**Minireview**

**Oligosaccharides Containing β1,4-Linked N-Acetylgalactosamine, a Paradigm for Protein-specific Glycosylation**

Stephen M. Manzella, Lora V. Hooper, and J. jacques U. Baenziger

From the Department of Pathology, Washington University School of Medicine, St. Louis, Missouri 63110

The carbohydrate moieties found on glycoproteins have long been recognized as having great potential to bear biologically important information. However, actual examples of systems in which oligosaccharides play defined physiological roles have remained limited. These oligosaccharides with known biologic functions typically have distinctive structural features and are generally confined to specific glycoproteins. Synthesis of structurally unique oligosaccharides on specific glycoproteins at defined times is essential if these structures are to fulfill their biologic purpose. Since cells produce many distinct oligosaccharides as newly synthesized glycoproteins pass through the endoplasmic reticulum and the Golgi, mechanisms are required to assure that the correct structures are added to the numerous glycoproteins being synthesized. Determining how synthesis of the vast array of oligosaccharides is brought about is essential for understanding the biologic importance of these complex structures.

Asn-linked oligosaccharides arise by processing of a common precursor structure, which is transferred en bloc from dolichol to the nascent peptide chain in the endoplasmic reticulum (1). As a result Asn-linked oligosaccharides have a common core region and differ primarily in the number and location of their peripheral branches and terminal modifications. Since all newly synthesized glycoproteins pass through the same subcellular compartments and are exposed to the same transferases, structural differences in oligosaccharides on individual glycoproteins and/or at individual glycosylation sites must in some fashion reflect the influence of the protein moiety on one or more glycosyltransferases. This suggests that key glycosyltransferases recognize features encoded within the peptide as well as the oligosaccharide of the target glycoprotein. Among the three glycosyltransferases thus far demonstrated to display peptide as well as oligosaccharide recognition, UDP-glucosylglycoprotein glucosyltransferase, UDP-N-acetylgalactosamine:lyosomal enzyme N-acetylgalactosamine-1-phosphotransferase, and UDP-GalNAc:glycoprotein hormone (β1,4-N-acetylgalactosaminyltansferase (β1,4-GalNACT), reviewed in Ref. 2), one of the most extensively characterized is the β1,4-GalNACT, which produces the terminal sequence GalNAcβ1,4GlcNAcβ1-R on glycoproteins that contain a specific peptide recognition determinant in addition to an appropriate oligosaccharide acceptor. The product of the β1,4-GalNACT may be further modified by the addition of sulfate, sialic acid, or fucose, thus producing a range of unique oligosaccharide structures defined by the presence of β1,4-linked GalNAc as illustrated in Fig. 1. Each of these structures has the potential to be recognized by a specific receptor or binding protein and thus mediate a distinct biological function. As will become apparent below, the β1,4-GalNACT is a key component of a well characterized system, which includes unique oligosaccharide structures, highly specific glycosyltransferases, and oligosaccharide-specific receptors. This is therefore an excellent model system for understanding protein-specific glycosylation.

**A Peptide Recognition Determinant, Which Directs Synthesis of Unique Oligosaccharides**

Lutropin (LH), a glycoprotein hormone produced in the anterior lobe of the pituitary, was the first glycoprotein shown to have Asn-linked oligosaccharides with one or two branches terminating with the sequence SO$_4$-4GalNAcβ1,4GlcNAcβ1,2Manα (3). Oligosaccharides on LH, thyrotropin (TSH), and the free glycoprotein hormone α subunit from a number of different animal species were subsequently shown to terminate with the same sequence (4–7). A growing number of glycoproteins, which are unrelated to the glycoprotein hormones but which bear oligosaccharides terminating with the sequence SO$_4$-4GalNAcβ1,4GlcNAcβ1,2Mana (6–10, 37, 39, 42), are synthesized in the pituitary as well as other tissues. Glycoproteins that bear oligosaccharides terminating with GalNAcβ1,4GlcNAcβ1-R (6–9, 37, 39, 40, 42–50), Siaα2,3GalNAcβ1,4GlcNAcβ1-R (7, 40, 42, 47–50), and GalNAcβ1,4(Fucu1,3)-GlcnAcβ1-R (40, 42, 44–46) have also been described from a number of sources. Furthermore, pro-opiomelanocortin, which contains Asn-linked oligosaccharides terminating with SO$_4$-4GalNAcβ1, 4GlcnAcβ1-R (8, 9), is the first instance of a glycoprotein bearing Ser/Thr-linked oligosaccharides terminating with this same sequence (10).

Even though a number of glycoproteins bearing oligosaccharides with the sequence GalNAcβ1,4GlcNAcβ1-R have been described, this carbohydrate structural motif is not common in vertebrates. The vast majority of other glycoproteins produced by the tissues or cells, which synthesize glycoproteins bearing β1,4-linked GalNAc, do not contain β1,4-linked GalNAc but instead contain β1,4-linked Gal, indicating that the addition of β1,4-linked GalNAc is generally a highly protein-specific process. A β1,4-GalNACT, which can account for the specific modification of Asn-linked oligosaccharides, is present in a limited number of tissues and cell lines, including those which are known to produce oligosaccharides with the β1,4-linked GalNAc motif (11). In contrast, β1,4-galactosyltransferase (β1,4-GalT), which transfers Gal in β1,4-linkage to virtually any terminal GlcnAc is expressed at relatively high levels in virtually all vertebrate tissues and cell lines (12, 13). Since the β1,4-GalNACT and β1,4-GalT compete for the same oligosaccharide acceptors (Fig. 1), preferential addition of GalNAc to an oligosaccharide must reflect recognition of the protein bearing the oligosaccharide.

In vitro model system was established for examining the protein specificity of the β1,4-GalNACT using human chorionic gonadotropin (hCG), hCG, which is closely related to LH but is synthesized in the placenta (14), binds to the same receptor as LH. hCG contains the peptide recognition determinant utilized by the β1,4-GalNACT but bears oligosaccharides that terminate with Siaα2,3Gal because neither the β1,4-GalNACT nor the GalNAcβ1,4-sulfotransferase (reactions 3 and 4, respectively, in Fig. 1) are expressed in human placenta (15). We established the existence of the peptide recognition determinant by comparing glycopeptides and glycopeptides bearing the different oligosaccharide, GlcnAcβ1,2Manβ1,4GlcNAcβ1Asn (the product of reaction 1 in Fig. 1), as acceptors for the addition of either Gal or GalNAc by transferases present in pituitary extracts. Gal is added to each of the glycopeptides tested with the same apparent $K_m$ of 1–2 mM and the same catalytic efficiency ($V_{max}/K_m$). In contrast, the apparent $K_m$ for addition of

---

*This minireview will be reprinted in the 1996 Minireview Compendium, which will be available in December, 1996. This work was supported by National Institutes of Health Grants R01-DK-41738 and R01-CA21923 to J.U.B.† To whom correspondence should be addressed. Tel.: 314-362-8730; Fax: 314-362-8888; E-mail: Baenziger@Pathology.WUSTL.edu.

‡The transferases used are β1,4-GalNACT, β1,4-N-acetylgalactosaminyltansferase, β1,4-GalT, β1,4-galactosyltransferase, CG, chorionic gonadotropin; hCG, human chorionic gonadotropin; FSH, follicitropin; LH, lutropin; TSH, thyrotropin; R, underlying oligosaccharide core.

‡‡S. M. Manzella, S. M. Dharmesh, C. Cohick, M. J. Soares, and J. U. Baenziger, unpublished data.
GalNAc or Gal. The product of reaction 1 is a common synthetic intermediate, which can be further modified by the addition of 1,4-linked Gal, reaction 2, or 1,4-linked GalNAc, reaction 3. The 1,4-GalT does not display any peptide specificity and will transfer Gal to any β-linked terminal GlcNAc with the same catalytic efficiency. In contrast the 1,4-GalNAcT recognizes features encoded in the peptide (i.e. the 1,4-GalNAc - recognition determinant) as well as the terminal β-linked GlcNAc. In the presence of the recognition determinant the catalytic efficiency for GalNAc addition to the same oligosaccharide is as much as 500-fold greater than for addition of either Gal or addition of GalNAc in the absence of the recognition determinant (15, 17, 18). The 1,4-Fucosyltransferase will transfer fucose to GlcNAc, in the presence of either 1,4-linked Gal (reaction 8) or GalNAc (reaction 5) (51), and α2,6-sialyltransferase will transfer sialic acid to either 1,4-linked Gal (reaction 7) or GalNAc (reaction 6) (52). In contrast, α2,3-sialyltransferase will add sialic acid to 1,4-linked Gal (reaction 9) but not GalNAc (37), while GalNAc-4-sulfotransferase will add sulfate to the 4-hydroxy of 1,4-linked GalNAc (reaction 4) but not to 1,4-linked Gal (53, 54). However, all of the oligosaccharides containing 1,4-linked GalNAc are distinct from those containing Gal and are largely confined to glycoproteins containing the 1,4-GalNAc recognition determinant, which are synthesized by cells expressing the 1,4-GalNAcT. GalNAc or Gal is transferred to oligosaccharides on glycoproteins such as transferrin with an apparent K_m of 5–10 μM. Thus, the catalytic efficiency for addition of GalNAc to oligosaccharides on certain glycoproteins is 100–500-fold greater than for transfer to the same oligosaccharides on other glycoproteins, indicating the presence of a specific 1,4-GalNAc recognition determinant on a particular subset of glycoproteins including CG (15–17). We have located two recognition determinants on CG, one on the a subunit and one on the β subunit (17), and have established a number of features that are critical for recognition by the 1,4-GalNAcT (18).

The recently solved crystal structure of CG (19, 20) including four Asn-linked oligosaccharides, illustrated in Fig. 2, has greatly enhanced our understanding of the 1,4-GalNAcT recognition determinant. It is immediately apparent that each of the four Asn-linked oligosaccharides on CG is nearly as large as the peptide portion of either the α or β subunit, and only the innermost 2 or 3 sugars are in direct contact with the peptide. Since the oligosaccharides are highly mobile rather than being fixed in space, their molecular dimensions are even greater relative to the peptide when considered in real time. It is also apparent that the peripheral sugars, Sia-Gal or SO_4GalNAc, are distant from the peptide, mobile, and highly exposed, as has been recently been established for at least one of the α subunit oligosaccharides by high resolution multinuclear NMR. The structures of LH, TSH, and follitropin (FSH), which have the same α subunit and highly homologous β subunits, are likely to be similar to that of hCG. The regions of the α subunit (Fig. 2, A and B) and the β subunit (Fig. 2B only) which include residues essential for recognition, have been highlighted in yellow. Residues in the α subunit critical for recognition by the 1,4-GalNAcT include the basic amino acids in the sequence Pro^5-Leu^6-Arg^7-Pro^8-Arg^9-Cys^10-Arg^11-lys^12-lys^13 (17). We have shown that the Arg^8 is essential for recognition and are currently examining the contribution of other residues within this sequence.

The evidence is compelling that the peptide recognition determinants on the glycoprotein hormones reduce the apparent K_m for GalNAc addition to the oligosaccharides by high resolution NMR, and that these structures are conserved among vertebrates. As a result residues critical for recognition determi-
brates ranging from fish to humans. Furthermore, the oligosaccharides on glycoprotein hormones from all classes of vertebrates terminate with GalNAc-4-SO_4. (24). This is the first instance in which a specific oligosaccharide structural motif has been shown to be maintained on a family of glycoproteins from different classes of vertebrates. Thus, the unique carbohydrate structural motif, like the sequence and structure of the glycoprotein hormone peptides, has been conserved during the evolution of vertebrate species.

LH is a major product of the gonadotroph and is one of only a few proteins produced by gonadotrophs or other cells in the pituitary that terminate with GalNAc-4-SO_4. The expression of β1,4-GalNACT and GalNAc-4-sulfotransferase in the gonadotroph is modulated in parallel to LH levels in response to circulating levels of estrogen (26). As estrogen levels fall, the expression of β1,4-GalNACT and GalNAc-4-sulfotransferase increases in concert with increased synthesis of LH. The coordinate regulation of LH synthesis and β1,4-GalNACT expression assures that the oligosaccharides on LH, but not other glycoproteins produced in the gonadotroph, always terminate with GalNAc-4-SO_4. In contrast, β1,4-GalNACT activity in other tissues including the submaxillary gland and kidney is not responsive to estrogen levels. Conservation of these sulfated oligosaccharide structures during evolution in conjunction with hormonal regulation of β1,4-GalNACT expression in gonadotrophs but not other cells further supports the view that these sulfated oligosaccharides play a central role in the biology of the glycoprotein hormones.

**The Biological Significance of GalNAc-4-SO_4 for the Glycoprotein Hormones**

Consistent with the high degree of regulation seen for the synthesis of sulfated oligosaccharides on LH, these oligosaccharides have been found to mediate a crucial biological function. We have shown that the sulfated oligosaccharides on LH regulate its circulating half-life following release into the blood (27). These oligosaccharides are recognized by a receptor in hepatic endothelial cells and Kupffer cells, which is specific for the terminal sequence SO_4GalNAcβ1,4GlcNAcβ1,2Manα1,R (28). Upon binding the GalNAc-4-SO_4 receptor the hormone is rapidly internalized and transported to lysosomes where it is degraded. The receptor is plentiful with 500,000 binding sites detectable at the surface of each endothelial cell and has an apparent Km of 1.63 × 10^-7 M for LH. The rapid and specific clearance of LH from the circulation on the basis of its terminal glycosylation was initially unexpected since rapid clearance of the hormone reduces its potency to initiate ovulation following a single intravenous injection (27). This seeming contradiction is resolved upon consideration of the properties of the LH/CG receptor in the ovary and the hormonal cycle. The LH/CG receptor is a member of the seven-transmembrane domain G-protein-coupled receptor family. Upon hormone binding the receptor is activated and CAMP is produced; however, at the same time, hormone binding causes rapid inactivation and internalization of the receptor (29). As a result, continuous stimulation would result in the entire population of LH/CG receptors becoming refractory to further activation. Thus, during the 24–48 h pre-ovulatory surge in circulating LH levels, the LH/CG receptor would not be maximally activated due to down-regulation. However, circulating LH levels rise and fall in a pulsatile manner. During the preovulatory surge it is the frequency and amplitude of these pulses that increases (30, 31). This pulsatile rise and fall reflects both the episodic release of LH from granules and its rapid clearance from the blood. Other hormones such as FSH are also released episodically but have a long half-life and do not display this pulsatile rise and fall in blood levels. We have therefore proposed that this pulsatile rise and fall in LH levels is essential to obtain maximal stimulation of the LH/CG receptor. Key to the pulsatile appearance of LH in the circulation is its rapid clearance from the bloodstream mediated by its oligosaccharide component.

The crucial role mediated by these oligosaccharides is highlighted by the fact that a number of animal species, including humans and horses, synthesize a glycoprotein hormone, CG, in their placenta during the early stages of pregnancy. Equine CG and LH arise for the same gene and have identical peptide sequences (32, 33). The Asn-linked oligosaccharides on equine CG terminate with Siaα2,3Galβ1,4GlcNAcβ1,R while those on equine
**Minireview:** \(\beta\)-1,4-Linked N-Acetylgalactosamine Oligosaccharides

LH terminate with SO\(_4\)-GalNAc\(\beta\)1,4GlcNAc\(\beta\)-1-R (34, 36). Consistent with the presence of sialic acid-bearing oligosaccharides, LG has a long circulatory half-life. Thus equine CG and LH are different glycoforms of the same protein, which we have shown differ in their rate and site of clearance from the circulation (34). Furthermore, LH is stored in granules and released episodically into the circulation in response to a releasing factor while CG, which is not stored in granules, is released continuously from placental trophoblasts. Thus, the major difference between LH, the hormone of the ovulatory cycle, and CG, the hormone of pregnancy, is the difference in their circulatory half-lives, which results in episodic and continual stimulation of the LH/CG receptor, respectively.

Other Roles for Oligosaccharides Containing \(\beta\)-1,4-Linked GalNAc

The sulfated oligosaccharides on LG illustrate how unique oligosaccharide structures can play crucial physiologic roles. We have defined and characterized many of the components of the physiologic system involving these oligosaccharides, including 1) their precise structures; 2) the transferases responsible for the synthesis of these structures; and 3) a receptor that specifically recognizes these sulfated oligosaccharides and mediates a specific biological function. Our results, furthermore, demonstrate that this system involves a high degree of regulation and specificity.

Not surprisingly many of these same elements are used at different times and under different circumstances for other biologic purposes. As was noted above, the number of glycophosphorylated proteins known to contain oligosaccharides terminating with the sequence GalNAc\(\beta\)1,4GlcNAc\(\beta\)-1-R has increased. The addition of SO\(_4\), \(\alpha\)1,3-linked fucose, or \(\alpha\),2,6-linked sialic acid (Fig. 1, products of reactions 4, 5, and 6, respectively) has the potential to produce three additional and unique oligosaccharide structures, each of which can potentially be recognized by a specific receptor similar to the hepatic GalNAc-4-SO\(_4\) specific receptor. Different glycoforms of the same protein may also arise at different times during development or in response to hormonal status (26, 41). Thus, these glycoforms may fulfill a variety of biological purposes.

The demonstration of protein-specific glycosylation by the \(\beta\)-1,4-GalNAc\(\beta\), which acts on the glycoprotein hormones, is particularly exciting because it provides a model for understanding how these and other distinct oligosaccharide structures are synthesized in a protein- and even site-specific manner. It also exemplifies a mechanism for the addition of unique structures at precise times to specific glycophosphorylated proteins. This is essential for oligosaccharides having biological roles, which involve encoding of specific information. Oligosaccharides are ideally suited for such a purpose because they are highly exposed and accessible at the surface of the proteins which bear them and because of their enormous structural diversity. It is likely that we have only gained a glimpse of the potential functions of carbohydrates thus far and that many new discoveries await those willing to embark on the study of this form of post-translational modification.

Acknowledgment—We thank Dr. Steve W. Homans, University of St. Andrews, St. Andrews, UK for providing us with the three-dimensional image of HCG and its oligosaccharides.