A Systematic Study of Site-specific GalNAc-type O-Glycosylation Modulating Proprotein Convertase Processing

Katrine Ter-Borch Gram Schjoldager, Malene B. Vester-Christensen, Christoffer K. Goth, Thomas Nordahl Petersen, Søren Brunak, Eric P. Bennett, Steven B. Levery, and Henrik Clausen

From the Center for Glycomics, Department of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences and Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark and the Department of Systems Biology, Center for Biological Sequence Analysis (CBS), Technical University of Denmark, DK-2800 Lyngby, Denmark

Background: GalNAc-type O-glycosylation is emerging as a co-regulator of proprotein convertase processing of proteins.

Results: Site-specific O-glycosylation by 20 polypeptide GalNAc transferases have wide co-regulatory functions in proprotein processing.

Significance: This is the first systematic study that paves the way for wider co-regulatory functions of O-glycosylation in protein processing.

Site-specific GalNAc-type O-glycosylation is emerging as an important co-regulator of proprotein convertase (PC) processing of proteins. PC processing is crucial in regulating many fundamental biological pathways and O-glycans in or immediately adjacent to processing sites may affect recognition and function of PCs. Thus, we previously demonstrated that deficiency in site-specific O-glycosylation in a PC site of the fibroblast growth factor, FGF23, resulted in marked reduction in secretion of active unprocessed FGF23, which cause familial tumoral calcinosis and hyperostosis hyperphosphatemia. GalNAc-type O-glycosylation is found on serine and threonine amino acids and up to 20 distinct polypeptide GalNAc transferases catalyze the first addition of GalNAc to proteins making this step the most complex and differentially regulated steps in protein glycosylation. There is no reliable prediction model for O-glycosylation especially of isolated sites, but serine and to a lesser extent threonine residues are frequently found adjacent to PC processing sites. In the present study we used in vitro enzyme assays and ex vivo cell models to systematically address the boundaries of the region within site-specific O-glycosylation affect PC processing. The results demonstrate that O-glycans within at least ±3 residues of the RXR motif in furin cleavage site may affect PC processing suggesting that site-specific O-glycosylation is a major co-regulator of PC processing.
Site-specific O-Glycosylation Modulates PC Processing

cleavage sites is emerging as another important co-regulator of PC-mediated processing. Originally, tissue-specific processing of pro-opiomelanocortin was suggested to be related to O-glycosylation at Thr48 (8–10). More recently, we demonstrated that deficient O-glycosylation of the threonine residue in the RTHR279 PC-processing site of FGF23 is the cause of the rare human disease, familial tumoral calcinosis (11). Moreover, O-glycosylation immediately C-terminal to the processing site in ANGPTL3 (RAPR274–TT) was found to regulate processing and was possibly related to regulation of serum lipids (12). Other studies have also indicated that O-glycans in close proximity to PC-processing sites may affect cleavage (13, 14). These studies all point to a regulatory role of site-specific O-glycosylation by the large family of GalNAc transferases, which provides a co-regulatory system with a high degree of differential regulation through the many isoforms.

Here, we have systematically addressed the question of the spatial range of the influence of O-glycans on furin processing. We used a comprehensive set of in vitro peptide assays and an ex vivo cell model where O-glycosylation can be manipulated to demonstrate that O-glycans within at least ±3 residues (P7–P3’) of the RXXR PC-processing site can inhibit cleavage. Bioinformatics analysis of secreted proteins further demonstrated prevalence for Ser and Thr residues in or immediately adjacent to many putative PC-processing sites suggesting a much wider regulatory role of O-glycosylation in PC processing.

EXPERIMENTAL PROCEDURES

Chemoenzymatic Synthesis of Glycopeptide Substrates for in Vitro Analysis—A series of 20-mer synthetic peptides (Schäfer-N, Sigma) were designed with a central furin RXXR sequon processing site and a “walk” of potential O-glycosylation sites (Table 1, Fig. 1). Recombinant glycosyltransferases were expressed as soluble secreted truncated proteins in insect cells and purified as described previously (15). Selection of suitable GalNAc transferases for site-specific O-glycosylation of peptides were performed in product development assays in 25 μl of 25 mM cacodylic acid sodium, pH 7.4, 10 mM MnCl2, 0.25% Triton X-100, 1.5 mM UDP-GalNAc (Sigma), 10 μg of acceptor peptides, and 0.5 μg of purified enzyme at 37 °C using MALDI-TOF to monitor products. Glycopeptides for further analysis were produced in reactions with a 100–200-μg scale and purified by HPLC using a Kinetex™ 2.6-μm C18 100-Å LC column (100 × 2.1-mm C18 column).

Characterization of O-Glycosylation Sites by Electron Transfer Dissociation-MS2—Products of O-glycosylation reactions were characterized by electrospray ionization-linear ion trap-Fourier transform mass spectrometry (ESI-LIT-FT-MS) in an LTQ-Orbitrap XL hybrid spectrometer (Thermo Scientific) equipped for electron transfer dissociation for peptide sequence analysis by MS/MS (MS2) with retention of glycan site-specific fragments. Samples were dissolved in methanol/water (1:1) containing 1% formic acid and introduced by direct infusion via a TriVersa NanoMate ESI-Chip interface (Advion BioSystems) at a flow rate of 100 nl/min and 1.4 kV spray voltage. Mass spectra were acquired in positive ion FT mode using parameters similar to previous studies (12), except at a nominal resolving power of either 30,000 or 60,000. Electron transfer dissociation-MS2 spectra were analyzed by comparison with theoretical c and z’ fragment m/z values calculated for all positional combinations of one HexNAc residue distributed on all potential S and T glycosylation sites in the sequence. Calculations were performed using the web-based Protein Prospector MS-Product software routine.

EYFP Reporter Expression Constructs—A previously described reporter construct (12) was modified to include an additional His tag and an extended MUC1 tandem repeat sequence. The modified reporter construct was designed with a central 20-amino acid PC processing target sequence flanked by short spacer sequences and His tags/EYFPs in N-terminal end and 3.5 MUC1-TR/His tag in C-terminal end (Fig. 2A). A series of constructs with targeting sequences identical to those used in in vitro peptide assays (Fig. 1) were developed by in-frame insertion of double-stranded oligonucleotides, encoding the respective peptide sequences, into the SacII and Bpu1102I restriction sites situated within the respective N- or C-terminal spacer sequences of the reporter construct (Fig. 2A).

Expression of Reporter Constructs in CHO ldlD Cells—CHO ldlD cells with defective UDP-Gal/UDP-GalNAc C4-epimerase (16) were grown in Ham’s F-12/Dulbecco’s modified Eagle’s medium (1:1, v/v) supplemented with 3% fetal bovine serum and 1% glutamine. Cells were transfected with 2 μg of pcDNA3.1 reporter constructs using Mirus TransIT®-CHO Transfection Kit according to the manufacturer’s instructions. Transient expression was analyzed 48 h post-transfection, and stable transfectants were selected in 0.4 mg/ml of Zeocin (Invitrogen) by monitoring expression of EYFP. Further analysis by immunocytopathy using anti-EYFP (Living colors, Clontech), anti-HIS (Santa Cruz Biotechnology), or anti-MUC1 mAbs were performed. A secondary screening strategy monitoring the secreted reporter construct by a capture ELISA was developed. Briefly, 96-well microtiter plates were coated with 3 μg/ml of anti-MUC1 5E10 (15) overnight at 4 °C, and incubated in blocking buffer, pH 7.4 (PO4, Na/K, 1% Triton, 1% BSA), for 2 h at room temperature. Following the wash, wells where incubated with 100 μl of cell culture supernatant for 4 h at RT and subsequently 16 h at 4 °C. Binding of reporter constructs were detected by incubating wells with 1 μg/ml of biotinylated HMFG2 for 2.5 h at RT followed by streptavidin-HRP for 1 h at RT and visualized by incubating with TMB+ (Dako). Adding 0.5 m H2SO4 stopped the reaction and the plate was read at 450 nm. To evaluate the effects of O-glycosylation stable CHO ldlD transfecnt cells were grown in medium supplemented with 0.1 mM Gal (G), 1 mM GalNAc (Gn), or a combination (G/Gn).

In Vitro Furin Cleavage Assay of Peptides and Glycopeptides—A modification of an in vitro cleavage assay with a human furin protease (Sigma) was used to analyze furin cleavage of RXXR-Thr-GalNAc walk peptides, pro-BNP, ST6GalNAc-I, inhibin α, integrin α, and Semaphorin 3B peptides and glycopeptides (11). The assay was performed in 100 mM Hepes, pH 7.5, 1 mM CaCl2, 0.5% Triton X-100, 2 mM DTT using 10 μg of peptide or glycopeptide substrate and 1 unit of enzyme in a total volume of 50 μl. The mixture was incubated at 37 °C, and product development was evaluated after 0.5, 5, and 20 h by MALDI-TOF.
Ex Vivo PC Processing in CHO ldlD Cells—CHO ldlD cells stably expressing PC target sequence reporter constructs were grown to 50% confluence in 6-well dishes, and the medium was replaced with fresh medium containing 50, 5, 1, or 0 mM of the cell-permeable furin inhibitor Dec-RVKR-CMK (EMD Chemicals) in dimethyl sulfoxide. After 24 and 48 h of growth, cell culture supernatant was sampled and assayed by NuPAGE and Western blotting. Briefly, cell culture medium (10 μl) was separated on NuPage Novex BisTris 4–12% gels (Invitrogen) and blotted onto nitrocellulose membrane for 60 min, followed by incubation in 5% BSA in Tris-buffered saline (TBS). Following a wash in TBS plus 0.05% Tween 20, the membranes were incubated in primary antibody (mouse anti-HIS, rabbit anti-EYFP, mouse anti-Tn-MUC1 mAb 5E5 or anti-MUC1 5E10) overnight at 4°C (15). Blots were developed with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium substrate after incubation with secondary alkaline phosphatase-conjugated antibodies for 1 h at room temperature. Assessment of processing was performed by scanning dried membranes in a flat bed scanner (highest resolution 800 dpi) using ImageJ 1.45i software.

Bioinformatic Analysis—We used the human proteome from the International Protein Index (version 3.65) (17) downloaded via the European Bioinformatics Institute homepage with 86,379 sequences and first identified a predicted secretome using version 3.0 of the SignalP prediction algorithm (18). Further selection was performed by combining the NN (neural network) and HMM (hidden Markov model) versions of SignalP. Subsequently, we applied the ProP prediction algorithm (19) to identify putative RXXR PC processing sites combined with the presence of Ser/Thr amino acids within P6–P2’ positions of the RXXR motif.

RESULTS

In Vitro Analysis of Furin Cleavage of a Library of GalNAc-Glycopeptides—To systematically explore the spatial range within which site-specific O-glycosylation can affect PC processing, we designed a series of ideal peptides with a central RXXR furin processing motif and a walk of O-glycosylation sites (Thr residues) from P8 to P4 (Table 1). Because there is little knowledge of conformational requirements for efficient substrates for either furin processing or GalNAc-T O-glycosylation, we designed an arbitrary set of related peptide sequences with a high degree of sequence and amino acid similarity and with use of small, uncharged residues (Ala, Gly, Pro, and Thr, except for the necessary Arg residues). The use of similar sequence design was expected to result in similar rates and efficiencies of furin digestion. To ensure efficient enzymatic O-glycosylation Pro residues were included. We also avoided dibasic motifs in or around the RXXR motif to enhance selectivity for furin as other PCs have a preference for dibasic motifs. As predicted we were able to use three GalNAc-T isoforms to produce GalNAc glycopeptides with a single GalNAc residue attached at the appropriate Thr site in all peptide designs.

Two peptides (P3 and P1’) were poor acceptors for in vitro O-glycosylation and only ~50% incorporation of GalNAc was achieved even after prolonged assays (16 h), however, we could still use these for analysis of processing. A time course analysis of in vitro furin cleavage of each of the peptides revealed that all peptides were substrates for in vitro furin cleavage as predicted.

| Sequencea | GalNAc-transferab | In vitro furin cleavage/Protection by O-glycosylationd | Ex vivo furin cleavage/Protection by O-glycosylation |
|-----------|------------------|--------------------------------------------------------|---------------------------------------------------|
| P8        | 87654321         | T1 T2 T3                                               | Y/P                                               |
| P7        | -                | x                                                       | Y/Y                                               |
| P6        | x                | x                                                       | Y/Y                                               |
| P5        | -                | x                                                       | Y/Y                                               |
| P3        | -                | -                                                       | Y/Y                                               |
| P2        | -                | -                                                       | Y/Y                                               |
| P1’       | -                | (x)                                                     | Y/Y                                               |
| P2’       | AAGPAAARPR       | ATPARGAAGA                                             | Y/N                                               |
| P3’       | AAGPAAARPR       | -                                                       | NA                                                |
| P4’       | AAGPAAARPR       | -                                                       | Y/N                                               |

a 20-mer peptides containing a RXXR minimal furin cleavage motif and a walk of threonine from position P8 to P4’ (grey, underlined).
b In vitro glycosylation of target sequences with GalNAc-T1, -T2, and -T3; full incorporation of 1 site. (x); partial incorporation of 1 site.c Peptide sequence is cleaved in vitro by furin.
d In vitro glycosylation provides protection against in vitro proteolytic cleavage by furin. Y = yes, N = no, and P = partial. NA (not analyzed).e Ex vivo analysis of PC processing in the CHO ldlD cell model. Y = yes, N = no, and P = partial. NA (not analyzed).
and some variation in velocities of the reactions were observed with P2\textsubscript{H11032} being the poorest substrate (not shown). Comparative analysis of furin cleavage of peptides and GalNAc glycopeptides in a time course monitored by MALDI-TOF revealed that O-glycosylation within the region from P8 to P3\textsubscript{H11032} adversely affected cleavage (Table 1). Position P8 at the most N-terminal end tested partially affected cleavage, whereas the most C-terminal position P4\textsubscript{H11032} tested had no effect on cleavage (Table 1 and Fig. 1). O-Glycans at positions P7 (Fig. 1\textsubscript{B}), P5 (Fig. 1\textsubscript{C}), and P2\textsubscript{H11032} (Fig. 1\textsubscript{E}) (as well as P6, P3, and P1\textsubscript{H11032} not shown) completely blocked cleavage, although perhaps surprisingly an O-glycan at P2 (Fig. 1\textsubscript{D}) only partially inhibited cleavage. The latter is, however, in agreement with our previous studies of PC processing of FGF23, where GalNAc at P2 in the RHTR\textsubscript{179} processing site only partially inhibited cleavage \textit{in vitro} (11). Importantly, an extended O-glycan structure with α2,6-linked sialic acid provided complete block of processing. In our \textit{in vitro} analysis, we limited studies to the simple GalNAc O-glycan because our
**Site-specific O-Glycosylation Modulates PC Processing**

**FIGURE 2.** *Ex vivo* glycosylation and PC processing of reporter constructs in CHO ldlD cells. A, schematic depiction of the design of reporter constructs. The chimeric reporter proteins consisted of the PC target sequences with a Thr residue walk in positions P8–P4’ flanked by an N-terminal EYFP module for visualization of expression, a C-terminal MUC1 3.5 tandem repeat sequence for evaluation of O-glycosylation efficiency, and histidine tags for purification. The vertical arrow indicates sites of furin cleavage, and circles indicate sites of O-glycosylation. B, NuPAGE Western blot analysis with anti-EYFP of representative clones expressing P5, P2, P1’, P2’, or P4’ reporter constructs. Due to extensive O-glycosylation in the MUC1 tandem repeat region, the reporter construct migrated at 62 kDa compared with 49 kDa for the construct without O-glycosylation. The PC-processed N-terminal fragments with EYFP migrate at 37 (nonglycosylated) or 38 kDa (glycosylated). Migration of O-glycosylated, nonglycosylated, and processed proteins are indicated by arrows, and the addition of sugars added to growth medium are indicated as G (Gal), Gn (GalNAc), and G/Gn (Gal/GalNAc). C, quantification of band intensities in Western blots was performed by ImageJ version 1.45i software and depicted as relative band intensities. Ratio of processed/nonprocessed nonglycosylated reporter was set to 100 and the same ratios of the glycosylated constructs are shown as percentages hereof. The quantification is based on several replications of Western blots, P8 (n = 5), P5 (n = 4), P2 (n = 5), P1’ (n = 4), P2’ (n = 6), and P4’ (n = 4).

**Ex vivo** cell model provides information of the effect of extended O-glycans as described below. The results are summarized in Table 1.

**Ex Vivo Analysis of PC Processing in the CHO ldlD Cell Model—**

We next proceeded with analysis of PC processing of a selection of the peptide sequences tested by the *in vitro* assay (including the most N- and C-terminal sites and sites most proximal to the cleavage site) in the CHO ldlD cell system where capacity for O-glycosylation can be modulated by addition of exogenous Gal and GalNAc sugars (20). We used a modification of a previously described chimeric reporter construct that facilitates rapid selection of stable clones as well as screening of processing and analysis of the influence of O-glycosylation (Fig. 2A) (12). We initially analyzed glycosylation and processing of the reporter constructs by transient expression, but found that both glycosylation and processing of secreted products were quite incomplete so we used stable clones for all studies reported (Fig. 2B).

Processing efficiency of constructs was assessed by mobility shifts using the ratio of cleaved (N-terminal) and uncleaved fragments produced by cells grown without (37/50-kDa bands) and with the capacity for O-glycosylation (37–38/62-kDa bands). As shown in Table 1 and Fig. 2B the *in vitro* furin cleavage assay results correlated well with *ex vivo* processing of the corresponding reporter constructs in CHO cells. All reporter constructs were cleaved completely or partially without O-glycosylation. With O-glycosylation processing, constructs P5, P1’, and P2’ were virtually abrogated, whereas processing of the P8 and P2 constructs was partially abrogated in agreement with the *in vitro* assay result. In contrast, O-glycosylation of construct P4’ did not block processing. There were no substantial differences in processing of constructs glycosylated with simple GalNAccs (cells grown in GalNAc alone) or glycosylated with sialylated core 1 (NeuAcα2,3Galβ1–3[± NeuAcα2,6]-GalNAcc) O-glycans (cells grown in Gal/GalNAc) (Fig. 2B).

The reporter system allows assessment of O-glycosylation by large shifts in mobility (pro-apoprotein ∼50 kDa and pro-glycoprotein ∼62 kDa) due to the C-terminal MUC1 tandem repeat sequence with multiple O-glycosylation sites. O-Glycosylation of the PC processing target region can also be identified by mobility shifts of the cleaved N-terminal fragment (N-terminal apoprotein ∼37 kDa and glycoprotein 38 kDa) when glycosylation is at P1–P8, whereas determination of glycosylation in positions P1’–P4’ requires further analysis. The O-glycosylation capacity of CHO ldlD cells, i.e. the repertoire of GalNAc-Ts expressed, is not known. However, a recent study reported expression analysis of GalNAc-Ts in the parent CHO-K1 cell line (21), which is likely to be similar to CHO ldlD. Only GalNAc-T2, -T7, -T11, and -T19 were expressed in K1. In agreement with this, we previously showed that glycosylation sites specific for the GalNAc-T3/T6 subfamily, including, e.g. the
Site-specific O-Glycosylation Modulates PC Processing

FGF23 PC-processing site, are not glycosylated by the endogenous GalNAc-Ts expressed in both CHO-K1 and ldlD cells (11, 12). Although we could not directly confirm glycosylation of the P1′–P4′ constructs by mobility shifts, the finding that constructs P1′–P3′ exhibited complete or partial blocking of processing is in accord with efficient glycosylation of the target sequences.

PC Processing of Reporter Constructs Is Performed by a Furin-like PC—Most PCs prefer cleavage at dibasic sequences, whereas furin exhibits rather strict specificity for RX(R/K)R but accepts RXXR. We therefore, as mentioned earlier, designed peptides and reporter constructs with cleavage sequence RXXR, with X = Pro, Ala, and Thr, only to ensure the absence of a dibasic motif, and thus excluding activity of other PCs that may be expressed in CHO cells. To further confirm that furin or furin-like activities were responsible for the observed processing we tested the effect of addition of the cell permeable furin-like protease inhibitor Dec-RVKR-CMK on processing in CHO ldlD cells. Cleavage of all tested reporter walker constructs (P8, P2, P2′, and P4′) was inhibited in a dose-dependent manner with almost complete inhibition at 5 μM (Fig. 3). These results indicate that the PC target sequences are processed by furin-like convertases in CHO ldlD. The convertases PACE4 and PC5/6, with similar substrate specificities as furin, are also inhibited by Dec-RVKR-CMK, but PACE4 is not believed to be expressed in CHO-K1 (22, 23) and presumably not in the subclone CHO ldlD.

Proteome-wide Analysis of Potential O-Glycosylation Around Potential RXXR Processing Sites—To probe the potential for co-regulation of PC processing by site-specific O-glycosylation we surveyed potential O-glycosylation sites in RXXR furin-like processing sites by searching the human proteins with signal sequences for occurrences of Ser/Thr residues within the P8–P5′ RXXR sequons (Fig. 4A). A high prevalence of Ser residues at P7–P5, P3, and P1–2′ were found, as well as less prevalent Pro at P6 and P2, and Thr even less prevalent at P6–5, 3, and P1–2′. Although there is no good prediction algorithm for single O-glycosylation sites these findings suggest that O-glycosylation around PC processing sites may be rather common. The analysis allowing Ser/Thr at P6–P2′ revealed 710 proteins with 1,264 RXXR sequons having Ser/Thr within P6–P2′, and further narrowing down the allowed positions for Ser/Thr residues reduced the number as shown in Fig. 4B. Analysis of the list of identified proteins reveals a wide range of all major protein classes known to undergo PC processing (Fig. 4C).

In Vitro Analysis of a Select Set of Identified Proteins—From the bioinformatics analysis we selected 11 functionally diverse proteins for in vitro analysis of O-glycosylation and furin processing (Table 2). FGF23 and ANGPTL3 identified previously (11, 12) were included as controls. Synthetic peptides were tested for in vitro O-glycosylation with the three most abundant GalNAc-T isoforms, and all but one (hepatocyte growth factor receptor) could be O-glycosylated by one or more of these isoforms indicating that the sequences may be O-glycosylated in vivo. Only four of the 11 proteins are known to be O-glycoproteins and in most cases the sites of O-glycans are unknown. In vitro furin cleavage assays demonstrated that 9 of the 11 sequences could be cleaved, whereas 8 of the 11 proteins are known to undergo PC processing. We observed, as expected, substantial differences in velocity and efficiency of cleavage of the different peptide sequences, which are exemplified by the time course analysis of the semaphorin 3B, integrin αE, and pro-BNP peptides (Fig. 5 and supplemental Fig. S1). Thus, the integrin αE sequence was completely cleaved at the first 3-h time point, where several other peptides were cleaved less than 50%.

We next tested in vitro cleavage of the same sequences with GalNAc incorporated by the most active GalNAc-T isoforms, and demonstrated that in all but one case (Endothelin-1), where O-glycosylation and furin cleavage were possible, O-glycosylation completely or partially protected in vitro furin cleavage as predicted (Table 2). Fig. 5A shows a time course furin digestion of fully O-glycosylated semaphorin 3B peptide where the GalNAc glycopeptide is completely protected from digestion. Fig. 5B shows an example of an integrin αE (integrin αE/CD103) peptide (mixture of 2 and 3 GalNAc residues), where the GalNAc glycopeptide with 3 GalNAc residues (Thr56, Thr60, Thr63) is completely protected from digestion, whereas the glycopeptide with 2 sites (Thr56, Thr63) and the unglycosylated peptide is cleaved. It is surprising that the glycoform with two GalNAc residues at positions P7 and P1′ was readily cleaved by furin. We have not identified a particular reason why the integrin αE peptide behaves differently than the other examples tested as well as the model peptides. The peptide was one of the best furin substrates tested, and the example may serve to illustrate that local conformation in addition to sequence and O-glycosylation also influence cleavage efficiency. An example of partial protection (pro-BNP) is shown in Fig. 6A, and additional examples (ST6GalNAc-I and Inhibin α) of complete protection are shown in Fig. 6, B and C.

DISCUSSION

Site-specific GalNAc-type O-glycosylation is emerging as an important co-regulator of PC processing of proteins, but our knowledge of and ability to predict the occurrence of this event is highly limited. In this study we first used a model system to delineate the sequence range within which site-specific O-glycosylation may affect PC processing in vitro and ex vivo in cells,
and demonstrate that O-glycans at least ±3 residues N- and C-terminal and of the RXXR furin recognition motif may affect processing. This allowed proteomewide analysis of sequences with the potential for PC processing and adjacent O-glycosylation, and the search yielded hundreds of candidate proteins despite that it was limited to the RXXR motif. It is estimated that ~3,500 proteins can potentially undergo PC processing (1, 7), and our results suggest that a major part of these may be co-regulated by O-glycosylation. The select group of proteins identified and analyzed in further detail supports this, where our predictions of PC processing, O-glycosylation, and modulation of processing by furin were largely correct.

The finding that site-specific O-glycosylation may regulate hundreds of PC processing events places the large GalNAc-T gene family in a central position in a wide area of biological pathways. Elucidating specific functions of each of the GalNAc-T isoforms in health and disease is not a straightforward task due to lack of reliable prediction methods. A number of studies including genomewide association studies link specific GalNAc-T isoforms with important biological functions and diseases, however, in most cases the molecular basis for the role of GalNAc-Ts in these remained obscure (24). The first and only specific mechanism uncovered so far was indeed dysregulated PC processing of the phosphaturic factor FGF23 caused by deficiency in the GalNAc-T3 isoform, which exclusively regulates O-glycosylation in the RHTR processing sequon of FGF23 (11). Perhaps surprisingly, the rare disease familial tumoral calcinosis may be caused by deleterious mutations in either the GALNT3 or FGF23 genes, which suggests that despite the broad function of the GalNAc-T3 enzyme as shown in Tables 1 and 2, this enzyme has limited essential nonredundant functions in O-glycosylation apart from the single Thr178 site in FGF23. We envision that other disease-causing defects in the GalNAc-T repertoire as predicted by genome-wide association studies will involve dysregulated PC-processing events as we recently proposed for GalNAc-T2 (12).

Our current knowledge of O-glycosylation of especially less abundant proteins and proteins with single O-glycosylation sites is quite limited, and sequence-based prediction models generally fail to predict isolated O-glycosylation sites in proteins (25). However, considerable data indicate that in vitro GalNAc-T enzyme analysis of short peptide substrates reasonably reflects glycosylation occurring in cells with the appropriate GalNAc-T repertoire as predicted by genome-wide association studies will involve dysregulated PC-processing events as we recently proposed for GalNAc-T2 (12).

![FIGURE 4. Bioinformatic search for RXXR furin processing sites with adjacent O-glycosylation sites. A, a sequence logo plot was generated from all known secreted mammalian proteins containing the RXXR motif. The upper panel shows the distribution of amino acids in positions P8 to P4 in relationship to RXXR and the lower panel shows an enlargement of the lower area to bring out the top 5 amino acids in the above mentioned positions. B, the bioinformatics search for RXXR revealed 1264 potential furin processing sites in 710 proteins where a Ser or Thr is found within position P6 to P2'. In P5 to P1' we found 896 sites in 573 proteins. If a search was made for the optimal furin cleavage site RXXR Ser or Thr in P6 to P2' were seen in 641 sites in 384 proteins and the corresponding number for P5 to P1' were 396 and 285, respectively. If the search was narrowed down to contain Ser or Thr in P2 or P3, 252 sites in 229 proteins were displayed. C, a schematic illustration of the many protein classes represented in our search result that can be affected by up to 20 different GalNAc-Ts and the 7 proprotein convertases.](image-url)
and identification of O-glycosites. With this strategy we have identified 400 new O-glycosites demonstrating that GalNAc O-glycosylation is far more prevalent than previously thought (27). Furthermore, we have tested 12 peptides derived from this search in in vitro O-glycosylation assays and found excellent correlation between in vitro enzyme assays and in vivo O-glycosylation (27). We hope that further characterization of the O-glycoproteome by this strategy will advance knowledge of the co-regulation of PC processing events by site-specific O-glycosylation.

In this study we further analyzed a set of protein examples identified by our proteome-wide search (Table 2). A known processing site in the pro-BNP (TLRAPR76SP) was susceptible to furin cleavage as well as O-glycosylation in our in vitro assays (Fig. 6A). Pro-BNP is produced by cardiomyocytes in the atrium and ventricle of the heart, and is proteolytically processed at Arg76 to yield NT-pro-BNP (N-terminal fragment of BNP precursor). GalNAc-T3 glycosylated Thr71 (P6) and Ser77 (P1/H), providing partial protection from furin cleavage in 16-h assays. This is in agreement with studies demonstrating that pro-BNP is O-glycosylated, and that Thr71 may be important for processing (13). The same authors also found evidence that pro-BNP is differentially glycosylated and processed in two different cell lines (HEK293 and CHO). Interestingly, we found that within the BNP processing site T71LRAPR S77, GalNAc-T3 initiates glycosylation at Thr71 and proceeds with glycosylation of Ser77 using its lectin domain. Thus, a GalNAc-T3 lectin mutant only glycosylated Thr71. This is similar to how GalNAc-T3 glycosylates Thr171 of FGF23 (TPIPRHTR179) before glycosylating Thr178 in a lectin-dependent manner (not shown). Glycosylation with a GalNAc-T3 lectin mutant revealed that GalNAc alone at P6 yielded protection in a 5-h cleavage assay (supplemental Fig. S1). Pro-BNP is up-regulated during pathophysiological stress and processed BNP is a prognostic biomarker in heart failure. It is possible that dysregulation of GalNAc-Ts plays a role for the altered processing of BNP in disease.

Recently, it was shown that activin stimulates maturation of inhibin/B and inhibin/\(\alpha\) through induction of furin or a furin-like PC, and thus acting in a positive feedback loop regulating FSH production (28, 29). We found that glycosylation of Thr234 (P2) in the processing site (RARR232ST) of inhibin/B interferes with in vitro furin cleavage (Fig. 6B). Furthermore, glycosylation was specifically performed by the GalNAc-T3 isoform, which, as inhibin/B, is highly expressed in the testis (30, 31).

Fibroblast growth factor 7/keratinocyte growth factor (FGF7/KGF) is secreted from mammalian cells as two biologically active forms, KGF1 (amino acids 1–163) and KGF2 (amino acids 23–163) due to PC processing in the RHTR23 sequon. The longer form is O-glycosylated at Thr22 (P2) and N-glycosylated at Asn14 (32). We found that O-glycosylation of Thr22 blocked processing suggesting that site-specific O-glycosylation in this case may co-regulate the ratio of short/long forms of KGF secreted (Table 2). We previously found that GalNAc O-glycosylation in the same RHTR179 motif in FGF23 only partially affected processing, whereas elongated

| Peptide | Accession no. | Sequence | GalNAc-transferase | In vitro cleavage | Protection by O-glycosylation | Known O-glycosylation | Known PC-processing |
|---------|---------------|----------|-------------------|-------------------|-------------------------------|-----------------------|--------------------|
| Pro-BNP | P16860        | T1111111 | -11111           | -11111           | Y/Y                           | Y(13,46)              | Y(13)              |
| Inhibin α | P05111       | T1111111 | -11111           | -11111           | Y/Y                           | Y(32)                 | Y(32)              |
| FGF7    | P2178        | T1111111 | 11111            | 11111            | Y/Y                           | Y(46)                 | Y(44)              |
| BMP7    | P18075       | T1111111 | 11111            | 11111            | N/P                           | N                    | N                  |
| Dystroglycan | Q14118     | 1111111 | -11111           | -11111           | N/NA                          | Y(46)                 | Y(44)              |
| Collagen XVIII | F93006   | 1111111 | 11111            | 11111            | N/NA                          | Y(45)                 | Y(45)              |
| HGF     | P08581       | 1111111 | -11111           | -11111           | Y/NA                          | N                    | Y(47)              |
| Integrin αE | F38570    | 1111111 | -11111           | -11111           | Y/NA                          | N                    | Y(39)              |
| ST6GalNAc1 | P05505    | 1111111 | 11111            | 11111            | N/NA                          | N                    | N                  |
| Control substrates |           |           |                  |                  |                               |                      |                    |
| FGF23   | Q9GZ9V       | 1111111 | -11111           | -11111           | Y/Y                           | Y(11)                 | Y(49)              |
| ANGPTL3 | Q9Y3C1       | 1111111 | -11111           | -11111           | Y/Y                           | Y(12)                 | Y(50)              |

TABLE 2 Summary of in vitro analysis of O-glycosylation and PC furin cleavage of selected protein examples

a Peptide sequences used containing cleavage site (down arrow) and potential sites of O-glycosylation (grey). Sites determined by Orbitrap ESI-ETD-FT-MS² are underlined and grey.
b In vitro glycosylation of target sequences with GalNAc transferase-T1, -T2, and -T3. No incorporation (--) and 1, 2, or 3 mol of GalNAc incorporated.
c In vitro glycosylation provides protection against in vitro proteolytic cleavage by furin, yes (Y), no (N) or not applicable (NA).
d Target protein is known to be O-glycosylated, reference.
e Target protein is known to undergo SPC processing, reference.
f Integrin αE is reported to be processed in RQR232, but two other potential cleavage sites exists.
g Two GalNAc glycoforms with 2 and 3 mol of GalNAc incorporated (third GalNAc on Thr60) were analysed in a mixture, and only the glycoform with 3 mol was protected from cleavage.
glycan structures were required for complete blocking of furin cleavage (11). This discrepancy may be due to the finding that the FGF7 peptide generally was a poorer substrate for in vitro furin cleavage compared with the FGF23 peptide.

Semaphorin 3B (Sema3B) (RNRR731 \textsuperscript{2}TH) was identified in our search, and we found that GalNAc-T3-glycosylated Thr\textsubscript{732} (P1/H11032), which inhibited in vitro processing by furin (Fig. 5A).

Sema3B was first identified together with semaphorin 3F as a tumor suppressor of small-cell lung carcinoma (33, 34), and later Sema3B was found to function as an endothelial cell repellent (35). Sema3B is inactivated by furin cleavage at multiple processing sites. The major furin site generates 20- and 80-kDa fragments, and Varshavsky and colleagues (35) show that processing at the most C-terminal site that generates /H11011-2- and 78-kDa fragments is sufficient to disrupt its function as an inhibitor of angiogenesis. A role of O-glycosylation in this process is intriguing.

Integrins are transmembrane cell surface \(\alpha/\beta\) heterodimeric receptors of which two, platelet glycoprotein IIb (36) and integrin \(\beta1\) (37), are known to be O-glycosylated and several \(\alpha\)-integrin subunits undergo PC processing by furin in the extracellular domain yielding disulfide-linked heavy and light chains (38). Integrin \(\alpha\varepsilon\) (CD103) heterodimerizes with integrin \(\beta7\) (\(\alpha\varepsilon\beta7\)) and constitutes the E-cadherin binding integrin, also known as mucosal lymphocyte-1 antigen, primarily expressed in intestinal intraepithelial lymphocytes. Integrin \(\alpha\varepsilon\) is reported to be processed in RQR\textsuperscript{178} (39), but contains two additional predicted SPC sites. One of these (RTKR\textsuperscript{62}TP) is located in the N-terminal region, and this was susceptible to furin cleavage as well as O-glycosylation in our in vitro assay. GalNAc-T3 incorporated 2 and 3 GalNAc residues at Thr\textsubscript{56} (P7), Thr\textsubscript{60} (P3), and Thr\textsubscript{63} (P1) where Thr\textsubscript{60} in the RTKR\textsuperscript{62} motif was only partially glycosylated (Table 2). Surprisingly, GalNAc glycosylation in the P7 and P1' positions did not protect against furin cleavage, whereas the additional glycosylation in P3 yielded complete protection (Fig. 5B). GalNAc glycosylation in positions P7 and P1' blocked processing of our model peptide sequences (Table 1) as did glycosylation of P1' in the

![FIGURE 5. MALDI-TOF time course analysis of in vitro furin cleavage of select examples. A, the semaphorin 3B peptide was readily cleaved in the RNRR\textsuperscript{731} motif, yielding N-terminal fragments of 1314 after 3 h (left panel), whereas the glycopeptide with one GalNAc residue at P1' was completely protected against cleavage even after 16 h incubation (right panel). B, the integrin \(\alpha\varepsilon\) peptide was completely cleaved in the RTKR\textsuperscript{62} motif after 3 h, yielding an N-terminal fragment of 1454 Da (left panel). The corresponding glycopeptide with two GalNAc residues at P7 and P1' was also cleaved after 3 h, whereas the glycopeptide with 3 GalNAc residues at P7, P3, and P1' was not cleaved after a 16-h incubation (right panel). Reactions were sampled at 0, 3, and 16 h.]

The full text includes additional details and figures not shown here, focusing on the role of site-specific O-glycosylation in modulating PC processing and the specific examples of proteins and peptides discussed.
semaphorin 3B sequence (Fig. 5A). There is no obvious explanation for this discrepancy, but local conformation and efficiency of substrate sequence may play roles and these are not considered in this study.

Finally, we identified a putative processing site in the stem region of the type 1 transmembrane Golgi-located sialyltransferase ST6GalNAc-I (PTRARRP). Many Golgi glycosyltransferases are known to be shed as catalytically active enzymes by cleavage in the stem region, but little is known about the mechanism and only two examples have been characterized to our knowledge. Thus, shedding a soluble active form of ST6Gal-I has been shown to be performed by BACE-1 (40). PC processing has been shown to regulate secretion of the β3GlCNAC transferase Lunatic fringe (41). We found that Thr74 (P2) was O-glycosylated in vitro and that this efficiently protected furin cleavage (Fig. 6C).

ST6GalNAc-I is a key regulator of the predominant cancer glycoform STn (NeuAc2,6GalNAc1-O-Ser/Thr) (42) and overexpression in cells override the normal elongation processes of O-glycosylation (37, 43). ST6GalNAc-I has not been reported to be O-glycosylated, but we have recently identified O-glycosylation sites in the stem regions of several Golgi glycosyltransferases with our SimpleCell strategy, including in ST6GalNAc-I, although not at the Thr74 site (27). This finding may help advance our understanding of the regulation of the resident time of glycosyltransferases in cells.

The dystroglycan and collagen XVIIIa1 peptides were both readily in vitro glycosylated but not cleaved by furin. Processing of both substrates has been described and it is plausible that PCs other than furin are responsible for potential processing of these sites (44, 45). Bone morphogenic protein 7 and endothelin-1 were both glycosylated and cleaved by furin in vitro but glycosylation did not yield protection against processing. The endothelin-1 peptide RLRR was not protected by O-glycosylation with a GalNAc residue at P8 (or P10 as specific site in N-terminal fragment was not confirmed).

In summary, our study predicts that site-specific O-glycosylation adjacent to PC processing sites is a major player in regulation of processing of proteins. We have only surveyed the simple RXXR motif, but further analysis of simpler dibasic motifs targeted by most PCs is expected to provide even further support for this prediction. It has long been known that O-glycans provide protection from proteolytic digestion with the most illuminating example being the LDL receptor (16, 20), but the understanding that O-glycosylation at specific sites play regulatory roles counteracting fine-tuned PC processing events that activates or inactivates bioactive proteins has only emerged recently (11, 12). The strategy and results of the survey conducted in this study provide a foundation for further studies into the co-regulatory role of site-specific O-glycosylation.

Acknowledgments—We kindly acknowledge Dr. Monte Krieger for generously sharing the CHO ldlD cell line and Dr. Nabil Seidah for helpful suggestions and review of the manuscript.

REFERENCES
1. Seidah, N. G. (2011) Ann. N.Y. Acad. Sci. 1220, 149–161
2. Seidah, N. G., Mayer, G., Zaid, A., Rousselet, E., Nassoury, N., Poirier, S.,
