Journal of Glycomics & Lipidomics

Editorial   **Open Access**

Remarkable Indications-How the Glycomics is Coded in Cells and Tissues

Jozef Nahalka*

Institute of Chemistry, Centre for Glycomics, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84538 Bratislava, Slovak Republic
Institute of Chemistry, Centre of Excellence for White-Green Biotechnology, Slovak Academy of Sciences, Trieda Andreja Hlinku 2, SK-94976 Nitra, Slovak Republic

It is presupposed that *N*-glycosylation encodes crucial information about the protein's structure-function, its age and its localization. Although *N*-glycosylation is often described as post-translational modification, recent studies show that it occurs co-translationally. During translation, ribosome is attached to a protein-translocation channel in the endoplasmic reticulum membrane and a nascent polypeptide is threaded through the channel, and oligosaccharyltransferase, adjacent to the channel, catalyses the attachments of oligosaccharide to the polypeptide chain before the protein acquire secondary or tertiary structure [1]. This commentary highlights connection between mentioned process of the glycoprotein folding and glycan–protein interactions on cell surfaces. Seems to be, both processes use special amino acid triplets (glycocodons) for programming of glycomics events, and this coding has relationship with the genetic code.

There is a belief that "eukaryotic cells form glycoproteins" and "prokaryotic cells form non-glycosylated proteins". In the eukaryotic cell, *N*-glycosylation occurs in the Endoplasmic Reticulum (ER) by Oligosaccharyltransferase (OSTs). However, OST homologues are also found in archaea and in bacteria [2,3]. It is possible to say that *N*-glycosylation is a phenomenon that integrates eukarya, bacteria and archaea, and the conservation of the structure and sequence of OSTs in these three forms of life suggests that similar glycosylation processes take place on the surfaces of bacterial and ER membranes, which is a simple confirmation of the endosymbiotic theory [4]. In other words, glyco-coding is not eukaryotic phenomenon, but it is older and had to originate before the last universal common ancestor. Even some researchers accept the theory that prebiotic peptides, which catalysed the first reactions in the primordial soup, established stereochemical principles that were taken later as bases for coding of biochemical reactions and cell life formation [5]. For example, homochiral L-dipeptides stereospecifically condensate glycolaldehyde to D-sugars [6]. The nature preferentially uses L-amino acids and D-sugars. In light of this, the types of peptide bonds that occur in monosaccharide recognizing proteins (lectins) were quantified recently [7]. This bioinformatics study revealed that the distribution of peptide bond types is quite similar for proteins (lectins) that recognize the same monosaccharide, and that typical pair of amino acids, which is related to a concrete monosaccharide, can be observed. The maximal occurrence of WS amino acids was detected for galactose (Gal), EW for fucose (Fuc), AY for mannose (Man), FS for *N*-acetylglucosamine (GlcNAc), MS and DD for *N*-acetylgalactosamine and MF for the glucose (Glc). When the search was simplified to GAVD amino acids, amino acids that are detected as main amino acids in carbonaceous meteorites [8] and main amino acids in Miller's experiments simulating amino acid synthesis in the primitive Earth [9], then GA amino acids were detected for Gal, DG for Fuc, GD for Man, VG for GlcNAc, VA and VV and DD for GalNAc, and AA for Glc. A comparison of hydropathy similarity of detected amino acids and the comparison with the genetic codon table revealed that modern amino acids pairs, related to monosugars, probably evolved from amino acids pairs established from prebiotic GAVD amino acid pool. Additionally is known that the sugar hydroxyl groups are well suited for directional acceptor/donor hydrogen (H) bonds (or the coordination of Ca^{2+} ions), and the polar amino acids, such as D, E, Q, N, and R, are well suited for cooperative H-bonding

[10]. It means that minimally one polar amino acid is needed to add to detected amino acid pairs for obtaining a basic peptide (glycocodon) which "recognizes" monosaccharide in a 3-D manner. As the result, AAR glycocodon can be recognized for Glc, GWN for Gal, QGD for Man and so on [7]. Human Galectin-3 in complex with LacNAc can serve as example how deduced glycocodons are used for Gal reading (Figure 1). The crystallized Carbohydrate Recognition Domain (CRD) shows that GW amino acid pair, inserted among polar amino acids, is used for Gal reading. It appears that GW detects hydrophobic patch over the C3-C6 galacto-positions and polar amino acids are responsible for the detection of sugar hydroxyl groups. Nowadays, the scientists start to understand how distant protein regions, that often do not adopt a well-defined structure, are involved in dynamic readout mechanism [11]. It was found that *N*-terminal non-CRD of the Galectin-3 has biological importance, the intact Gal-3 has a 3.8-fold higher affinity than the crystallized carbohydrate recognition domain (CRD) [12]. In figure 1, the sequences composed from the potential glycocodons of non-CRD are marked (Y is considered as polar, it shares the same second base in the DNA codons for polar amino acids; the biomolecular
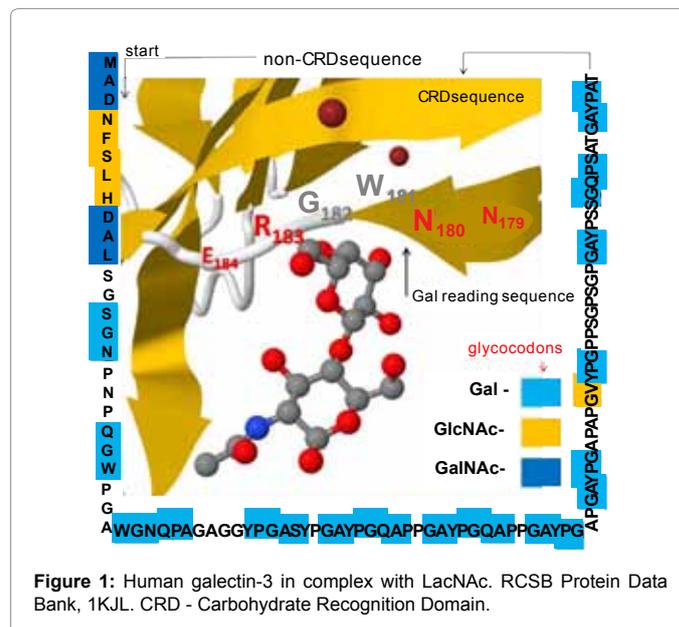

Figure 1: Human galectin-3 in complex with LacNAc. RCSB Protein Data Bank, 1KJL. CRD - Carbohydrate Recognition Domain.

Corresponding author: Jozef Nahalka, Institute of Chemistry, Centre for Glycomics, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84538 Bratislava, Slovak Republic, Tel: +421-2-59410255; Fax: +421-2-59410222; E-mail: nahalka@savba.sk

behavior of Y, in the context of the polypeptide chain, is probable polar). It seems that the non-CRD segment modulates conformational preference and flexibility of CRD by triplets – glycocodons distributed in the non-CRD.

Lectins, adhesins and other proteins, which are exposed on cell membrane surfaces, are responsible for glycomics coding in cell-cell interactions, but on the level of glycoprotein synthesis, OSTs are mainly responsible for implementation of influence of some glycan-sequences on the protein structure formation. The first structure of completely crystallized OST from *Campylobacter lari* (PGIB, 3RCE) showed that the conserved WWD triplet defines the substrate specificity, but it is not directly involved in catalysis [3]. In the case OSTs, similar glycocodon system seems be involved in the coding of glyco-programe. $W_{463}W_{464}D_{465}$ is hydropathically close to AAD and GWD, the glycocodons identified for glucose and galactose, and following sequence $Y_{466}G_{467}Y_{468}$ is close to mannose and fucose glycocodons, and $E_{457}D_{458}Y_{459}V_{460}V_{461}A_{462}$ sequence, that goes in front of $W_{463}W_{464}D_{465}$, can be divided into $E_{457}D_{458}Y_{459}$ and $Y_{459}V_{460}V_{461}$ glycocodons and $V_{461}A_{462}$ peptide bond specific for GalNAc (Y is considering as a "polar amino acid" according to the second base in the DNA codons). OSTs transfer isoprenoid-pyrophosphate-linked oligosaccharides to specific sequence motifs (sequons) on the asparagine side-chain of proteins. According to the glycocodon theory [7], there should be a correlation between the sequon and the glycocodon for directly connecting monosaccharide of the transferred oligosaccharide with the sequon's asparagine. The sequon for eukaryotic N-glycosylation is either N-X-S or N-X-T, where X is any amino acid except proline, S denoting serine and T threonine. Seems that N-X-S matches NFS glycocodon deduced for *N*-acetylglucosamine, and interestingly, adding the F to N–X–T motif doubles the number of proteins that can be stabilized by glycosylation without having to alter the native reverse turn type [13]. For bacterial PGIB, DQNATF has been recognized as an optimal acceptor sequence [14], DQN can be a GalNAc-glycocodon and TF is close to the FS - SF identified for GlcNAc.

Theories about the origin of genetic code can be divided to RNA world theories, protein world theories, co-evolution theories and stereochemical theories. All these theories take different approach with aim to find "the first word", the most ancient codons. Recent results confirm older proposals that a system of four codons ("gnc", n=a, g, c, u) and four amino acids (G, A, V, D) could be the original genetic code [7,15]. There is a connection, present day glycocodons can be simplified to triplets composed from GAVD amino acids. It indicates that prebiotic [GAVD]-peptides established "specificity" for each monosaccharide, and during evolution, it is likely that [GAVD]-glycocodons were transformed to novel glycocodons by positive selection for the increased diversity and functionality of "sugar-protein language" that can be made with a larger amino acid alphabet. Nevertheless, evolution process holds hydropathic similarity, amino acids in glycocodons are substituted with amino acids with similar polar properties, what minimizes errors in established sugar-protein interactions [7]. In conclusion, it says that the origin of the glycocode is related to the origin of the genetic code.

References

1. Gilmore R (2011) Structural biology: Porthole to catalysis. Nature 474: 292-293.

2. Yurist-Doutsch S, Magidovich H, Ventura VV, Hitchen PG, Dell A, et al. (2010) N-glycosylation in Archaea: on the coordinated actions of *Haloferax volcanii* AglF and AglM. Mol Microbiol 75: 1047-1058.

3. Lizak C, Gerber S, Numao S, Aebi M, Locher KP (2011) X-ray structure of a bacterial oligosaccharyltransferase. Nature 474: 350-355.

4. Sagan L (1967) On the origin of mitosing cells. J Theor Biol 14: 255-274.

5. van der Gulik P, Massar S, Gilis D, Buhrman H, Rooman M (2009) The first peptides: the evolutionary transition between prebiotic amino acids and early proteins. J Theor Biol 261: 531-539.

6. Weber AL, Pizzarello S (2006) The peptide-catalyzed stereospecific synthesis of tetroses: A possible model for prebiotic molecular evolution. PNAS 103: 12713-12717.

7. Nahalka J (2012) Glycocodon theory-the first table of glycocodons. J Theor Biol 307: 193-204.

8. Pizzarello S, Shock E (2010) The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. Cold Spring Harb Perspect Biol 2: a002105.

9. Johnson AP, Cleaves HJ, Dworkin JP, Glavin DP, Lazcano A, et al. (2008) The miller volcanic spark discharge experiment. Science 322: 404.

10. Gabius HJ, André S, Jiménez-Barbero J, Romero A, Solís D (2011) From lectin structure to functional glycomics: principles of the sugar code. Trends Biochem Sci 36: 298-313.

11. Fuxreiter M, Simon I, Bondos S (2011) Dynamic protein-DNA recognition: beyond what can be seen. Trends Biochem Sci 36: 415-423.

12. Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, et al. (2002) Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim Biophys Acta 1572: 232-254.

13. Price JL, Powers DL, Powers ET, Kelly JW (2011) Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. PNAS 108: 14127-14132.

14. Schwarz F, Huang W, Li C, Schulz BL, Lizak C, et al. (2010) A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol 6: 264-266.

15. Nahalka J (2011) Quantification of peptide bond types in human proteome indicates how DNA codons were assembled at prebiotic conditions. J Proteomics Bioinform 4: 153-159.