A limited number of glycoproteins including luteinizing hormone and carbonic anhydrase-VI (CA6) bear N-linked oligosaccharides that are modified with β1,4-linked N-acetylgalactosamine (GalNAc). The selective addition of GalNAc to these glycoproteins requires that the β1,4-N-acetylgalactosaminyltransferase (βGT) recognize both the oligosaccharide acceptor and a peptide recognition determinant on the substrate glycoprotein. We report here that two recently cloned βGTs, βGT3 and βGT4, that are able to transfer GalNAc to GlcNAc in β1,4-linkage display the necessary glycosyltransfer specificity in vivo. Both βGTs transfer GalNAc to N-linked oligosaccharides on the luteinizing hormone α subunit and CA6 but not to those on transferrin (Trf). A single peptide recognition determinant encoded in the carboxyl-terminal 19-amino acid sequence of bovine CA6 mediates transfer of GalNAc to each of its two N-linked oligosaccharides. The addition of this 19-amino acid sequence to the carboxy terminus of Trf confers full acceptor activity onto Trf for both βGT3 and βGT4 in vivo. The complete 19-amino acid sequence is required for optimal GalNAc addition in vivo, indicating that the peptide sequence is both necessary and sufficient for recognition by βGT3 and βGT4.

Unique carbohydrate structures found on a number of glycoproteins are believed to contribute to biological functions ranging from transport of lysosomal enzymes to lysosomes (1, 2), to regulation of circulatory half-life of hormones (3, 4), to cellular recognition (5, 6). The addition of unique carbohydrate structures to select glycoproteins requires a “recognition determinant,” a feature encoded in the peptide portion of these glycoproteins, to be recognized by one or more of the transferases responsible for their synthesis. In some cases key features of the peptide for such recognition determinants have been identified (7–11). However, understanding the molecular basis for recognition ultimately requires that a recognition determinant that does not itself comprise the site of modification can be added to a structurally unrelated glycoprotein and confer selective modification of this unrelated glycoprotein.

A limited number of glycoproteins bear N-linked or O-linked oligosaccharides that are modified with β1,4-linked GalNAc rather than β1,4-linked Gal. We first described such structures on the pituitary glycoprotein hormone LH, where it is further modified with sulfate to form terminal β1,4-linked GalNAc-4-SO₄ (12). We have recently shown that these sulfated structures contribute to regulation of the level of steroid hormone produced because they determine the circulatory half-life and thereby the circulating levels of LH in vivo (13–15). Other glycoproteins that have different functions in vivo such as carbonic anhydrase-VI (CA6) (16), glycolelin (17), prolactin-like proteins (18), proopiomelanocortin (19), SorLA/LR11 (20), sialoadhesin (21), and tenasin-R (22) are also modified with β1,4-linked GalNAc.

In in vitro studies we demonstrated that one or more βGTs present in the pituitary (23) and other tissues (24) are protein-specific, adding GalNAc to oligosaccharide acceptors on proteins that contain a peptide recognition determinant. The catalytic efficiency for this addition is 500-fold higher than for GalNAc transfer to the same oligosaccharide acceptors on proteins lacking the determinant in their peptide sequence (23, 25). We reported that the basic amino acids within the sequence PLRSKK that is present in the pituitary glycoprotein hormone α subunit are critical elements of the peptide recognition determinant conferring transfer of GalNAc to N-linked oligosaccharides in vitro (26, 27). Two recently cloned βGTs, βGT3 (28, 29), and βGT4 (30) that are closely related transfer GalNAc in β1,4-linkage to terminal GlcNAc. They are expressed in tissues that have previously been determined to express protein-specific GalNAc transferase activity, but they have not yet themselves been shown to be protein-specific either in vitro or in vivo.

We now report the development of unique chimeric glycoprotein constructs that allow us to examine the efficiency of GalNAc addition to N-linked carbohydrates on these glycoproteins in the complex milieu of the Golgi. The chimeric proteins consist of either Renilla (31–33) or Gaussia (34, 35) luciferase.

* This work was supported by National Institutes of Health Grant RO1-DK41738 (to J. U. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby indicated as advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
1 Present address: Graduate School of Biomedical Sciences, University of Texas Health Science Center San Antonio, 7703 Floyd Curl Dr., MC 7819, San Antonio, TX 78229-3900.
2 Present address: Transkaryotic Therapies, Inc., 700 Main St., Cambridge, MA 02139.
3 To whom correspondence should be addressed: Washington University School of Medicine, Dept. of Pathology and Immunology, 4940 Parkview Place, St. Louis, MO 63110. Fax: 314-362-8888; E-mail: Baenziger@wustl.edu.
Protein-selective Glycosylation Determinant

and glycoproteins that do or do not bear N-linked oligosaccharides modified with β1,4-linked GalNAc. By expressing these chimeric glycoproteins in cells that express either βGT3 or βGT4, we were able to quantitatively compare their extent of modification with β1,4-linked GalNAc in vivo. Using this approach we show that βGT3 and βGT4 are indeed protein-specific, recognizing a peptide determinant in the protein substrate to allow efficient and protein-selective transfer of GalNAc to their oligosaccharides. We have identified a 19-amino acid sequence that is both necessary and sufficient to mediate protein-specific GalNAc addition by βGT3 and βGT4. The addition of this sequence to a glycoprotein that is not recognized by either βGT3 or βGT4 converts the glycoprotein into one that is selectively modified with β1,4-linked GalNAc in vivo.

MATERIALS AND METHODS

Luciferase Constructs—Construction of the Renilla luciferase (RLuc) (32, 33) chimeras α-RLuc, transferrin-RLuc (Trf-RLuc), RLuc-α, and RLuc-CA6 and the Gaussia luciferase (GLuc) (34, 35) chimeras GLuc-α, GLuc-Trf, and GLuc-CA6 will be described in detail elsewhere. Each expression plasmid was made from pcDNA3.1 with a cytomegalovirus promoter to drive expression (Invitrogen). The chimeric glycoproteins α-RLuc and Trf-RLuc consist of the glycoprotein hormone α subunit or Trf followed by RLuc epitope-tagged at its carboxyl terminus with V5His. RLuc-CA6 and RLuc-α were constructed using the pSec-Tag expression plasmid (Invitrogen) and consist of the IgG leader sequence followed by RLuc and CA6 or α epitope-tagged at their carboxyl termini with MycHis. pCMV-GLuc was obtained from New England Biolabs. The GLuc constructs in each case consist of GLuc followed by α, Trf, or CA6 and the epitope MycHis. Because GLuc is a secreted protein, no additional leader is required, and the leader sequences of α, Trf, and CA6 were omitted from the constructs. All of the cDNAs were amplified using KlenTaq Long and Accurate DNA polymerase (KTLa polymerase) (36, 37). The sequence encoding additional amino acids from CA6 was added to Trf using KTLA polymerase and ribocloning as described (38). A schematic of the key luciferase chimeras generated including the locations of the glycosylation sites and the amino acid sequences that constitute the recognition sequences is shown in Fig. 1.

Cell Lines—HEK 293T cells were maintained in Pro293A-CDM medium (Lonza). Individual clones of Flp-In™ CHO (Invitrogen) cells expressing βGT3 (βGT3/CHO) or βGT4 (βGT4/CHO) under control of the EF-1α promoter were prepared and characterized as described elsewhere. βGT3/CHO and βGT4/CHO were maintained in Ham’s F-12 medium, 10% fetal bovine serum, 2 mM glutamine, 600 µg/ml hygromycin. Transfections were performed using Lipofectamine-2000 (Invitrogen).

Biotinylated Wisteria Floribundia Agglutinin (WFA)—WFA was stored in aliquots at −20 °C until use.

Assay for GalNAc Addition to Luciferase Chimeras—Microplate—96-well plates were coated with Streptavidin (Roche Applied Science) by adding 100 µl of 25 mM sodium carbonate buffer, pH 8.5, containing 1 µg of streptavidin and incubating at 37 °C for 3 h. The plates were then washed six times with PBS containing 0.1% bovine serum albumin (BSA) using a Bio-Rad Immunowasher. Biotinylated WFA (0.25 µg in 100 µl of PBS/well) was incubated with the immobilized streptavidin overnight at 4 °C. Each well was washed six times with 300 µl of cold PBS, 0.1% BSA and then blocked by incubating for 30 min at 25 °C with 300 µl of PBS, 5% BSA. After washing six times with PBS, 0.1% BSA, aliquots of luciferase chimeras containing 50,000 light units (LU) of luciferase activity in 100 µl of PBS, 0.1% BSA were incubated with the biotinylated WFA-coated wells for 4 h at 4 °C in the presence or absence of 50 mM GalNAc. The wells were washed six times with PBS, 0.1% BSA, and 20 µl of PBS was added to each well. The amount of bound luciferase activity was measured using a Wallac Victor2 luminometer by injecting 50 µl of luciferase assay buffer containing freshly diluted coelenterazine (New England Biolabs) into each individual well and determining the LU produced over a period of 10 s. The GalNAc-specific LU bound were calculated by subtracting the LU bound in presence of 50 mM GalNAc from the LU bound in the absence of GalNAc. The background in each case was less than 250 LU and was subtracted from the analyses shown.

6 N. M. J. Blake, Y. Mi, E. L. Oates, M. Beranek, and J. U. Baenziger, submitted for publication.
FIGURE 2. The KRKKEK sequence near the carboxyl terminus of CA6 is required for optimal GalNAc addition in vivo. CA6(Wt) and CA6(Mu1) epitope-tagged at their carboxyl termini with V5His were expressed in HEK 293T cells. Culture medium containing equal amounts of CA6(Wt) and CA6(Mu1) (M) was incubated with WFA immobilized on agarose. After removing medium containing unbound CA6 (UB), the WFA-agarose was washed with equal volumes of buffer (W1). Bound CA6 was eluted with four successive incubations with buffer containing 50 mM GalNAc (E1–E4). Any CA6 that remained bound to the WFA-agarose was then recovered by heating the WFA-agarose in SDS-PAGE loading buffer (E5). The amount of CA6(Wt) and CA6(Mu1) in equal aliquots of M and UB fractions and larger aliquots of W1 and E1–E5 was determined following SDS-PAGE using Pre-Cast NuPAGE® gels (Invitrogen) and electrophoretic transfer to nitrocellulose. The amount of epitope-tagged protein present was determined by Western blot analysis using mouse anti-V5 antibody (Invitrogen) and Alexa Fluor 680-labeled goat anti-mouse IgG (Molecular Probes). IR-induced fluorescence was used for detection. CA6(Wt) synthesized by CHO cells that do not express βGT activity was not bound by WFA-agarose (not shown), indicating that binding is GalNAc-dependent and that the inability to elute all of the bound CA6(Wt) with GalNAc reflects the inefficiency of elution rather than nonspecific binding.

RESULTS

Protein-selective Addition of GalNAc to CA6—Bovine CA6, a secreted form of carbonic anhydrase, bears N-linked oligosaccharides that are modified with β1,4-linked GalNAc when it is expressed by salivary or lacrymal glands (16). We have previously shown that HEK 293T cells endogenously express protein-specific βGTs that are able to modify the N-linked oligosaccharides on LH and other glycoproteins with β1,4-linked GalNAc (20, 24, 39–41). When native CA6(Wt) epitope-tagged with V5His at its carboxyl terminus is expressed in HEK 293T cells, at least 51% of the secreted CA6(Wt) is bound by immobilized WFA, a lectin that binds oligosaccharides bearing terminal β1,4-linked GalNAc (42, 43) (Fig. 2). Even though the β1,4-linked GalNAc added to N-linked oligosaccharides on glycoproteins expressed by HEK 293T cells can be further modified with either SO₄ or α₂,6-linked sialic acid, little CA6(Wt) remains in the unbound fraction. The major fraction of bound CA6(Wt) is selectively eluted with GalNAc (Fig. 2, lanes E1–E4). The CA6(Wt) eluted by warming in SDS-PAGE loading buffer reflects the inefficiency of elution with GalNAc because CA6(Wt) expressed in CHO cells does not contain any GalNAc and is not retained by immobilized WFA (not shown). Thus, the N-linked oligosaccharides on CA6(Wt) expressed in HEK 293T cells are modified with β1,4-linked GalNAc, suggesting that, like the glycoprotein hormone α subunit, CA6(Wt) has a recognition determinant that results in its selective modification with GalNAc when expressed by HEK 293T cells.

The 19-amino acid sequence located at the carboxyl terminus of bovine CA6 contains 8 basic amino acids (see CA6(Wt) in Fig. 1). Modeling programs such as that of Chou and Fasman (44) predict that this sequence is likely to form an α helix. This region of CA6 may therefore resemble the PLRSKK sequence of the glycoprotein hormone α subunit that forms an α helix in the dimeric form of the hormone (45, 46) and is essential for recognition by pituitary βGT in vitro assays (26). The sequence KRKKEK was therefore deleted from the carboxyl-terminal 19-amino acid sequence of CA6 (see CA6(Mu1) in Fig. 1). In contrast to CA6(Wt), less than 13% of the CA6(Mu1) produced by HEK 293T cells is bound by immobilized WFA and eluted from the WFA by GalNAc. Furthermore, little CA6(Mu1) is subsequently eluted by SDS-PAGE loading buffer (see Fig. 2). CA6 therefore contains a recognition determinant that is utilized by the βGT(s) expressed by HEK 293T cells, and the KRKKEK sequence is a critical element of this recognition sequence.

Luciferase Chimeras Replicate the Properties of Native Glycoproteins in Vivo—Western blot analysis for quantitation of the amount of a glycoprotein that is modified with terminal β1,4-linked GalNAc when expressed in different cells is cumbersome. We therefore prepared chimeric glycoproteins that consist of either RLuc or GLuc and the glycoprotein of interest (Fig. 1). The RLuc chimeras consist of the glycoprotein followed by RLuc at the carboxyl terminus followed by the V5His epitope. α-RLuc, Trf-RLuc, and CA6-RLuc are efficiently secreted into the culture medium following transfection of cultured HEK 293T and CHO/Flp-In cells (not shown). The extent of GalNAc addition to each of these luciferase chimeras can be compared by immobilizing biotinylated WFA onto streptavidin coated 96-well plates, capturing GalNAc-containing chimeras onto the immobilized WFA, removing unbound chimera, and quantitating the amount of luciferase activity that has been bound using coelenterazine.

α-RLuc and Trf-RLuc Were Transfected into CHO/Flp-In cells expressing a single copy of either βGT3 (Fig. 3A) or βGT4 (Fig. 3B), the amount of luciferase activity bound, expressed as LU bound per 50,000 LU of input, was 6-fold greater for α-RLuc than Trf-RLuc for both βGT3/CHO and βGT4/CHO cells.
Mutation of the glycosylation site (NVT to QVT) immediately adjacent to the PLRSKK sequence that we previously determined is essential for recognition using in vitro assays (26) completely abolishes GalNAc addition to α(QVT)-RLuc. In contrast, mutation of the glycosylation site more distant from the PLRSKK sequence (NHT to QHT) reduces but does not abolish GalNAc addition to α(QHT)-RLuc. Thus, α-RLuc is efficiently modified by both βGT3 and βGT4 in vivo, whereas Trf-RLuc is not efficiently modified by either βGT3. The properties of α-RLuc and Trf-RLuc agree with our in vitro studies showing the free α subunit contains a peptide sequence that is recognized by βGT and as a result is modified with a catalytic efficiency that is 500-fold greater than for glycoproteins that do not contain a recognition determinant for the βGT (23, 25–27). In addition, the oligosaccharide added to NVT appears to be the predominant target for GalNAc addition in the α-RLuc chimera.

The Ca6-RLuc chimera expressed by either HEK 293T cells, βGT3/CHO cells, or βGT4/CHO cells are not bound by immobilized WFA, indicating GalNAc is not added to this construct (not shown). Because the KRKKEK sequence that is essential for efficient GalNAc addition to Ca6(Wt) is located at the carboxyl terminus, we prepared chimeras in which the RLuc preceded the α and Ca6 sequences. RLuc-α, like α-RLuc, is secreted into the medium and is modified with GalNAc when expressed in either βGT3/CHO or βGT4/CHO cells, indicating the location of the RLuc is not critical for recognition of the α subunit by either βGT3 or βGT4 (compare Fig. 3, A and B, with Fig. 4, A and B). In contrast, RLuc-Ca6(Wt), unlike Ca6(Wt)-RLuc, is efficiently modified with GalNAc when it is expressed in either βGT3/CHO or βGT4/CHO cells (Fig. 4, A and B). Therefore, the presence of the MycHis epitope at the carboxyl

![FIGURE 4. Comparison of RLuc-α, RLuc-Ca6, and Trf-RLuc chimeric glycoprotein constructs as substrates for GalNAc addition in vivo.](Image)

![FIGURE 5. Comparison of GLuc-α, GLuc-Ca6, and GLuc-Trf chimeric glycoprotein constructs as substrates for GalNAc addition in vivo.](Image)

terminus of Ca6(Wt) does not interfere with recognition by either βGT3 or βGT4, whereas the presence of the much larger luciferase sequence at the carboxyl terminus of Ca6 is sufficient to prevent recognition. As was seen with Ca6(Wt), deletion of the KRKKEK sequence from RLuc-Ca6(Wt) to generate RLuc-Ca6(Mu1) markedly reduces GalNAc addition by βGT3 (Fig. 4A) and by βGT4 (Fig. 4B). Mutation of each individual Asn glycosylation site alone reduces but does not abolish GalNAc addition to either RLuc-Ca6(QLT) or RLuc-Ca6(QET). Mutation of both sites completely abolishes GalNAc addition to RLuc-Ca6(QLT/QET). The results obtained with RLuc-Ca6 indicate that βGT3 and βGT4 utilize the same recognition determinant in vivo and are able to transfer GalNAc to Asn-linked oligosaccharides at two different glycosylation sites on Ca6.

Recognition by βGT3 and βGT4 Can Be Conferred onto Transferrin by Adding Sequences from Ca6—GLuc is a naturally secreted form of luciferase that is significantly more stable than RLuc following secretion into the medium of cultured cells (34, 35). In each case >95% of the chimeric glycoprotein is secreted into the medium following transfection. Levels of expression were similar for each of these constructs with 1–10 × 10^6 light units of luciferase activity present per 5 μl of medium. The constructs shown in Fig. 1 were prepared and compared with RLuc chimeras (Fig. 5). As was seen with the RLuc chimeras, Gluc-α is modified with GalNAc when it is expressed in either βGT3/CHO or βGT4/CHO cells, whereas Gluc-Trf is not efficiently modified with GalNAc. Mutation of the key basic residues in α from PLRSKK to PLESEE (Fig. 1) reduces GalNAc addition to Gluc-α(PLESEE) by either βGT3 or βGT4 to the level seen for GLuc-Trf (Fig. 5, A and B). Adding the carboxyl-terminal 19 amino acids from Ca6 to Gluc-Trf yields a form of Trf, GLuc-Trf-Ca6(1–19) that is modified to a greater extent than Gluc-α with GalNAc by both βGT3 and
βGT4 (Fig. 5, A and B). The amount of GalNAc added to GLuc-Trf by βGT3 and βGT4 is increased 8-fold and 18-fold, respectively, in the presence of the carboxyl-terminal 19-amino acid sequence from CA6. The addition of only the carboxyl-terminal 9 amino acids that contain the KRKKEK sequence, GLuc-Trf-CA6(11–19), increased GalNAc addition to GLuc-Trf by βGT3 and βGT4 2.0- and 2.4-fold, respectively. Thus, the KRKKEK sequence does not have all the information required for recognition by either βGT3 or βGT4.

Because the first 10 amino acids within the 19-amino acid sequence added to GLuc-Trf-CA6(1–19) also contain basic residues, we prepared GLuc-Trf-CA6(1–10) that does not contain the final 9 amino acids including the KRKKEK sequence. Like GLuc-Trf-CA6(11–19), GLuc-Trf-CA6(1–10) is modified with GalNAc when expressed in βGT3 CHO and βGT4 CHO cells (Fig. 6, A and B) but not to the same extent as GLuc-Trf-CA6(1–19). Furthermore a construct containing the carboxyl-terminal 14 amino acids of CA6, GLuc-Trf-CA6(6–19) was also examined (Fig. 6, C and D). Like GLuc-Trf-CA6(1–10) and GLuc-Trf-CA6(11–19), GLuc-Trf-CA6(6–19) is modified to greater extent than GLuc-Trf but not to the same extent as GLuc-Trf-CA6(1–19) (Fig. 6, C and D). Thus, the full 19-amino acid sequence from the carboxyl terminus of CA6 is needed to for optimal modification of N-linked oligosaccharides on Trf chimeras by βGT3 and βGT4 in vivo.

**DISCUSSION**

Our studies using RLuc and GLuc glycoprotein chimeras establish that βGT3 and βGT4 are protein-selective glycosyltransferases in vivo. Both βGTs selectively transfer GalNAc to N-linked oligosaccharides on the pituitary glycoprotein hormone α subunit and on CA6 but not to the identical oligosaccharide acceptors on glycoproteins such as Trf. Adding the 19-amino acid sequence from the carboxyl terminus of CA6 to the carboxyl terminus of Trf converts GLuc-Trf into a glycoprotein that is selectively modified by βGT3 and by βGT4. This outcome is particularly remarkable because Trf has no structural relationship to either the α subunit or to CA6. The crystal structure of human serum transferrin has recently been solved (47). The two N-linked glycosylation sites at Asn⁶¹¹ and Asn⁶¹³ are located on adjacent loops of peptide and are both in close proximity to the carboxyl terminus of Trf. The addition of the recognition sequence to Trf may serve to mediate transfer of GalNAc to both N-linked oligosaccharides as it does in CA6. Because this 19-amino acid sequence is sufficient to confer recognition onto RLuc-Trf in vivo by both βGT3 and βGT4, it is likely that the same key residues serve to mediate recognition by both βGTs. Furthermore, because peptide recognition and GalNAc transfer to terminal GlcNAc represent distinct interactions, the addition of this sequence to virtually any glycoprotein will likely confer recognition by βGT3 and βGT4 and selective modification of accessible N-linked oligosaccharides with β1,4-linked GalNAc in vivo.

The N-linked oligosaccharide acceptor that is modified with β1,4-linked GalNAc on glycoproteins such as the glycoprotein hormone LH and CA6 is identical in structure to the N-linked oligosaccharide acceptors on glycoproteins that are not modified with GalNAc but are instead modified with β1,4-linked Gal. This led us to hypothesize the existence of a peptide recognition determinant located on glycoproteins that could be selectively modified with β1,4-linked GalNAc-containing structures. We confirmed this by comparing glycoproteins that do and do not contain a recognition determinant as acceptors for GalNAc addition in in vitro assays using solubilized enzymes. Now we have definitively demonstrated that in vivo the same recognition determinant is utilized by cells expressing βGTs endogenously and that the presence of the recognition determinant is necessary for efficient modification with β1,4-linked GalNAc.

Using in vitro analyses, we identified the basic amino acids in the sequence PLRSKK that is found in the α subunit as being essential for recognition by the βGT activity present in the pituitary. The lack of GalNAc transfer to Gluc-α(PLESEE) by either βGT3 or βGT4 now provides the in vivo demonstration that the basic amino acids in the PLRSKK sequence are indeed critical elements of the recognition determinant. Similarly, the marked reduction in GalNAc transfer by βGT3 and βGT4 to GLuc-CA6(Mu1) that has had the KRKKEK sequence deleted indicates that basic amino acids are critical elements of the recognition determinant in CA6. The results obtained with
Gluc-α(PLESEE) and Gluc-CA6(Mu1) show that the basic amino acids within these sequences are necessary for recognition in vivo.

The conversion of Trf, a glycoprotein that has no structural relationship to glycoproteins known to be modified with β1,4-linked GalNAc, into a substrate for GalNAc addition by βGT in vivo allows us for the first time to describe a sequence that is sufficient to mediate recognition. The addition of the carboxy-terminal 19 amino acids of CA6, LRRFIEQKITKRKKEKYWP, to the carboxyl terminus of Trf results in a glycoprotein Gluc-Trf-CA6(1–19) that becomes modified with GalNAc by both βGT3 and βGT4 to the same or a greater extent than Gluc-α in vivo. The 19-amino acid sequence is therefore sufficient to confer recognition by βGT3 and βGT4 in vivo. Gluc-Trf-CA6(1–10), Gluc-Trf-CA6(11–19), and Gluc-Trf-CA6(6–19) are also modified with β1,4-linked GalNAc in vivo but not to the same extent as Gluc-Trf-CA6(1–19). Thus, the first 10, the last 9, and even the last 14 amino acids of this 19-amino acid sequence are individually able to confer some recognition by βGT3 and βGT4 in vivo; however, the full 19-amino acid sequence is significantly more effective.

This 19-amino acid sequence is therefore sufficient for recognition in vivo by both βGT3 and βGT4, and the entire sequence is necessary to obtain maximal transfer of GalNAc in vivo.

The inability of CA6 (11–19) to confer as much transfer of GalNAc to Gluc-Trf as CA6 (1–19) indicates that a cluster of basic residues alone is not sufficient to mediate recognition. A number of different programs for the prediction of secondary structure including Chou-Fasman (44, 48), Garnier-Robson (49), Hierarchical Neural Network (50), and each of the secondary prediction algorithms available at www.compbio.dundee.ac.uk indicate that this sequence forms an α helix. Thus, it is possible that the requirement for the entire 19-amino acid sequence reflects the need to attain a specific secondary structure that places specific basic residues in the correct spatial relationship to interact with βGT3 and βGT4.

The ability to confer recognition by both βGT3 and βGT4 onto Gluc-Trf by adding the carboxyl-terminal 19 amino acids from CA6 indicates that the key elements of this sequence that are recognized by these closely related transferases are similar if not identical. The fact that a single recognition determinant can serve to mediate GalNAc addition to two distinct N-linked oligosaccharides on CA6 and can confer recognition onto an unrelated glycoprotein, i.e. Trf, indicates that the peptide recognition determinant does not interact directly with the oligosaccharide acceptor. Rather the peptide recognition determinant and the oligosaccharide acceptor interact with βGT3 and βGT4 independently.

βGT3 and βGT4 are large glycosyltransferases consisting of 987 and 1035 amino acids, respectively. The carboxyl-terminal regions of βGT3 and βGT4, consisting of 226 and 231 amino acids, respectively, are 68% identical and contain sequences that are characteristic of β1,4-glycosyltransferases (28, 30). This region is presumed to encode the actual catalytic activity; however, neither the region that mediates peptide recognition nor the function of the other regions of these transferases is known. Because the 19-amino acid sequence from the carboxyl terminus of CA6 is sufficient to mediate recognition in vivo and in vitro, it is now possible to devise strategies that will allow us to identify the key features of this determinant that are required for optimal recognition. It should also be possible to locate the regions of βGT3 and βGT4 that bind the recognition determinant.

The synthesis of glycoproteins bearing N-linked structures containing β1,4-linked GalNAc is a highly regulated process in vivo, requiring the expression of both the appropriate transferases and glycoproteins that have a recognition determinant. In the case of the glycoprotein hormone LH, the structure of the N-linked sugars plays a critical role in determining in vivo clearance rates and potency (14, 15). Structures with β-linked GalNAc are also present on other glycoproteins such as the low density lipoprotein receptor homolog SorLA/LR11 (20) that has been implicated as a risk factor for development of Alzheimer disease (51) and tenasin-R, an extracellular matrix component in the central nervous system (22). The highly regulated and selective addition of GalNAc generates unique structures that may play a number of different roles in vivo. Defining the key features of this 19-amino acid recognition determinant will allow us to identify additional glycoproteins that may also bear this unique modification. The addition of β1,4-linked GalNAc is among the best characterized forms of protein-selective carbohydrate modification. Understanding how the selective addition of GalNAc is determined will provide a useful paradigm for understanding how regulation of the synthesis of specific carbohydrate structures contributes to biologic function in vivo.

REFERENCES

1. Ghosh, P., Dahms, N. M., and Kornfeld, S. (2003) Nat. Rev. Mol. Cell Biol. 4, 202–212
2. Kornfeld, S. (1990) Biochem. Soc. Trans. 18, 367–374
3. Manzella, S. M., Hooper, L. V., and Baenziger, J. U. (1996) J. Biol. Chem. 271, 12117–12120
4. Hooper, L. V., Manzella, S. M., and Baenziger, J. U. (1996) FASEB J. 10, 1137–1146
5. Haltiwanger, R. S., and Lowe, J. B. (2004) Annu. Rev. Biochem. 73, 491–537
6. Lowe, J. B., and Martel, J. D. (2003) Annu. Rev. Biochem. 72, 643–691
7. Mi, Y.-L., Shapiro, S. D., and Baenziger, J. U. (2002) J. Biol. Chem. 277, 36052–36059
8. Mi, Y.-L., Shapiro, S. D., and Baenziger, J. U. (2002) J. Clin. Investig. 109, 269–276
9. Baenziger, J. U., Kumar, S., Brodbeck, R. M., Smith, P. L., and Beranek, M. C. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 334–338
10. Geurts, E. D., Van Halbeek, H., Boime, I., and Baenziger, J. U. (1985) J. Biol. Chem. 260, 15523–15530
11. M. A., Shapiro, S. D., and Baenziger, J. U. (2002) J. Biol. Chem. 277, 36052–36059
