Glycan-RNA: a new class of non-coding RNA

Xiuling Li¹,², Tiing Jen Loh³, Jia Jia Lim³, Phei Er Saw¹,²,* and Yong Liao⁴,*

This commentary is based on a recent publication on the novel finding of a new class of non-coding RNA called glycan-RNA by Flynn et al.

Scientists have identified a new class of RNA functionalized with carbohydrates, which is present on cell surfaces. They have named this novel RNA glycoRNA according to its structure [1]. To better understand the characteristics of glycoRNA, we first describe basic information on RNA and glycosylation; we then discuss related topics on the identification and development of glycoRNAs.

RNAs

Ribonucleic acid (RNA), a type of nucleic acid, forms complex compounds with high molecular weight and takes deoxyribonucleic acid (DNA) as a template to produce protein while replacing DNA as a scaffold of genetic codes in some viruses, although RNA is limited to four bases [2, 3]. RNA functions primarily in cellular protein translation and synthesis, serving as a messenger carrying genetic information between DNA and ribosomes, helping ribosomes correctly assemble proteins, acting as biological catalysts, and participating in transcriptional and post-transcriptional genetic regulation [4, 5].

Of the many types of RNA, messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) are the three best known and most commonly studied. RNA can be broadly classified into coding RNA and noncoding RNA (ncRNA). Two classes of ncRNAs exist—housekeeping ncRNAs (tRNA and rRNA) and regulatory ncRNAs—which are further divided according to size: those longer than 200 nucleotides are denoted long ncRNAs (lncRNA), whereas those shorter than 200 nucleotides are denoted small ncRNAs [6]. According to their biogenesis, structure and action, lncRNAs can be divided into many different classes (Figure 1) [7]. Small ncRNAs are subclassified into small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small-interfering RNA (siRNA), PIWI-interacting RNA (piRNA) [8–10] and small nucleolytic ribozymes. These ribozymes are functional RNAs with enzymatic catalytic activity, which contain a loop or hairpin or hammerhead structure [11]. A large family of circular RNAs with self-cleaving functionality [12, 13] includes hammerhead ribozymes, hairpins, hepatitis delta virus RNA, glm3, twister, twister sister, hatchet, pistol and Varkud-satellite [14, 15].

RNA-associated interactions broadly include RNA-RNA interactions, RNA-DNA interactions and RNA-protein interactions, which participate in a variety of cell pathological and physiological activities, such as cell growth, proliferation, differentiation and death [16]. RNA-RNA interactions include those between miRNA and mRNA, which result in mRNA degradation [17, 18]; those between tRNA and mRNA, which result in translation of genetic information [19]; and those involving RNA splicing, maturation and decay. Ribosomes are composed of rRNA and protein, and they interact with DNA in the transcription of the genetic code. RNA-binding proteins further participate in the regulation of gene expression and cellular functions [20]. Different types of RNA have unique structures that endow them with functionality (Figure 2).

In the field of transcriptomics, more than 140 types of RNA modification have been identified in four nucleobases [21].

1Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
2Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
3Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia
4Key Laboratory of Molecular Biology for Infectious Diseases, Ministry of Education, Institute for Viral Hepatitis, and Department of Infectious Diseases, the Second Affiliated Hospital, Chongqing Medical University, Chongqing, China

*Correspondence to:
Yong Liao,
E-mail: yongliao@cqmu.edu.cn;
Phei Er Saw,
E-mail: caipeie@mail.sysu.edu.cn

Received: October 15 2021
Revised: November 23 2021
Accepted: May 22 2022
Published Online: August 9 2022

Available at:
https://bio-integration.org/
In addition to the 5′ cap and 3′ poly(A) tail, which have been extensively researched, five primary classes of RNA modification have been identified: adenosine methylation, cytosine modifications, uridine isomerization, ribose modification and cytidine acetylation [22–24]. Functional RNAs can use base alterations to regulate their structure and function, and their modification is required for stability and proper biogenesis [25–31]. Through splicing, translation and decay, RNA modification ultimately modulates protein production [21], in a process called post-transcriptional modification. Beyond the chemical modification of ribose and nucleosides, RNA glycosylation [1] is a newly identified class of RNA modification, thus expanding the domain of epigenetics.

**Glycosylation and glycans**

In glycosylation, a carbohydrate-like glycosyl donor is covalently attached to a hydroxyl or functional group of protein or lipid molecule, thereby forming a glycoconjugate through a series of enzymatic reactions in the Golgi apparatus and endoplasmic reticulum [32]. Glycosylation, the most complex post-translational modification of proteins and lipids, usually occurs at Ser/Thr residues (O-linked) or Asn-X-Ser/Thr consensus sequences (N-linked), thus forming a highly heterogeneous array of glycan structures on proteins or lipids [32, 33]. Glycoconjugates are formed...
from glycans and proteins or lipids through covalent bonding. According to their linkages with protein, lipid or glycan moieties, glycans are categorized into various types. The major glycoconjugate types on the cell membrane are glycoproteins, proteoglycans, glycosphingolipids and glycosyl-phosphatidylinositol (GPI)-anchored glycoproteins [34, 35]. Among the N-linked glycans and O-linked glycans, protein glycosylation involves the addition of glycosaminoglycans (GAGs), phosphorylated glycans, C-mannosylated tryptophan residues and GPI anchors to peptide backbones [36]. Glycosphingolipids are formed through lipid glycosylation in the secretory pathway; they serve as components of plasma membranes and are functionally associated with the formation of lipid rafts [37] (Figure 3).

Glycans are branched molecules composed of monosaccharides assembled through chemical bonds. Glycans are synthesized from seven monosaccharides: fucose (Fuc), sialic acid (SA), glucose (Glc), galactose (Gal), mannose (Man), N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc). Fuc and SA are usually found at the termini of glycosylated glycan chains [38]. In fucosylation and sialylation, Fuc and SA, respectively, are attached to a functional group or the hydroxyl of a protein.

Glycosylation enhances protein stabilization destabilizing unfolded proteins, on the basis of thermodynamic analysis [39]. N-glycosylation helps protein fold into the correct three-dimensional form to enable biological functions, such as cell-cell interactions and cell signaling [40, 41]. N-linked glycosylation can prevent the deamidation of glycoproteins and glycopeptides—a common pathway of protein degradation [42]. Glycosylated proteins are more stable than non-glycosylated proteins, owing to longer half-lives after deamidation. The mechanism by which glycosylation arrest proteins deamidation is that large N-glycans sterically hinder attack of the amine groups of side chains by backbone nitrogen, thereby preventing formation of a hydrolyzable succinimide intermediate that result in hydrolysis [43, 44]. Nascent glycoproteins bind membrane-bound calnexin and soluble calreticulin, two endoplasmic reticulum chaperones, which retain correctly unfolded proteins in the endoplasmic reticulum until the terminal glucose residue of the sugar chain is eliminated by glucosidase, and glycoprotein assembly is correctly completed [45, 46]. Protein glycosylation influences protein folding, degradation and trafficking to the proper destinations, such as targeting of receptors to the cell surface [47].

Glycosylation can alter the interactions between ligands and receptors, thus regulating signal transduction [35]. Sialylation is a switch-off signal of integrin glycan receptor combining with endogenous lectin galectin-1 (Gal-1) ligand, which may enable tumor cells to develop Gal-1 dependent anoikis resistance [48]. Obesity-induced insulin resistance is associated with the activity of FcγRIIB, an endothelial immunoglobulin G (IgG) receptor, which can be invalidated by sialylated IgG [49]. Glycosylation-dependent cell adhesion molecules are members of the L-selectin ligand family, whose sialylation and fucosylation modifications stimulate lymphocyte homing through tethering and rolling adhesion, as well as the integrin-activation pathway [50]. Blood-group antigens, the terminal epitopes generated by sialic acid and fucose residues on the sugar chains, are glycans conjugated with proteins [51, 52], thus indicating that glycans can be antigenic.

As immune system effectors, glycol forms are key in the synthesis, stability, signal recognition, function regulation

---

Figure 3  Major types of glycoconjugates on the cell membrane. Glycoproteins are a glycoconjugate class in which amino acid residues bind glycans and glycan chains via oxygen and nitrogen atoms of O-glycans and N-glycans. Proteoglycans are a class of glycoproteins characterized by long GAG chains with extended long disaccharide repeats (up to 50 disaccharide units) formed by GlcNAc or GalNAc linked to proteins. Glycosphingolipids consist of glycans and cellular membrane lipids. GPI-anchored glycoproteins are embedded in the outer plasma membrane and contain a glycan covalently attached to phosphatidylinositol. GlycoRNAs are a class of glycoconjugate in which glycans, such as Fuc or Sia, are conjugated to small conserved ncRNAs located at the cell surface. Geometric symbols depict types of sugars. Glc, glucose; GlcNAc, N-acetylgalactosamine; GlcA, glucuronic acid; Gal, galactose; GalNAc, N-acetylgalactosamine; IdoA, iduronic acid; Man, mannose; Sia, sialic acid; Fuc, fucose; Xyl, xylose.
and protein–protein interaction of immune proteins [53, 54].

The unassembled heavy chains of major histocompatibility complex I (MHC I) interact with membrane bound calnexin, and Asn residues must be modified with sugar chains for initial assembly. The transmembrane glycoprotein tapasin acts as a scaffold integrating a complex of MHC I, calreticulin, ERp57 and TAP transporter, thus forming mature MHC I [55, 56]. The integration of antigenic glycopeptides on peptide moieties and glycans defines the specificity of T cells [57].

Through the classical MHC relative way, monosaccharide and disaccharide alterations of glycopeptides are recognized by CD8+ and CD4+ T cells [58, 59].

Glycosylation modifications of proteins on the cell surface have been found to be associated with multidrug resistance [60]. In acute myeloid leukemia (AML) mouse models, inflammatory mediators, such as TNF-α and G-CSF [61, 62] released from AML cells alter the glycosylation of cell-surface ligand sugar chains by coordinating the activities of sialyltransferase and furcosyltransferase. Consequently, AML cell binding to endothelial E-selectin receptors [63] may cause chemotherapy resistance [64] through altering the endothelial niche microenvironment and activating the Akt/NF-κB/mTOR pathway. The expression of relative glycosylases is associated with DNA methylation and histone modulation [65, 66].

CD63 is a protein that can recruit or link receptor tyrosine kinases to integrins and Src family kinases, thus resulting in the development of cancer malignancy [67] via the N-glycosylation modification. Glycosylated CD63 is transported to the cell surface, where it has functions in invasion and drug resistance of breast cancer cells, whereas non-glycosylated CD63 remains trapped in the endosomal system, thus resulting in tumor cell non-malignancy [68]. N-glycan is a major component of the epidermal growth factor receptor (EGFR), which can regulate its cell surface expression and cause a drug resistance phenotype to EGFR inhibition [69].

N-glycosylated P-glycoprotein is more stable than non-glycosylated P-glycoprotein, and consequently is anchored to the cell membrane and promotes multidrug resistance [70–72].

Aberrant glycosylation of proteins may lead to disintegration of the cell membrane. In Aspergillus, the abnormal synthesis and transport of GDP-mannose, a mannosylation precursor, modulates cell viability, cell antigen phenotypes and cell membrane integrity by altering the function of GDP-mannose transporters and related catalytic enzymes [73, 74].

Glycosylation often occurs on proteins, lipids and glycans [34, 35], whereas RNA is usually not considered a major target of glycosylation. However, a new study has recently found that mammalian cell types and animals use RNA as a third scaffold substrate for glycosylation (Figure 3).

GlycoRNA: glycans are directly linked to RNA

Research has indicated a new biological phenomenon in the RNA field in which conserved small ncRNAs have been found to bear N-glycans. The term glycoRNAs has been coined to describe those small highly sialylated and fucosylated ncRNAs at the cell surface [31]. GlycoRNAs are a new breakthrough in the fields of RNA and glycobiology. GlycoRNA assembly depends on the canonical N-glycan biosynthetic machinery, which catalyzes sialic acid and fucose enrichment. These RNAs are localized on cell surface, where they interact with anti-dsRNA antibodies and members of the sialic acid-binding immunoglobulin-type lectins (Siglec) receptor family. GlycoRNAs are a class of small ncRNAs with a common set of transcripts, such as Y RNA, snRNA, rRNA, snoRNA and tRNA. The Y5 RNA transcript is strongly enriched in the pool of candidate glycoRNAs.

Y5 RNA is a member of family Y ncRNAs in human genes with a highly conserved stem-loop structure [75, 76]; it is transcribed by RNA polymerase III [77], and it frequently binds Ro60 protein, La protein or their orthologs in a loop-dependent manner [78, 79]. The localization of Y5 RNA is mostly nuclear [80, 81], and its transport to the cytoplasm is triggered by Ro60-dependent nuclear export, a lack of La-binding to the 3' end of Y RNA or its trimming [82, 83].

Other classes of glycoRNA transcripts, such as sn/snoRNAs, are localized to the cytoplasm and nucleus, respectively. Flynn et al. [31] have discovered that glycoRNAs are associated with the cellular membrane but not the soluble cytosol or nucleus, thus suggesting that interactions with glycan moieties on RNA transcripts might influence RNA localization and lead to yet-unknown functional changes.

Evidence indicates that glycoRNA glycans are structurally related to those found on proteins. Thus, study of the transferases participating in protein glycosylation may clarify the associations between RNA and glycoRNA-associated glycans. Glycan structures on glycoRNA have been defined as N-glycans, and only N-glycosylation enzymes can regulate the biosynthesis of glycoRNA. An apparent dose-dependent loss of glycoRNA label would occur after treatment with endoglycosidases, which are highly selective for N-glycans, and small molecule inhibitors of N-glycan trimming enzymes. Partial loss of label or even no effect has been observed after treatment with weakly selective N-glycan digesting enzymes or other enzymes, such as O-glycosidase and mucinase. Glycosylation-associated enzymes have been found to mediate glyco-complex biosynthesis in both the RNA and protein fields. Glycosyltransferase-encoding genes are epigenetically regulated by DNA methylation and histone acetylation, and produce specific type N-glycans in the proteome [84–86].

DNA methylation of genes also shapes the subclasses of IgG by modulating IgG glycan synthesis [87]. Because glycans have high homology, on the basis of proteomics and RNA nucleic omics, DNA methylation and histone acetylation can reasonably be speculated to tune the biosynthesis of glycoRNA through mechanisms similar to those of glycoproteins.

The team used an azide-labeled precursor of sialic acid, peracetylated N-azidoacetylmannosamine (Ac 4ManNAz), as an azidosugar to label living cells; they found that azide reactivity was focused on highly purified RNA preparations rather than other cell lysates [31]. The team had established methods based on metabolic labeling and biorthogonal chemistry to study protein-associated glycans several years prior [88–90]. The preliminary work was based on the hypothesis that labels with precursor sugars decorated with an azide group, called azidosugars, would be incorporated into cellular glycans and would undergo a biorthogonal reaction with...
Glycosylation-associated ncRNA vs. glycoRNA

Before identification of this new class of ncRNA, the relationship between RNA and glycans was known: ncRNA regulates glycosylation of proteins and the function of glycosyltransferase, which remodels glycans and influences cell activity. Protein glycosylation is a primary post-translational modification that substantially influences protein folding, localization, stability and activity. Ranging from simple monosaccharide modifications of nuclear transcription factors to highly complex branched polysaccharide modifications of cell-surface receptors, glycosylation encompasses diverse sugar addition to proteins.

N-acetylgalactosamine transferase (GALNT) is an enzyme that initiates the cascade of mucin type O-linked glycosylation, whose presence at the cell surface can lead to metabolic disorders and cancers [91]. Li et al. [92] have found that the lncRNA SNHG7 acts as a competing endogenous RNA that sponges miR-34a, thus blocking binding of GALNT and miR-34a. Without the limitation of miR-34a, GALNT expression in cancer tissues is strengthened, thus leading to cancer proliferation, invasion and metastasis in the context of aberrant O-glycosylation [93].

Specific fucosyltransferases (FUTs) play major roles in malignant cancer processes by catalyzing aberrant fucosylation. Xu et al. [94] have confirmed that the exosome-derived lncRNA MALAT1 directly competes for miR-26a/26b binding sites and increases FUT4 expression and fucosylation levels, thus promoting metastasis of colorectal cancer. The lncRNA HOTAIR, associated with poor clinical prognosis, sponges miR-326 and consequently regulates FUT levels by modifying the fucosylation of the E-selectin ligand [95]. Some microRNAs can also regulate other members of the FUT family, thereby altering glycan production [96, 97].

Being the substrate of sialyltransferase ST6GAL1, sialylation of activated epidermal growth factor receptor could be mediated and regulated by the ZFAS1/miR-150/ST6GAL1 axis, thus conferring a multi-drug resistance phenotype via the activated PI3K/AKT pathway [98]. ST6GAL1 and the lncRNA HOTAIR are direct targets of miR-214, and HOTAIR regulates the expression of ST6GAL1 by sponging miR-214 [99]. ST6GAL1 leads to metabolic sialylation of c-Met though the JAK2/STAT3 pathway, thus promoting colorectal cancer malignancy. Sialyltransferase family members interact with miRNA by altering glycosylation patterns, thus affecting the progression of breast cancer [96].

The newly discovered glycoRNAs differ from glycosylation-associated ncRNAs. Glycosylation-associated ncRNAs regulate glycan-associated protein expression and alter glycosylation patterns, thus inducing disease occurrence. These sugars are not directly linked to ncRNA under the reaction between ncRNA and related proteins. GlycoRNAs are a new class of ncRNA in which RNA is directly linked with glycans; the glycoRNA carries genetic information and is present on the cell surface, thus distinguishing it from glycosylation-associated ncRNA.

In glycosylation, glycans decorate other biological polymers, thus allowing cells to construct extensive molecular forms from the same DNA blueprint. Scientists had long believed that only proteins and lipids can be linked to carbohydrates. However, new research from Flynn R.A. et al. has indicated that RNA can also be glycosylated, and the glycosylation-modified nucleic acids are located on cell surfaces. This report has provided the first observation of this feature in the RNA field.

GlycoRNAs are located on the outer cell membrane, where they bind Siglecs; therefore, glycoRNAs may have roles in immune signal transduction. Siglecs are an immune receptor family associated with various diseases, such as systemic lupus erythematosus. Because of their ability to interact with anti-dsRNA antibodies and Siglec receptors, glycoRNAs may become a new serum marker of several diseases, thus enabling rapid clinical disease diagnosis. RNA modified with glycans may be sensitive to immunotherapy medicines [60] and consequently may serve as drug targets to increase drug response and drug enrichment in lesions. Immunotherapy, radiotherapy, chemotherapy and other methods may be used to alter the genetic information or glycan structure of glycoRNAs on the cell surface, thus therapeutically altering the lesion microenvironment. GlycoRNAs are expected to become a new type of immune response signal receiver, beyond protein receptors, in drug therapy.

Intriguingly, an emerging paradigm suggests that synthetic and clickable sugars that label glycoproteins and glycolipids can also modify small noncoding RNAs, thus making RNA the third scaffold for glycosylation. The first evidence indicating that highly sialylated and fucosylated glycoRNAs (RNA-glycan conjugates) are displayed on the cell surface, and can bind Siglec receptors and be decorated with N-glycans, has elevated studies of RNA biology, glycochemistry and the glycome (particularly the sialome and fucosome) to a new level of complexity. Although understanding and deciphering of the chemical structures and molecular functions of specific glycoRNAs on the cell surface remain limited, the discovery of glycoRNAs in mammalian cells may enable wider use and integration of glycoRNA data in traditional glycomics and omics workflows in cell biology, and enhance understanding of human diseases, thus enabling the design of novel glycan-based natural and glycoRNA mimetic therapeutics in the future.


Barvier B, Erbani F, Fivesh C, Davies JM, Tay J, et al. Endothe-


