Epigenetic regulation of glycosylation is the quantum mechanics of biology

Gordan Lauc a,b, Aleksandar Vojta c, Vlatka Zoldoš c,*

a Department of Biochemistry and Molecular Biology, Faculty of Pharmacy and Biochemistry, University of Zagreb, A. Kovačića 1, 10000 Zagreb, Croatia
b Genos Glycobiology Laboratory, Handrova 2/11, 10000 Zagreb, Croatia
c Department of Molecular Biology, Faculty of Science, University of Zagreb, Horvatovac 102a, 10000 Zagreb, Croatia

A R T I C L E   I N F O
Article history:
Received 24 April 2013
Received in revised form 22 July 2013
Accepted 23 August 2013
Available online 31 August 2013

Keywords:
Glycosylation
Epigenetics
Adaptation
Evolution

A B S T R A C T
Background: Most proteins are glycosylated, with glycans being integral structural and functional components of a glycoprotein. In contrast to polypeptides, which are fully encoded by the corresponding gene, glycans result from a dynamic interaction between the environment and a network of hundreds of genes.

Scope of review: Recent developments in glycomics, genomics and epigenomics are discussed in the context of an evolutionary advantage for higher eukaryotes over microorganisms, conferred by the complexity and adaptability which glycosylation adds to their proteome.

Major conclusions: Inter-individual variation of glycome composition in human population is large; glycome composition is affected by both genes and environment; epigenetic regulation of “glyco-genes” has been demonstrated; and several mechanisms for transgenerational inheritance of epigenetic marks have been documented.

General significance: Epigenetic recording of acquired characteristics and their transgenerational inheritance could be important mechanisms used by higher organisms to compete or collaborate with microorganisms.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

With the ability to sequence genomes in a matter of hours, along with the accompanying technical advances in other “omics,” biology is ripe for a scientific revolution analogous to the one which transformed the field of physics in the early 20th century. Newton’s laws of motion are still as useful today as they were in the 17th century when first formulated. However, certain properties of matter could not be explained until a complete paradigm shift took place with the introduction of quantum mechanics. Biology today faces a similar challenge: with the theory of Darwinian evolution by natural selection still undisputed as a cornerstone of modern biology, certain aspects of adaptation to selection pressure cannot be adequately explained by changes in single protein structures alone. Rather, the complexity, which lies inside as well as outside of the genome itself and in the intricate network of interactions belonging to other “omics,” begins to emerge as an important evolutionary force. Two frequently overlooked “omics” – glycomics and epigenomics – are the missing pieces of the puzzle and a key to better understanding of biology, which might soon prove as important to that discipline as the introduction of quantum mechanics had been for physics. Despite the scarcity of hard evidence, the big picture is already emerging from the recent studies, which makes this exciting new field, i.e. epigenetics of glycosylation, ripe to be reviewed in the context of evolution.

One major point which simple Darwinian model of evolution fails to explain adequately is the huge difference in the rate of reproduction between prokaryotic microorganisms and higher eukaryotes, plants and animals in particular. For example, the majority of animals will have at most – and only if we consider extreme examples – several thousands of surviving offspring in their lifetimes, while a single bacterium can generate billions of progeny bacteria in a single day. Clearly, if higher eukaryotes are to keep up with this evolutionary arms race without getting overrun by sheer numbers, they have to look for a source of diversity and rapid adaptation elsewhere, but not in their reproductive capacity. This matching of the evolutionary rate (and the speed of general adaptation) is achieved in higher eukaryotes by modifying their proteins not only by direct change in the amino acid sequence – which takes a full generation to be established – but by attaching other molecules, such as glycans, to their surface, changing their function in this way and enormously increasing diversity, thus compensating for their slower reproduction rate.

Another point, which Darwinian evolution does not explain adequately, is the shaping of development and functional integration of trillions of cells in a multicellular organism. The way cells are organized into higher-order structures (tissues, organs) is written in the genome, but in a way that is not nearly as explicit as how the structure of a protein is encoded. An intricate and both functionally and structurally complex system such as the human brain is fully defined by a set of
In contrast to polypeptides, which are fully de 
the big difference between a polypeptide and a glycoprotein is that 
large dynamic network of both genetic and environmental factors 
function of a protein.
covalently linked glycans. Both the polypeptide and the glycan part of a protein participate in its structure and function and it is incorrect to consider them separately while studying the 
importance of regulatory elements (vs. protein-coding genes) in defining the blueprint of an organism.
The genome, with its structural (protein-coding) and regulatory 
elements, defines an organism by serving as a template giving rise to 
a complex network of interacting biological molecules. Such modular 
biological networks capture some key properties of life — robustness 
et al) are epigenetics — which adds to genome 
another layer of information about when, where and how a coding 
sequence will be read, and glycosylation — with its capacity to significantly 
alter protein structure and function.
Posttranslational modifications alter and enrich protein structure 
and function. By far the most complex of these modifications is 
glycosylation. The vast majority of human proteins are glycosylated 
with most proteins targeted to the cell’s membrane system getting 
the core glycan attached during their synthesis in the endoplasmic 
reticulum (N-linked glycosylation), with further processing and 
O-linked glycosylation occurring in the Golgi apparatus. Glycan 
parts of proteins perform numerous important structural and func-
tional roles [6]. Actually, once the glycan part is added to the poly-
peptide backbone, it becomes completely irrelevant whether —OH, 
—NH₂ or —COO⁻ groups belong to the polypeptide or the glycan 
part. They all together form the integral molecular structure 
(Fig. 1) that performs specific physiological functions [7]. However, 
the big difference between a polypeptide and a glycoprotein is that 
there is no direct genetic template for glycan parts of glycoproteins. In contrast to polypeptides, which are fully defined by nucleotide 
sequence in the corresponding genes, glycans are defined by a 
large dynamic network of both genetic and environmental factors 
[8,9]. In addition to genetic polymorphisms in the participating 
genes, regulation of gene expression, posttranslational modifications, 
and the activity of the corresponding proteins work together to 
determine the structure of a glycan. Through this process, the 
environment participates in shaping the final structure of a glycoprotein.

Biosynthesis of glycans requires many monosaccharide building 
blocks and their availability significantly affects structure of glycans 
and composition of the glycome [10]. Altered pH in Golgi [11,12], 
oxygen concentration [13] and many other external factors also 
affect protein glycosylation. Subcellular localization of enzymes, 
activated monosaccharide donor substrates and glycan acceptor 
substrates can also affect the final outcome [14]. We are only beginning 
to understand the details of the intricate enzymatic network which 
controls the manner in which proteins are glycosylated [15]. Recently 
initiated genome wide association studies (GWAS) of the human 
glycome [16–18] started to identify new and unexpected genes which are involved in this process and further progress in this field 
is expected to map the complex network of genes which regulates 
protein glycosylation [19].

All these “glyco-genes” (glycosyltransferases, glycosidases and 
other genes involved in complex biosynthetic pathways of glycans) 
are regulated on the transcriptional level not only by general transcription 
factors, but also by chromatin-modifying activities including ATP-
dependent remodeling complexes as well as histone modifying 
complexes, which add/remove covalent groups (phosphate, acetyl and 
methyl groups, etc.) to/from histone tails. These chromatin-modifying 
activities act in concert with DNA methylation to create epigenetic 
information, which not only determines gene transcription status, but 
also changes this status in response to external and intrinsic signals, in 
order to achieve appropriate functional change in protein glycosylation 
(Fig. 2). The mediator role of epigenetic mechanisms between genes, 
environment [20] and the final glycoprotein structure and function, 
has a great potential for evolution of multicellular life [9]. For example, 
the repertoire of glycan structures that can be produced by epigenetic 
changes in glyco-gene expression can be very large [21]. The addition 
of glycans to polypeptide backbones increases the complexity of the 
proteome by several orders of magnitude. This increased structural 
capacity and its dynamic flexibility enables complex eukaryotes to 
perform numerous complex functions. For example, fine tuning of IgG 
function and the regulation of the cell surface half-life of membrane 
proteins seem to be, at least by a large part, regulated by alternative 
glycosylation [22,23]. The role of alternative glycosylation in the 
function of the important developmental regulator Notch has also 
been well documented [24]. Recent population studies of both total 
plasma glycine [25] and glycine of an individual protein [26] revealed 
great inter-individual differences in glycome composition, while 
individual glycome composition was remarkably stable [27,28]. Up to

Fig. 1. Examples of glycoprotein structure. Three examples of glycoprotein structure (A — Fc fragment of immunoglobulin G, B — toll like receptor 8 and C — promeprin) are provided in a form of molecular models. The polypeptide backbone is shown in gray, and glycans in color, but it should be noted that glycoprotein is an integral structure composed of a polypeptide with covalently linked glycans. Both the polypeptide and the glycan part of a protein participate in its structure and function and it is incorrect to consider them separately while studying the function of a protein.
50% of the observed variations were heritable [25] with limited effects of directly acting environmental factors on the majority of glycans [29].

A particular gene expression pattern is established by epigenetic marks and then memorized, meaning inherited through cell divisions [30]. However, the epigenome also provides the genome with certain plasticity, owing to the possibility of epigenetic marks to change rapidly in response to environment (the so-called epigenetic on/off switch) and to the reversible nature of this change. These short-term memory epigenetic effects [31] are mostly achieved by quick alterations in histone marks, rather than changes in DNA methylation. DNA methylation also changes during a lifetime, either stochastically or in response to environmental factors. However, this change is less rapid and represents long-term memory effects [32], since an addition of the methyl group to a cytosine is a more stable epigenetic mark than are histone modifications. In order to have impact on evolution, the newly established alterations in the gene expression pattern should be passed through gametes to the next generation, with alterations in DNA methylation being a plausible mechanism. The body of evidence demonstrating transgenerational epigenetic inheritance in both plants and animals is growing rapidly [33–40]; this phenomenon is developing into an exciting topic in the field of epigenetics. In their extensive review, Jablonka and Raz [41] gave an impressive table of over hundred examples of inherited epigenetic variations for organisms ranging from Caenorhabditis elegans to humans. Molecular mechanisms for transgenerational epigenetic inheritance have been extensively studied, but full molecular characterization of epigenetic transfer through gametes to the next generation/ generations is not yet available for any organism.

Most of the studies performed in animals have identified incomplete epigenetic resetting of DNA methylation as the most probable mechanism for transfer of epigenetic information through gametes [42–44]. In order to pass to the next generation, DNA methylation variations (i.e. epialleles) have to slip through the two waves of epigenetic reprogramming — during gametogenesis and early embryogenesis [42]. There are some valuable examples of DNA methylation-mediated trans-generational inheritance by incomplete erasure of methylation marks [34,43,45,46]. The recent prominent study of Skinner and coworkers [40] has shown how subtle environmentally induced changes in cytosine methylation can have a dramatic effect on the transcriptome of different tissues mediated by “epigenetic control regions,” even in the F3 generation. DNA methylation is mechanistically interrelated with other chromatin components such as histone modifications and/or action of small non-coding RNA molecules. However, data are scarce for histone-mediated transgenerational inheritance by incomplete replacement of histones by protamines [47], retention of the centromeric histone H3 variant CENP-A in mammalian sperm [48], or by direct modifications to sperm chromatin [39]. These, and other epigenetic inheritance systems, such as self-sustaining feedback loop, structural inheritance and small RNAs [41], have nevertheless not been as rigorously explored in animal models as they have been in plants [49–51].

An exciting (albeit the least explored) epigenetic inheritance system is the action of small non-coding RNAs of various origins. Evidence is accumulating that this mechanism can be responsible for epigenetic effects lasting through multiple generations. In rats and some other mammals, epigenetic effects mediated through non-coding RNAs are recorded within 3–4 generations [44,52,53] and in some insects even for 10–15 generations [41]. Recent outstanding study in C. elegans has shown that small interfering virus-derived viRNAs are involved in transgenerational epigenetic inheritance through 30 generations in the absence of the genetic template and even in the absence of

![Fig. 2. Schematic representation of adaptation to changes in environment through epigenetic regulation of glyco-genes. Epigenetic mechanisms are mediators between the environment and the glyco-genes, which is reflected in the final glycan structures on the membrane glycoproteins interacting with microbes. Glyco-gene expression is epigenetically regulated by DNA methylation and by chromatin modifications, which include covalent modifications of histone tails (phosphorylation, acetylation, methylation, etc.) and ATP-dependent remodeling complexes; small non-coding RNAs also play a regulatory role. Glyco-gene transcriptional status can change in response to external and intrinsic signals, in order to achieve appropriate functional change in protein glycosylation, which can become adaptive through long-term epigenetic effects and even passed to the next generation.](image-url)
Table 1 (continued)

<table>
<thead>
<tr>
<th>Gene/pathway</th>
<th>Conclusion</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sialyltransferase genes</td>
<td>Regulation of sialyltransferase genes by promoter methylation is involved in differential regulation of selectin ligand expression in naïve versus CD4+ T cells.</td>
<td>[63]</td>
</tr>
<tr>
<td>Sulfate transporter gene</td>
<td>Decrease of DTDST expression results in decreased expression of sialyl 6-sulfo Lewis(x) and increased expression of sialyl Lewis(x) in colon cancer.</td>
<td>[65]</td>
</tr>
<tr>
<td>Transcription factor</td>
<td>Hypermethylation of HNF1A is associated with the increase in the proportion of highly branched glycans in the plasma N-glycome suggesting that HNF1A regulates transcription of glyco-genes involved in glycogen branching.</td>
<td>[67]</td>
</tr>
</tbody>
</table>

The functional small RNA-generating machinery [35]. Mammalian spermatocytes and oocytes are filled with piwi-interacting RNAs (piRNAs) [54,55], responsible for silencing of retrotransposons and other repetitive elements in germ line cells. Therefore, these and some other similar, yet undiscovered, RNA molecules could be candidates for the transgenerational epigenetic inheritance through germ-line cells in humans. Glycans are synthesized through complex biochemical pathways in which many genes are involved. The final glycan structure is as much influenced by genetic polymorphisms as by environmental factors where epigenetic mechanisms play mediator role between environment and the glyco-gene expression. Indeed, many glyco-genes with a role in normal development [56,57] are epigenetically regulated (Table 1). These glyco-genes show different epigenetic regulation in normal cells and in cancer [58,59], a connection which is sometimes established through the influence of epigenetically controlled glyco-genes on other cellular processes such as apoptosis [60]. There are many examples of expression of cancer-specific glycans in many types of cancer such as colon cancer, where these are the products of epigenetic deregulation either by promoter methylation [61,62] or by histone modifications [63]. Other examples include bladder [64], ovary [65], gastric [62] and pancreatic [66] cancer. Also, epigenetic deregulation of other glycosylation-related genes, such as transcription factors, is shown to have an effect on glycome composition and the disease outcome [67]. Treatments of cells in culture with epigenetic inhibitors reveal that N-glycome profiles drastically change, which is an indication that many glyco-genes and glycosylation-related genes are regulated both by DNA methylation and histone modifications [65,68,69]. Finally, tissue-specific epigenetic control of glyco-genes has been recently found in brain [57], which implies a role of glycosylation in development. Hard evidence is thus accumulating to support the very important role of epigenetically controlled glycosylation in differentiation and adaptation. Also, epigenetic regulation of protein glycosylation might represent an important road from homeostasis to complex diseases such as diabetes [67], cardiovascular diseases, or cancer. Evolutionary significance of epigenetic variations and epimutations has been widely discussed by Jablonka and Raz [41]. By combining these mechanisms for the inheritance of “acquired” characteristics, with the power of glycosylation machinery to create novel structures, higher organisms could have generated a powerful mechanism for creation of large structural variability through environmentally mediated, transgenerationally inherited epigenetic changes. The evolutionary impact of this mechanism could be immense. Glycans are the main receptors for virtually all pathogenic and commensal microorganisms and higher organisms have complex mechanisms

Table 1
Examples of recent research in the field of epigenetics of glycosylation. The presented original research highlights the recent advances in elucidating the role of epigenetic deregulation of glyco-gene expression in cancer or differentiation.

<table>
<thead>
<tr>
<th>Gene/pathway</th>
<th>Conclusion</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>β3Gal-T5/ expression of Lewis antigens in colon cancer</td>
<td>β3Gal-T5 galactosyltransferase influences expression of oligosaccharide products, such as Lewis antigens. β3Gal-T5 native promoter is regulated by cross-talk between DNA methylation, histone methylation and acetylation, and epigenetic deregulation of this gene is involved in colon cancer. Loss of heterozygosity and hypermethylation of ABO gene associate with abnormal expression of the A antigen in transitional cell carcinoma of the bladder.</td>
<td>[61]</td>
</tr>
<tr>
<td>ABO gene/expression of A antigen in bladder cancer</td>
<td>[64]</td>
<td></td>
</tr>
<tr>
<td>Cell membrane glycoconjugates</td>
<td>[58]</td>
<td></td>
</tr>
<tr>
<td>HeLa cell surface N-glycome</td>
<td>The pattern of N-glycosylation of HeLa cell surface proteins is changed by epigenetic inhibitors suggesting that cell surface glycosylation is regulated by both DNA methylation and histone acetylation. The effects are reversible if HeLa cells are grown in a drug-free environment indicating that this regulation is stable.</td>
<td>[68,69]</td>
</tr>
<tr>
<td>GnT-Iva/pancreatic cancer</td>
<td>N-acetylgalactosaminyltransferase-Iva is down-regulated by DNA methylation and histone acetylation.</td>
<td>[66]</td>
</tr>
<tr>
<td>Glycosyltransferase genes including B4GALNT2 and ST3GAL5/expression of gastrointestinal cancer specific antigens</td>
<td>Promoter hypermethylation of B4GALNT2 and ST3GAL5 silences the expression of these genes which may be related to aberrant glycosylation and expression of cancer-associated carbohydrate antigens in gastric and colon cancers.</td>
<td>[62]</td>
</tr>
<tr>
<td>GnT-IX/brain-specific expression of O-mannosyl type of glycans</td>
<td>GnT-IX gene, exclusively expressed in the brain, encodes for N-acetylgalactosaminyltransferase IX responsible for branched type O-mannosyl glycans involved in neural cell adhesion and migration through catenin signaling. GnT-IX expression is regulated by brain-specific chromatin activation (by histone modifications) and DNA binding proteins CTCF and NeuroD1.</td>
<td>[57]</td>
</tr>
<tr>
<td>Gene for α2→6 sialyltransferase/expression of Lewis antigens in colon cancer</td>
<td>DNA methylation and/or histone acetylation induce strong expression of Lewis(a) antigens in colon cancer suggesting that gene for α2→6 siaxyltransferase can be epigenetically regulated.</td>
<td>[74]</td>
</tr>
<tr>
<td>Fucosylation-related genes/TRAIL-induced apoptosis in cancers</td>
<td>TRAIL ligand induces apoptosis in cancer but not in normal cells and the deficiency of fucosylation leads to resistance to TRAIL-induced apoptosis. Cellular fucosylation is regulated by DNA methylation of fucosylation-related genes in cancer cells.</td>
<td>[60]</td>
</tr>
<tr>
<td>GMDS, FX and MGAT5/altered N-glycosylation pattern in ovarian cancer epithelial cells (OVCAR3)</td>
<td>Hypomethylation of GMDS and FX causes reduction of core fucosylation, while hypomethylation of MGAT5 causes increase of branching and sialylation of OVCAR3 secreted glycoproteins.</td>
<td>[65]</td>
</tr>
<tr>
<td>FUT3/Lewis antigens in gastric cells</td>
<td>FUT3 overexpression due to promoter hypomethylation is responsible for overexpression of Lewis antigens in gastric cells.</td>
<td>[75]</td>
</tr>
</tbody>
</table>
to modulate these interactions [70,71]. Human populations exposed to pathogens develop resistance mechanisms, which are poorly understood, but the presence of these mechanisms is clearly evident from devasting effects of relatively benign diseases like smallpox, chicken pox, or measles which decimated native American populations after being transferred from Europe [72]. It is tempting to speculate that the European resistance to diseases endemic to the Old World resulted from gene expression patterns developed as an adaptation to specific pathogens, leading to adaptive glycosylation in the immune system, which was passed to the next generations by epigenetic mechanisms. This speculation is not as far-fetched as it might sound at first glance, given the well-known role of epigenetic mechanisms in the transmission of traits from parents to their offspring. For example, at the moment it is only possible to focus on the main effects of GRE modifier genes, such as those encoding a variety of N-linked glycosylation enzymes.
K. Moriwaki, M. Narisada, T. Imai, S. Shinzaki, E. Miyoshi, The effect of epigenetic...


