncogenic processes facilitates the formation of the sLe^a and sLe^x antigens and promotes metastasis in gastrointestinal cancers (58). Additionally, when human colon cancer cells were treated with butyrate, a histone acetylase inhibitor, neither alterations in the Sd^a antigen nor *B4GALNT2* gene expression was observed (53). These results suggest that it is the DNA methylation rather than histone deacetylation that contributes to the downregulation of *B4GALNT2*.

The first discovered human blood group system, ABO, derives from the existence of three glycosyltransferase alleles, A, B and O, at the same locus. Allele A is an N-acetylgalactosaminyltransferase (*A3GALNT* gene), allele B is a galactosyltransferase and allele O is a nonfunctional truncated variant (59). Promoter hypermethylation of the *A3GALNT* gene was reported to be associated with the loss of blood group A antigen expression in bladder cancer, oral squamous cell carcinoma and gastric cancer cell lines (60–62). More recently, Kawamura and colleagues compared the expression levels of 'glyco-genes', including glycosyltransferases and glycosidases in normal gastrointestinal mucosa and in gastric and colorectal cancer cells (53). Inhibition of DNA methyltransferase was found to significantly increase expression of several N-acetylgalactosaminyltransferase genes including *A3GALNT*, *B4GALNT1* and *B4GALNT2* (53). These results suggested that aberrant methylation can lead to a cancer-associated reduction of the level of A antigen in colon cancers. Moreover, the methylation status of the *A3GALNT* gene was found to correlate well with the expression of the blood group A and B determinants in gastric cancer cell lines (53). These results suggest that DNA methylation contributes to a cancer-related silencing of the *A3GALNT* gene in gastrointestinal cancer cells.

Fucosyltransferases

FUT3 **(fucosyltransferase 3)** is a member of the fucosyltransferase family which catalyzes the addition of fucose to precursor polysaccharides in the final step of Lewis antigen biosynthesis. Fucosylated glycosphingolipids which comprise the Lewis histo-blood group system take part in embryogenesis, tissue differentiation, tumor metastasis, inflammation and bacterial adhesion. This enzyme is essential for the expression of Le^a and Le^b antigens on erythrocytes but is not expressed in hematopoietic cells. By contrast, it is widely expressed in the epithelial cells of the digestive tract, including salivary gland, mammary gland, esophagus, stomach, pancreas, small intestine and rectum (25).

Aberrant expression of Lewis antigens has been demonstrated in gastric carcinoma and can be partly due to an overexpression of the Lewis (*FUT3*) enzyme. Indeed, human α-1, 3/4 fucosyltransferase III (*FUT3*) participates in synthesis of all Lewis antigens (Le^a, SLe^a, Le^b, Le^x, SLe^x, Le^y) among which those that are altered in carcinomas also have a crucial role in the extravasation of cancer cells. Once again, *FUT3* gene expression was shown to be regulated by promoter methylation (54).

The *FUT7* (α-1,3-fucosyltransferase, FucT-VII) gene encodes a Golgi stack membrane protein that is involved in the formation of sialyl-Lewis X antigens and can direct the synthesis of the E-selectin binding sialyl X moiety. E- and P-selectin ligands are fucosylated oligosaccharides that enable trafficking of leukocytes and T cells into inflammed areas. The generation of selectin ligands depends on the coordinated action of a set of glycosyltransferases, particularly expression of fucosyltransferase (FucT) VII and core-2 glucosaminyltransferase (C2 GlcNAcT). The tissue distribution of *FUT7* is limited to leukocytes and high-endothelial venules (25).

Serpa and colleagues reported that cell cycle arrest prevented transcriptional activation of glycosyltransferases involved in the generation of selectin ligands (54). By performing artificial DNA demethylation with 5-aza-2′-dC, they strongly increased the frequency of selectin ligand expression in the cells. This data suggests that DNA methylation keeps transferase genes inaccessible in naive T cells. Hence, the induction of selectin ligands could be epigenetically modulated in lymphocytes during differentiation (63).

Sialyltransferases

ST6GalNAc6 is identified biochemically as a sialyltransferase responsible for the synthesis of 2,3/2,6 disialyl Le^a determinant (55). This gene belongs to the family of sialyltransferase genes, which products (sialyltransferases) modify proteins and ceramides on the cell surface to alter cell-cell or cell-extracellular matrix interactions. Sialyl Le^a determinant is also known to serve as a ligand for E-selectin and to mediate cancer metastasis. Concordantly, the expression of sialyl Le^a is known to be increased in cancers of the digestive organs. In contrast, disialyl Le^a, which has an extra sialic acid attached to the C6-position of penultimate GlcNAc in sialyl Le^a, is expressed preferentially on nonmalignant colonic epithelial cells and its expression decreases significantly with malignant transformation. The transcription of a gene encoding the α-2,6-sialyltransferase (responsible for disialyl Le^a synthesis in colon cancer cells) was markedly downregulated in cancer cells compared with nonmalignant epithelial cells, which is in line with the decreased expression of disialyl Le^a and increased expression of sialyl Le^a in cancer (55). Treatment of human colon cancer cells with butyrate or 5-aza-C strongly induced disialyl Le^a expression. This result implies that the observed downregulation of ST6GalNAc6 gene expression in cancer cells could be a consequence of epigenetic changes such as histone deacetylation and DNA methylation (55).

ST3GAL6 **(ST3 β-galactoside α-2,3-sialyltransferase 6)** catalyzes the transfer of sialic acid to terminal positions of glycoprotein and glycolipid carbohydrate groups. These ter-

minal sialic acid residues are widely distributed in many cell types. Kawamura and colleagues found this gene to be hypermethylated and downregulated in gastric cancer cell lines. These results suggest that epigenetic changes can occur in a group of 'glyco-genes' including *ST3GAL6* in gastric cancer tissues, which can eventually induce aberrant glycosylation and expression of cancer-associated carbohydrate antigens by silencing the enzyme activity responsible for antigen expression (53).

Sulfotransferases

Sulfotransferase 1A1 (*SULT1A1*) is a cytosolic enzyme with diverse tissue distribution. It catalyzes the sulfate conjugation of hormones, neurotransmitters, drugs and xenobiotic compounds. Among others, this reaction can result in inactivation of estrogens required for the maintenance of growth by most human breast cancers. Kwon and colleagues have shown a higher methylation density of the *SULT1A1* gene in cancer tissue and a comparatively lower methylation density in normal and benign breast tissues. After evaluating the mRNA expression of the *SULT1A1* gene, the data indicated that DNA methylation in the *SULT1A1* gene had a significant impact on the transcriptional silencing of the gene (56).

Heparan sulfate biosynthetic enzymes are key components in generating a myriad of distinct heparan sulfate fine structures that carry out multiple biological activities. The *3-OST-2* (heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2) gene encodes an enzyme involved in the final modification step of heparan sulfate proteoglycans (HSPGs) and it directly influences signal transduction. HSPGs play a role in cell growth: adhesion and migration by interactions with grow factors, morphogens, cytokines and extracellular matrices. A CpG island in the 5′ region of the *3-OST-2* gene was found to be hypermethylated, whereas its exon 2 was hypomethylated (51). After treatment with 5-aza-2′-dC the methylation was removed from the 5′ region of the *3-OST-2* gene, and consequently a loss of *3-OST-2* expression was observed. Silencing of the *3-OST-2* gene suggests that altered modification of HSPGs is involved in cancer development and occasionally even in progression.

Expert opinion and outlook

After completion of the human genome project, the focus of biomedical science shifted to proteomics, metabolomics and other 'omics'. Unfortunately, the majority of studies are still being performed on recombinant proteins, which either lack glycans altogether or are associated with some nonphysiological glycans. Because over 60% of all proteins are actually glycoproteins (3), this is a significant drawback. The complex shape, functionality and dynamic properties of glycans allow these molecules to function in intermolecular interactions as encoders of biological information. N-glycosylation is conserved between all three domains of life, and even complex eukaryotes first synthesize common N-glycan precursors with nine mannose residues, which are then being trimmed down to core $Man_3GlcNAc_2$ glycan by removal of six mannose residues. This process of adding and removing six mannose residues is metabolically very costly but apparently is a necessary process.

In spite of differences in genealogy and biosynthetic mechanisms of glycan and polypeptide parts of a glycoprotein, once synthesized, glycoprotein functions as a single unit and other interacting molecules cannot differentiate, whether they are binding to a protein or a carbohydrate part of the molecule. The most prominent peculiarity of glycans, which makes them very different from polypeptides, is that our genome does not contain templates for individual glycan structures. Instead, structures of glycans are encoded in a dynamic network of hundreds of genes and their products. Although production of a novel polypeptide structure requires changes in a DNA sequence, which threatens valuable genetic heritage, novel glycan structures can be produced by simple modification in gene expression, activity or localization of the corresponding enzymes. Moreover, because these changes are amenable to environmental influences (64), they are less predetermined than classical Mendelian mutations.

The variability of glycan structures both between and within a species by far exceeds the variability of proteins and other macromolecules. Although small differences between individual genes and proteins can hardly account for the versatility of life forms, when combined into a complex pathway such as protein glycosylation these small differences can generate a large difference in structures, which then perform different physiological functions. The immediate influence of environmental factors is also very important in this context. Although changes in protein structure can arise only as a consequence of mutations that irreversibly alter genetic information and can be validated only in the second generation, changes in glycan structures are reversible and can be repeatedly evaluated within the same organism, as manifested by crosstalk between intestinal glycans and commensal intestinal bacteria (65).

Epigenetic regulation is a universal tool that higher organisms use to adapt to environmental changes. Although environmental factors (such as diet) influence enzymatic processes only while they are directly present, their prolonged effects can be achieved through epigenetic modifications which are inherited through numerous cycles of cell division. By modulating expression of relevant genes, epigenetic information can influence glycan structures and make the organism more fit for a given environment (food sources, specific commensal or pathogenic microorganisms, etc.). Transgenerational inheritance of epigenetic information is still an insufficiently understood process (66, 67). Owing to the incomplete erasure of epigenetic modifications during the genome-wide resetting, which normally occur during mammalian early development (68), some epialleles (epigenetic variants) can be transmitted to the offspring. It is tempting to speculate that if an acquired (through the lifespan of an individual) epigenetic regulatory status of genes responsible for glycosylation would be transmitted to a germ cell, this could be an essential evolutionary mechanism, which would

enable adaptation of complex organisms to environmental changes, while preserving the precious genetic heritage.

Highlights

- The majority of eukaryotic proteins are glycosylated and the addition of glycans has important structural and regulatory roles.
- N-glycosylation is essential for life and its absence is embryonically lethal.
- Glycoproteome is several orders of magnitude more complex than the proteome.
- Glycan structures are not defined by single genes but by a complex dynamic network of hundreds of genes.
- Glycome profiles are generally stable in an individual, but highly variable between individuals and species.
- Changes in epigenetic regulation of glycosylation generate novel biological structures without changes in the genome of an organism. These structures confer novel functions and contribute to the adaptation to environmental changes.
- Inheritance of epigenetic regulation of protein glycosylation is an intriguing mechanism that could contribute to the evolutional adaptivity of complex organisms.

Conflict of interest statement

None declared.

Acknowledgements

Research in the authors' laboratories is supported by grants #309-0061194-2023 (to G.L.), and #119-1191196-1224 (to V.Z.) from the Croatian Ministry of Science, Education and Sport, and by FP6-EuroPharm, FP7-EuroGlycoArrays and FP7-GlycoBioMarkers grants from the European Community.

References

1. Weerapana E, Imperiali B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. Glycobiology 2006; 16: 91R–101R.
2. Calo D, Kaminski L, Eichler J. Protein glycosylation in Archaea: sweet and extreme. Glycobiology 2010; 20: 1065–76.
3. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1999; 1473: 4–8.
4. Drickamer K, Taylor ME. Evolving views of protein glycosylation. Trends Biochem Sci 1998; 23: 321–4.
5. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 1993; 3: 97–130.
6. Cummings RD. The repertoire of glycan determinants in the human glycome. Mol Biosyst 2009; 5: 1087–104.
7. Sutton BJ, Phillips DC. The three-dimensional structure of the carbohydrate within the Fc fragment of immunoglobulin G. Biochem Soc Trans 1983; 11: 130–2.
8. Harada H, Kamei M, Tokumoto Y, Yui S, Koyama F, Kochibe N, Endo T, Kobata A. Systematic fractionation of oligosaccharides of human immunoglobulin G by serial affinity chromatography on immobilized lectin columns. Anal Biochem 1987; 164: 374–81.
9. Lee RT, Lauc G, Lee YC. Glycoproteomics: protein modifications for versatile functions. EMBO Rep 2005; 6: 1018–22.
10. Marek KW, Vijay IK, Marth JD. A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. Glycobiology 1999; 9: 1263–71.
11. Skropeta D. The effect of individual N-glycans on enzyme activity. Bioorg Med Chem 2009; 17: 2645–53.
12. Dennis JW, Lau KS, Demetriou M, Nabi IR. Adaptive regulation at the cell surface by N-glycosylation. Traffic 2009; 10: 1569–78.
13. Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. Nat Rev Immunol 2007; 7: 255–66.
14. Marth JD, Grewal PK. Mammalian glycosylation in immunity. Nat Rev Immunol 2008; 8: 874–87.
15. Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. Cell 2006; 126: 855–67.
16. Lauc G. Sweet secret of the multicellular life. Biochim Biophys Acta 2006; 1760: 525–6.
17. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. Anal Biochem 2008; 376: 1–12.
18. Huhn C, Selman MH, Ruhaak LR, Deelder AM, Wuhrer M. IgG glycosylation analysis. Proteomics 2009; 9: 882–913.
19. Selman MH, McDonnell LA, Palmblad M, Ruhaak LR, Deelder AM, Wuhrer M. Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem 2010; 82: 1073–81.
20. Moloney DJ, Shair LH, Lu FM, Xia J, Locke R, Matta KL, Haltiwanger RS. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. J Biol Chem 2000; 275: 9604–11.
21. Lee YC, Lee RT. Carbohydrate–protein interactions: basis of glycobiology. Acc Chem Res 1995; 28: 321–7.
22. Lauc G, Heffer-Lauc M. Shedding and uptake of gangliosides and glycosylphosphatidylinositol-anchored proteins. Biochim Biophys Acta 2006; 1760: 584–602.
23. Haltiwanger RS, Lowe JB. Role of glycosylation in development. Annu Rev Biochem 2004; 73: 491–537.
24. Sharon N. Lectins: carbohydrate-specific reagents and biological recognition molecules. J Biol Chem 2007; 282: 2753–64.
25. Taniguchi N, Honke K, Fukuda M, editors. Handbook of glycosyltransferases and related genes. Tokyo: Springer Verlag, 2002.
26. Abbott KL, Nairn AV, Hall EM, Horton MB, McDonald JF, Moremen KW, Dinulescu DM, Pierce M. Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. Proteomics 2008; 8: 3210–20.
27. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J Biol Chem 2008; 283: 17298–313.
28. Ovsyannikova IG, Haralambieva IH, Dhiman N, O'Byrne MM, Pankratz VS, Jacobson RM, Poland GA. Polymorphisms in the vitamin A receptor and innate immunity genes influence the antibody response to rubella vaccination. J Infect Dis 2010; 201: 207–13.

29. Dhiman N, Haralambieva IH, Kennedy RB, Vierkant RA, O'Byrne MM, Ovsyannikova IG, Jacobson RM, Poland GA. SNP/haplotype associations in cytokine and cytokine receptor genes and immunity to rubella vaccine. Immunogenetics 2010; 62: 197–210.
30. Brown RS. Autoimmune thyroid disease: unlocking a complex puzzle. Curr Opin Pediatr 2009; 21: 523–8.
31. Freeze HH. Genetic defects in the human glycome. Nat Rev Genet 2006; 7: 537–51.
32. Haeuptle MA, Hennet T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. Hum Mutat 2009; 30: 1628–41.
33. Freeze HH. Human disorders in N-glycosylation and animal models. Biochim Biophys Acta 2002; 1573: 388–93.
34. Knežević A, Polašek O, Gornik O, Rudan I, Campbell H, Hayward C, Wright A, Kolčić I, O'Donoghue N, Bones J, Rudd PM, Lauc G. Variability, heritability and environmental determinants of human plasma N-glycome. J Proteome Res 2009; 8: 694–701.
35. Knežević A, Gornik O, Polašek O, Pučić M, Novokmet M, Redžić I, Rudd PM, Wright AF, Campbell H, Rudan I, Lauc G. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. Glycobiology 2010; 20: 959–69.
36. Gornik O, Wagner J, Pučić M, Knežević A, Redžić I, Lauc G. Stability of N-glycan profiles in human plasma. Glycobiology 2009; 19: 1547–53.
37. Pučić M, Pinto S, Novokmet M, Knežević A, Gornik O, Polašek O, Vlahoviček K, Wei W, Rudd PM, Wright AF, Campbell H, Rudan I, Lauc G. Common aberrations from normal human N-glycan plasma profile. Glycobiology 2010; 20: 970–5.
38. Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000; 403: 41–5.
39. Timmermann S, Lehrmann H, Polesskaya A, Harel-Bellan A. Histone acetylation and disease. Cell Mol Life Sci 2001; 58: 728–36.
40. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, Waghorn K, Zoi K, Ross FM, Reiter A, Hochhaus A, Drexler HG, Duncombe A, Cervantes F, Oscier D, Boultwood J, Grand FH, Cross NC. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 2010; 42: 722–6.
41. Moss TJ, Wallrath LL. Connections between epigenetic gene silencing and human disease. Mutat Res 2007; 618: 163–74.
42. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol 2009; 10: 192–206.
43. Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature 2001; 414: 277–83.
44. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 2006; 6: 107–16.
45. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007; 128: 683–92.
46. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW. Epigenetic stem cell signature in cancer. Nat Genet 2007; 39: 157–8.
47. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 2007; 39: 237–42.
48. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Perez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet 2005; 37: 391–400.
49. Kondo Y. Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. Yonsei Med J 2009; 50: 455–63.
50. Cameron EE, Bachman KE, Myohanen S, Herman JG, Baylin SB. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat Genet 1999; 21: 103–7.
51. Miyamoto K, Asada K, Fukutomi T, Okochi E, Yagi Y, Hasegawa T, Asahara T, Sugimura T, Ushijima T. Methylation-associated silencing of heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2 (3-OST-2) in human breast, colon, lung and pancreatic cancers. Oncogene 2003; 22: 274–80.
52. Ide Y, Miyoshi E, Nakagawa T, Gu J, Tanemura M, Nishida T, Ito T, Yamamoto H, Kozutsumi Y, Taniguchi N. Aberrant expression of N-acetylglucosaminyltransferase-IVa and IVb (GnT-IVa and b) in pancreatic cancer. Biochem Biophys Res Commun 2006; 341: 478–82.
53. Kawamura YI, Toyota M, Kawashima R, Hagiwara T, Suzuki H, Imai K, Shinomura Y, Tokino T, Kannagi R, Dohi T. DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. Gastroenterology 2008; 135: 142–51.
54. Serpa J, Mesquita P, Mendes N, Oliveira C, Almeida R, Santos-Silva F, Reis CA, LePendu J, David L. Expression of Lea in gastric cancer cell lines depends on FUT3 expression regulated by promoter methylation. Cancer Lett 2006; 242: 191–7.
55. Miyazaki K, Ohmori K, Izawa M, Koike T, Kumamoto K, Furukawa K, Ando T, Kiso M, Yamaji T, Hashimoto Y, Suzuki A, Yoshida A, Takeuchi M, Kannagi R. Loss of disialyl Lewis(a), the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewis(a) expression on human colon cancers. Cancer Res 2004; 64: 4498–505.
56. Kwon MS, Kim SJ, Lee SY, Jeong JH, Lee ES, Kang HS. Epigenetic silencing of the sulfotransferase 1A1 gene by hypermethylation in breast tissue. Oncol Rep 2006; 15: 27–32.
57. Lau KS, Dennis JW. N-Glycans in cancer progression. Glycobiology 2008; 18: 750–60.
58. Wang HR, Hsieh CY, Twu YC, Yu LC. Expression of the human Sd(a) beta-1,4-N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. Glycobiology 2008; 18: 104–13.
59. Yoshida A. Identification of genotype of blood group A and B. Blood 1980; 55: 119–23.
60. Chihara Y, Sugano K, Kobayashi A, Kanai Y, Yamamoto H, Nakazono M, Fujimoto H, Kakizoe T, Fujimoto K, Hirohashi S, Hirao Y. Loss of blood group A antigen expression in bladder cancer caused by allelic loss and/or methylation of the ABO gene. Lab Invest 2005; 85: 895–907.
61. Gao S, Worm J, Guldberg P, Eiberg H, Krogdahl A, Liu CJ, Reibel J, Dabelsteen E. Genetic and epigenetic alterations of the blood group ABO gene in oral squamous cell carcinoma. Int J Cancer 2004; 109: 230–7.
62. Kominato Y, Hata Y, Takizawa H, Tsuchiya T, Tsukada J, Yamamoto F. Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region. J Biol Chem 1999; 274: 37240–50.

63. Syrbe U, Jennrich S, Schottelius A, Richter A, Radbruch A, Hamann A. Differential regulation of P-selectin ligand expression in naive versus memory CD4+ T cells: evidence for epigenetic regulation of involved glycosyltransferase genes. Blood 2004; 104: 3243–8.
64. Lauc G, Zoldos V. Protein glycosylation – an evolutionary crossroad between genes and environment. Mol Biosyst 2010; DOI: 10.1039/C0MB00067A.
65. Hooper LV, Gordon JI. Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology 2001; 11: 1R–10R.
66. Richards EJ. Inherited epigenetic variation – revisiting soft inheritance. Nat Rev Genet 2006; 7: 395–401.
67. Rakyan VK, Beck S. Epigenetic variation and inheritance in mammals. Curr Opin Genet Dev 2006; 16: 573–7.
68. Morgan HD, Santos F, Green K, Dean W, Reik W. Epigenetic reprogramming in mammals. Hum Mol Genet 2005; 14 (Spec No 1): R47–R58.