Stony Brook University

Mapping *O*-Fucosylation and *O*-Glucosylation on Human Notch3 Epidermal Growth Factor-Like (EGF) Repeats Using Nano-LC-ESI-MS/MS

A Thesis Presented

by

Ga-Ram Hwang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

December 2012

Stony Brook University

The Graduate School

Ga-Ram Hwang

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Dr. Robert S. Haltiwanger – Thesis Advisor
Professor and Chair, Department of Biochemistry and Cell Biology

Dr. Deborah Brown – Second Reader
Professor, Department of Biochemistry and Cell Biology

This thesis is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Thesis

Mapping *O*-Fucosylation and *O*-Glucosylation on Human Notch3 Epidermal Growth Factor-Like (EGF) Repeats Using Nano-LC-ESI-MS/MS

by

Ga-Ram Hwang

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

2012

The Notch protein is a transmembrane signaling protein responsible for regulating several important pathways among all metazoans including cell proliferation, differentiation, and death. Notch exists as one protein in *Drosophila*, but has four homologs in mice and humans (Notch1-4). A defining component of the Notch protein is the presence of 29-36 tandem epidermal growth factor-like (EGF) repeats, small protein motifs defined by the presence of six cysteines forming three disulfide bonds. Defects in Notch have been linked to various diseases like Alagille syndrome and cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). CADASIL is responsible for early onset of dementia in patients aged 40-50, along with migraines and stroke. CADASIL is characterized by the presence and accumulation of granular osmiophilic material (GOMs) and Notch3 extracellular domain in close proximity to vascular smooth muscle cells, eventually leading to the degradation of vascular smooth muscle cells. Previous evidence showed that mutations in Notch3 are related to

the pathogenesis of CADASIL, mainly in EGF repeats 1-5 and 10-11. A majority of these mutations have been identified as missense mutations leading to the addition or deletion of cysteine residues, which is vital component of *O*-linked glycosylation in Notch. The two main types of *O*-linked glycosylation on Notch are *O*-fucosylation initiated at the consensus sequence $C^2XXXX(\underline{S}/\underline{T})C^3$ by Pofut1, and *O*-glucosylation initiated at the consensus sequence $C^1X\underline{S}X(P/A)C^2$ by Poglut1/Rumi, where C^{1-3} are three of the 6 conserved cysteines within the EGF repeat. After *O*-fucosylation, the *O*-fucose can be elongated by Fringe, which adds a GlcNAc to the fucose, followed by subsequent additions by other enzymes, resulting in the tetrasaccharide Siaα2-3Galβ1-4GlcNAcβ1-3Fucα1-O-Ser/Thr. Similarly, *O*-glucose can be elongated by Gxylt1 or Gxylt2, followed by Xxylt1, resulting in the trisaccharide Xylα1–3Xylα1–3Glcβ1-O-Ser. *O*-linked glycosylation is known to be essential for the function of Notch. In addition, previous work showed that CADASIL mutants in mouse Notch3 EGF repeats 1-5 did not affect addition of *O*-fucose to Notch EGF repeats, but elongation of *O*-fucose by Fringe was reduced. These results suggest CADASIL mutations may influence glycosylation of mouse Notch3. My project has been to analyze *O*-fucose and *O*-glucose modifications on wild type human Notch3 as a precursor to analyzing whether CADASIL mutations in the human protein alter their structure. Using mass spectral methods I have identified *O*-fucose and *O*-glucose glycans on the majority of predicted sites, and *O*-glucose was elongated to its trisaccharide form at most sites. However, glycosylation at several “non-canonical” sites were also identified. More work needs to be done to identify the remaining glycosylation sites on Notch3, and then the effects of CADASIL mutations on the efficiency of modification at individual sites can be addressed.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BSA	Bovine Serum Albumin
CADASIL	Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
CNL	Constant Neutral Loss
Cys	Cysteine
Di-AP	Di-Ammonium Phosphate
ECD	Extracellular Domain
EGF	Epidermal Growth Factor-like
EIC	Extracted Ion Chromatogram
Fuc	Fucose
Gal	Galactose
Glc	Glucose
GlcNAc	*N*-acetyl-D-glucosamine
GOMs	Granular Osmiophilic Deposits
Gxylt1	Glucoside Xylosyltransferase 1
Gxylt2	Glucoside Xylosyltransferase 2
HA	Human Influenza Hemagglutinin
His	Histidine
hN3	Human Notch3
HPLC	High-Performance Liquid Chromatography
ICD	Intracellular Domain

mN1	Mouse Notch1
mN3	Mouse Notch3
MS	Multiple Sclerosis
NANA	Sialic Acid
Nano-LC-ESI-MS/MS	Nano-Liquid Chromatography-Electrospray Ionization-Mass Spectrometry/Mass Spectrometry
NRR	Negative Regulatory Region
Pofut1	Protein *O*-fucosyltransferase 1
Poglut1	Protein *O*-glucosyltransferase
SB	Sample Buffer
SCD	Spondylocostal Dystosis
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Ser	Serine
T-ALL	T Cell Acute Lymphoblastic Leukemia
TCEP	*Tris*(2-carboxyethyl)phosphine
Thr	Threonine
Tris-HCl	Tris(hydroxymethyl)aminomethane-Hydrochloride
VSMCs	Vascular Smooth Muscle Cells
WT	Wild-type
Xyl	Xylose
Xxylt1	Xyloside Xylosyltransferase 1

Acknowledgments

I would like to thank Dr. Jan Manent and Dr. Spyros Artavanis-Tsakonas of the Artavanis-Tsakonas lab for graciously providing the human Notch3 ECD samples used for this lab. Next, I want to thank the members of the Haltiwanger lab for all of the assistance and support that they've given me over this past year. I want to thank Dr. Deborah Brown for having time to read this thesis. I want to also thank Dr. Neta Dean for all of the help that she gave me in our Master's program. Next, I want to thank my mentor, Esam Al-Shareffi, for all of the guidance and help that he gave during my time in the Halitwanger lab. Also, I would like to thank Dr. Robert Haltiwanger, for it was through his lab and guidance that this thesis was possible. Finally, I want to thank my family for always being there to support me.

Chapter 1: Introduction

The Notch protein is a transmembrane signaling receptor protein responsible for regulating several important pathways in all metazoans (1). It is a single-pass Type 1 transmembrane protein, consisting of a large extracellular domain (ECD), followed by a transmembrane region and an intracellular domain (ICD) responsible for downstream signaling. The ECD contains 29-36 tandem epidermal growth factor-like (EGF) repeats and a negative regulatory region (NRR). Notch EGF repeats are around 40 amino acids long containing exactly six cysteines per EGF repeat. As shown in Figure 1, the six cysteines in an EGF repeat form three intra-chain disulfide bonds as follows: Cys1 to Cys3, Cys2 to Cys4, and Cys5 to Cys6 (2). Although only one Notch protein exists in *Drosophila*, four homologs of Notch exist in mammals (Notch1-4) (2, 3). Two of Notch's EGF repeats (EGF 11-12 for Notch1 or 2, EGF 10-11 of Notch3) have been shown to be necessary and sufficient for ligand binding (4). Notch interacts with the transmembrane ligands of neighboring cells such as Delta (known as Delta-like in humans) and Serrate (known as Jagged in humans). Through Notch's interaction with these ligands, Notch is cleaved and its ICD is released, which then travels to the nucleus to regulate the transcription of its downstream targets. Some of the pathways regulated by the Notch pathway are cell proliferation, cell fate, differentiation, and cell death (1).

Mutations of Notch or its downstream pathway components have been implicated in numerous diseases. In humans, T cell acute lymphoblastic leukemia (T-ALL) results from mutations in Notch1 (4). Mutations in Notch2 have been implicated in causing Alagille syndrome, a developmental disorder with defects in heart and liver formation, while mutations in Notch3 cause CADASIL (cerebral autosomal dominant arteriorpathy with subcortical infarcts

and leukoencephalopathy) (4). Other diseases linked to defects in the Notch signaling pathway include Spondylocostal Dystosis (SCD) and Multiple Sclerosis (MS) (4). In addition, although not a disease, elimination of Notch1 or Notch2 in mice results in embryonic lethality, demonstrating the importance of Notch in mammalian development (2). However, for this thesis, the focus will be placed on CADASIL.

CADASIL is a hereditary cerebral small vessel disease caused by dominant mutations in the NOTCH3 gene (5). Clinical features include migraines, transient ischemic attacks or strokes, psychiatric disorders, and cognitive impairment with an onset among patients around the age of 40 – 50. An important characteristic of CADASIL is the presence of granular osmiophilic deposits (GOMs) within the basal membrane of dermal media, peripheral nerves, and muscle arteries (5) along with the progressive loss of vascular smooth muscle cells (VSMCs). However, the presence and number of GOMs appears to not have correlation to the severity of damage to VSMCs. Furthermore, the relationship between GOMs and VSMCs is still largely unknown, specifically, how GOMs cause changes to the vessel wall and damage VSMCs (6). In addition to the presence of GOMs on the VSMCs of CADASIL patients, an aggregation of Notch3 ECD multimers have also been detected in close proximity of GOMs, although the exact relation between GOMs and Notch3 ECD is still unknown. Although not much is known about the composition of GOMs, they do not contain amyloid, elastin, chromatin, calcium or iron (6). As a result of the accumulation of Notch3 ECD, degeneration of vascular smooth muscle cells occurs. Therefore, pathogenesis of CADASIL is characterized by the accumulation of GOMs and Notch3 ECD around vascular smooth muscle cells, eventually leading to their degradation (6).

An interesting characteristic of mutations of NOTCH3 that lead to CADASIL is that the vast majority are missense mutations located within the EGF repeat region of Notch3's ECD,

resulting in either the addition or loss of cysteine residues. Specifically, mutations to EGF repeats 1-5 and 10-11 have been found to be major hotspots leading to CADASIL (7). Since the three intrachain disulfide bonds per EGF repeat are vital to the proper folding of EGF repeats, an addition or loss of cysteine residues, resulting in an odd number of cysteine residues in the EGF repeat, would lead to improper folding of EGF repeats (3). Also, by creating an odd number of cysteines residues, a cysteine residue would be left unpaired, and this could lead to the aggregation of Notch3 ECD through pairing among each other. Furthermore, an important characteristic of Notch's EGF repeat region is that it has been observed to be glycosylated, specifically through *O*-linked glycosylation, with the cysteines playing a vital role in determining its glycosylation fate (8). Therefore, there may be a link between CADASIL mutations and the glycosylation of Notch3.

Glycosylation is one of the more common types of co/post-translational modifications found on various proteins. The two major types of glycosylation are *N*-linked glycosylation and *O*-linked glycosylation (9, CC). *N*-linked glycosylation operates through the linking of carbohydrates (primarily *N*-acetylglucosamine) to the side-chain of asparagine residues that fall under the consensus sequence of NX(S/T), where X can be any amino acid except proline. In *O*-linked glycosylation, carbohydrates are linked to the side-chain of serine or threonine residues. Unlike *N*-linked glycosylation, a universal consensus for *O*-linked glycosylation does not exist (10). However, consensus sequences do exist for certain subclasses of *O*-linked glycosylation. In Notch proteins, specific consensus sequences exist for *O*-linked glycosylation among the EGF repeats of the Notch protein.

As observed in mouse Notch1 (mN1), an important characteristic of Notch's EGF repeat region is the presence of *O*-fucosylation and *O*-glucosylation on the EGF repeats at a high, but

variable stoichiometry (8). *O*-Fucose is added by protein *O*-fucosyltransferase 1 (Pofut1) to serines or threonines found in the consensus sequence $C^2XXXX(\underline{S/T})C^3$, where C^2 and C^3 are the second and third cysteine residues of the EGF repeats (Figure 1). Meanwhile, *O*-glucose is added by protein *O*-glucosyltransferase (Poglut), also referred to as Rumi, to serines in the consensus sequence $C^1X\underline{S}X(P/A)C^2$, where C^1 and C^2 are the first and second cysteine residues of the EGF repeats (Figure 1). Glycosylation by both enzymes requires properly folded EGF repeats (11). In addition to the addition of the first sugar, *O*-fucose and/or *O*-glucose, to the EGF repeats, these *O*-glycans can then be elongated by subsequent additions of carbohydrates on the glycan (2, 8, 11). After *O*-fucosylation, the *O*-fucose can be elongated with a GlcNAc by β1,3-*N*-acetylglucosaminyltransferases of the Fringe family followed by subsequent additions by other enzymes, ultimately resulting in the tetrasaccharide Siaα2-3Galβ1-4GlcNAcβ1-3Fucα1-O-Ser/Thr (Figure 1) (3). After *O*-glucosylation, the *O*-glucose can then be elongated with the addition of xylose by glucoside xylosyltransferases (Gxylt1 or Gxylt2) followed by an additional xylose by xyloside xylosyltransferase (Xxylt1), ultimately resulting in the trisaccharide Xylα1–3Xylα1–3Glcβ1-O-Ser (Figure 1) (11). In addition to *O*-fucosylation and *O*-glucosylation, there has also been evidence reported of an *O*-GlcNAc modification on Ser or Thr residues between the fifth and sixth cysteine residues of the Notch EGF repeats (Figure 1) (12).

Since the majority of CADASIL missense mutations either result in the addition or loss of cysteine residues which could disrupt folding of an EGF repeat, and since *O*-fucosylation and *O*-glucosylation is dependent on the presence of a consensus sequence within a properly folded EGF repeat, this suggests that *O*-glycosylation on the EGF repeats of Notch3 could be altered in CADASIL mutants. However, previous work in our lab suggested that *O*-fucosylation is not affected in CADASIL mutants (13). Nevertheless, Fringe-mediated elongation of *O*-fucose

glycans was impaired in CADASIL mutants of mammalian Notch3 [mouse Notch3 (mN3)] (13). This is an important factor since Fringe-mediated elongation of *O*-fucose results in a change in ligand-dependent activation and subsequent regulation of Notch signaling (3). Nonetheless, we do not know how CADASIL mutations affect the addition of *O*-glucose to EGF repeats nor how they effect elongation of *O*-glucose to the trisaccharide.

Since CADASIL occurs in humans, it is important to conduct a study on *O*-glucosylation and its subsequent elongations on human Notch 3 (hN3). Based on previous studies with *O*-fucosylation and *O*-glucosylation on mN1 (8), we expect to also see *O*-fucosylation and *O*-glucosylation present at a high, but variable, stoichiometry on the EGF repeats of hN3. Using purified hN3 from HEK 293T cells, we were able to determine the structures of *O*-glucose glycans at a majority of predicted sites and their subsequent elongation on the ECD of wild-type hN3. Furthermore, *O*-fucosylation of wild-type hN3 compared to their predicted sites, in the absence of Fringe, was also studied. In addition, because of the presence of an *N*-linked glycosylation site in the hN3 ECD, we also tested the possibility of the *N*-linked glycan hindering detection of the target *O*-linked glycopeptides. These studies provide the foundation for future studies on the effects of CADASIL mutations on *O*-glycosylation of human Notch3.

Chapter 2: Materials and Methods

Materials

Purified wild-type hN3 ECD protein was a generous gift from Dr. Spyros Artavanis-Tsakonas (Harvard Medical School). Before being sent to our lab, the hN3 ECD protein was purified from Human Embryonic Kidney (HEK 293T) cells using the HA and 6x-His tags fused to the end of the hN3 ECD amino acid sequence (Figure 2). All other reagents were of the highest quality available.

hN3 ECD Glycopeptide Analysis via Nano-LC-ESI-MS and MS/MS

The methods described and used below were followed as described (8). Using seven volumes of acetone, approximately 1 µg aliquots of hN3 ECD were acetone precipitated, vortexed, and stored at -20 °C overnight. Protein precipitate was collected as a pellet after centrifugation at 14,000x *g* at 4 °C for 15 minutes. Protein precipitate was then reduced with 8 µl 2X SB/TCEP (50 µl Laemmli Sample Buffer without Reducing Agents + 2 µl 0.5 M TCEP), followed by 30 minutes of carbamidomethylation (in the dark) with 100 mM iodoacetamide (20 mg iodoacetamide + 1 ml 50 mM Tris-HCl, pH 8.0). Protein concentration and purity were assessed using SDS-PAGE and GelCode Blue Stain Reagent alongside BSA standards. hN3 ECD gel bands were cut out into several small cubes and destained using two cycles of destaining with 30% acetonitrile (B&D) in 25 mM ammonium bicarbonate under 10 minute vortexes. The supernatant was then removed and gel pieces were washed in 1 ml of 50%

methanol in 20 mM di-ammonium phosphate (di-AP, pH 8.0) and vortexed twice at room temperature for 20 minutes. A third wash of 50% methanol in 20 mM di-AP was then done at 4 °C overnight before doing a fourth wash and vortex cycle for 30 minutes at room temperature the following day. Gel pieces were then dried using 500 µl 100% acetonitrile. After gel pieces were dried, acetonitrile was removed and the gel pieces were subjected to in-gel digestion using combinations of proteases (300 ng per protease) trypsin, chymotrypsin (Princeton Separations), Asp-N (Sigma), and/or V8 (Sigma, Protease Profiler) for no more than 16 hours at 37 °C using the corresponding enzyme buffers: 20 mM di-AP (trypsin, Asp-N) and 25 mM ammonium bicarbonate (pH 8.0, chymotrypsin, V8). Specifically, the following combinations of proteases were used: chymotrypsin, trypsin, chymotrypsin + trypsin, Asp-N, and V8. Protein digests were then resuspended in 5% formic acid, and then desalted and concentrated using ZipTip C18 cartridges (Millipore). Digests were then diluted in 15% of 95% acetonitrile in 0.1% acetate before being loaded onto the ion trap mass spectrometer (Agilent model 6340) for nano-LC-ESI-MS and MS/MS analysis with collision-induced dissociation (CID) using the settings described below.

hN3 digests (2 µl) were injected onto a Zorbax 300 SB-C18 chip with a 40 nl enrichment column and a 43 mm x 75 µm separation column (Agilent). Samples were loaded on the enrichment column at 4.0 µl/min in 5.0% buffer B and then fractionated on the separation column at 450 nl/min with a 26 min non-linear gradient from 5.0 to 95% buffer B (buffer A: 0.1% formic acid; buffer B: 95% acetonitrile in 0.1% formic acid). Effluents from the HPLC chip were then sprayed into the ion trap mass spectrometer, operating in positive ion mode, where MS/MS was set to be performed on the three most abundant ions per scan. In addition, 30 second active exclusion was set to be triggered after two consecutive spectra with a recurring

ion. In addition, the following settings were used for MS/MS: capillary voltage, 1700-1950 V, end plate Offset, -500 V; dry gas, 5.0 liters/min at 325 °C; trap drive, 100; smart target, 500,000; maximum accumulation time, 150 ms; and scan range, 300 –2200 m/z.

PNGase F Cleavage of N-linked Glycans

The methods described and used below were followed as described (14). Aliquots designated for PNGase F (produced in-house) digestion were subjected to PNGase F digestion first, before undergoing the same methods used for glycopeptide analysis, mentioned above. After acetone precipitation and collection of the protein pellet, the protein pellet is resuspended in 25 µl of 1% SDS, 1% β-mercaptoethanol and boiled at 100 °C for 5 minutes. The samples are then cooled and 225 µl of 50 mM Tris-HCl (pH 8.6), 0.7% Nonidet P-40, and Protease inhibitor cocktail (1 Protease inhibitor tablet per 10 ml buffer) was added to the sample. Samples were digested with 5 U of PNGase F for at least eight hours at 37 °C before going through the same methods listed above for glycopeptide analysis.

Site-Mapping of O-linked Glycopeptides

The methods described and used below were followed as described (8). Prior to analyzing data from the mass spectrometer, putative sites of *O*-linked glycosylation in hN3 EGF repeats were identified using the amino acid sequence of the ECD from hN3 provided by the Artavanis-Tsakonas laboratory. EGF repeats were predicted to be *O*-fucosylated if the consensus sequence $C^2XXXX(S/T)C^3$ (8) was seen in the EGF sequence, and *O*-glucosylated if the consensus

sequence $C^1XSX(P/A)C^2$ (11) was observed (Figure 2). Some EGF repeats are predicted to be both *O*-fucosylated and *O*-glucosylated (8).

In addition to predicting the sites of *O*-fucosylation and *O*-glucosylation on the EGF repeats, *in silico* digests were performed on hN3 ECD with various proteases using the PeptideMass tool from the ExPASy Bioinformatics Resource Portal website (http://web.expasy.org/peptide_mass/). PeptideMass calculates the masses of the predicted peptides, taking into account the added mass of carbamidomethyl groups for each cysteine. In addition to carbamidomethylated cysteines, missed cleavages of up to 3 were used in PeptideMass' settings to take into account possible aberrant proteolytic cleavages as a result of presence of glycans (10). Also, after obtaining the mass spectrometry data, a mass list was exported in mascot generic format (.mgf) and submitted to either GPM's X! P3 (http://ppp.thegpm.org/tandem/thegpm_ppp.html) or Matrix Science's MASCOT (http://www.matrixscience.com/search_form_select.html) online search engines for confirmation that the protein submitted was indeed our target protein (hN3). Peptides identified from these online database searches were used to determine sequence coverage of our in-gel digests.

From the mass spectrometry data, Constant Neutral Loss (CNL) and Extracted Ion Chromatogram (EIC) searches were used to identify which ions represent glycopeptides. CNL searches allow us to search MS/MS spectra for ions that have lost a mass consistent with that of one of our known glycans (neutral loss). For *O*-fucosylation, CNL searches of 73, 48.7, 36.5 were used (Figure 3). From the filtered MS/MS spectra, if a dominant peak is found that loss one of the aforementioned masses (73, 48.7, or 36.5) from the parent ion, this suggests that the MS/MS spectra may be an *O*-fucosylated peptide. In addition, the charge state of the *O*-fucosylated peptide in the MS/MS spectra can be determined, based on the mass loss, denoting a

charge state of +2, +3, or +4 for 73, 48.7 and 36.5, respectively (Figure 3). The ion representing the unmodified peptide in the MS/MS spectra was then used to mathematically determine the mass of the peptide in the +1 charge state and compared to the *in silico* digests of the ECD. In addition, the actual b- and y-fragment ions of the peptide were compared to theoretical b- and y-fragment ions for the peptide, with an error of up to 0.5 mass units, obtained from the University of California, San Francisco's MS-Product (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct), to confirm the peptide's identity. After confirming the identity of the peptide, an EIC search was used to give an estimate of the relative quantities of the different possible glycoforms that may be present for the peptide (*e.g.* unmodified, mono-, di-, tri-, and tetrasaccharide glycoforms). Searches for *O*-fucose disaccharide and longer chains were not focused on in this study because previous evidence suggests that Fringe elongation of *O*-fucose is not abundant in HEK 293T cells (Kakuda S and Haltiwanger RS, unpublished). Similar CNL and EIC searches were used for identifying *O*-glucosylated peptides. Monosaccharide forms of *O*-glucosylated peptides were determined by searching for a neutral loss of 81, 54, and 40.5 in the +2, +3, and +4 forms, respectively (Figure 3). In addition, trisaccharide forms were detected by an order of losses in the following pattern (A-[A+B]-[A+B+C]; first-second-third, right to left): 66-132-213, 44-88-142, and 33-66-106.5 in the +2, +3, and +4 forms, respectively.

Site-Mapping of Aberrant Cleavages

If putative glycopeptides were found using the CNL method described above, but did not match with the predicted masses found from *in silico* digests, it was assumed that an aberrant

cleavage had occurred. As a result, the amino acid sequence provided by the Artavanis-Tsakonas lab was used with the FindPept tool from the ExPASy Bioinformatics Resource Portal website (http://web.expasy.org/findpept/) with the determined mass of the unmodified peptide found in the MS/MS spectra. Putative peptide sequences provided by FindPept were then input into MS-Product for confirmation by comparing b- and y-peptide fragment ions.

Chapter 3: Results

Total Sequence Coverage and Summary

Based on the sequence of hN3 ECD (Figure 2), a total sequence coverage map was made from the peptide sequences identified by GPM, MASCOT, and manual CNL searches for glycopeptides (Figure 4). Approximately 64.4% of hN3 ECD sequence was found (1042 out of 1619 of total amino acids in the Notch3 ECD). Of the region encoding EGF repeats, approximately 63.0%, of the amino acids were identified (841 out of the 1334 amino acids coding for EGF repeats). Figure 5 summarizes the glycans mapped to predicted sites of hN3, 24 of which are predicted to be *O*-glycosylated. Glycopeptides from 16 out of the 24 predicted sites were identified. In addition, one *O*-fucose modified peptide was found from a site not predicted to be modified in EGF repeat 23 (Figure 5). Table 1 summarizes the peptide sequences of the identified glycopeptides along with the predicted and actual masses of the unmodified peptides, and the mass difference between the found glycopeptide and unmodified peptide, where all masses were denoted in the +1 charge state. In addition, peptides were matched with their amino acid sequence numbers and corresponding EGF repeats (Table 1). It should be noted that lack of observation of the glycopeptides from the remaining predicted sites does not indicate that those EGF repeats are not *O*-glycosylated, but rather, the glycopeptides were simply not observed at the time of this study.

Human Notch3 EGF Repeats Modified with O-Fucose Monosaccharide

Using CNL searches of 73, 48.7, 36.5 to represent the +2, +3, and +4 charge states of fucose, respectively, along with corresponding masses for di-, tri-, and tetrasaccharide forms (Figure 3), *O*-fucosylated peptides were discovered and compared to predicted masses from *in silico* digests of hN3 ECD. Using the *O*-fucosylation consensus sequence $C^2XXXX(\underline{S}/\underline{T})C^3$, glycopeptides from 10 of hN3's EGF repeats are predicted to only contain this consensus sequence with no *O*-glucose consensus sequence (Figure 5, all red). Out of these ten sites, *O*-fucosylated glycopeptides were identified from five, in EGFs 6, 7, 8, 24 and 28 (Figures 12A, 13A, 14A, 25A, 26A). EIC searches for naked, mono-, di-, tri-, and tetrasaccharide glycoforms of the peptides showed that all 5 glycopeptides were found to be mainly in the monosaccharide form (Figures 12B, 13B, 14B, 25B). The glycopeptide from EGF repeat 28 showed the presence of a di- and trisaccharide form and will be discussed at a later section (Figure 26B). CNL and EIC searches of the matching di-, tri- and tetrasaccharide glycoforms also showed lack of elongated forms of *O*-fucosylated peptides. This is consistent with previous data that Notch proteins derived from HEK 293T cells show essentially no elongation of *O*-fucose on Notch1 by Fringe family members. As a result, *O*-fucosylated peptides cannot be elongated to the GlcNAcβ1-3Fucα1-O-Ser/Thr disaccharide form, which is required for further elongation to the tri- and tetrasaccharide forms by corresponding enzymes (Kakuda S and Haltiwanger RS, unpublished).

Human Notch3 EGF Repeats Modified with the O-Glucose Trisaccharide

Similar to the methods used for discovering *O*-fucosylated peptides, trisaccharide *O*-glucosylated peptides were discovered using CNL searches of 213, 142, and 106.5 for the

respective +2, +3, and +4 charge states denoting loss of the full *O*-glucose trisaccharide in the MS/MS (Figure 3). In addition, unlike finding peptides with *O*-fucose monosaccharides, a distinct pattern of neutral losses could be seen in the order xylose, xylose, glucose in order to definitively identify the glycopeptide. Since *O*-glucosylated peptides are elongated in a specific order, observing a different fragmentation pattern is unlikely (8). Using the consensus sequence $C^1X\underline{S}X(P/A)C^2$, 10 of the EGF repeats in hN3 are predicted to be only *O*-glucosylated with no *O*-fucose (Figure 5, all blue). Out of these ten sites, *O*-glucosylated glycopeptides were identified from seven sites: EGFs 2, 3, 12, 13, 16, 23, and 32. The *O*-glucose was elongated to the Xylα1–3Xylα1–3Glcβ1-O-Ser trisaccharide form in most cases (Figures 7A, 10A, 17A, 18A, 19A, 20A, 21A), but EGF repeat 32 was only found in the monosaccharide form which will be discussed later (Figure 28A). Furthermore, EGF repeat 23 was found to also be *O*-fucosylated, which will also be discussed later (Figure 24A). EIC searches of corresponding glycopeptides showed that EGFs 12, 13, and 16 were predominantly found in the Xylα1–3Xylα1–3Glcβ1-O-Ser trisaccharide glycoform (Figures 17B, 18B, 19B, 20B, 21B). However, corresponding EIC searches of the naked, mono-, di-, and trisaccharide glycoforms of the peptides from EGFs 2 and 3 showed that these peptides were not found predominantly in the trisaccharide form, but instead, were predominantly in the Glcβ1-O-Ser monosaccharide form (Figures 7B, 10B).

Human Notch3 EGF Repeats Modified with the O-Glucose Monosaccharide

As mentioned above, EGF repeats 2 and 3 were found to be *O*-glucosylated, but were not elongated to the predicted trisaccharide form by Gxylt1, Gxylt2, and Xxylt2 at a high efficiency. As a result, the EICs of the two EGFs were found to be predominantly in the monosaccharide

form (Figures 7B, 10B). To confirm these findings, CNL searches for loss of the *O*-glucose monosaccharide were performed using 81, 54, and 40.5 for the +2, +3, and +4 charge states, respectively, to identify peptides that were only *O*-glucosylated, but not further elongated (Figure 3). Using this search, of the ten EGF repeats predicted to be *O*-glucosylated, *O*-glucose monosaccharide glycopeptides corresponding to EGF repeats 2 and 3 were found (Figures 6A, 8A, 9A, 11A). In addition, an *O*-glucose monosaccharide glycopeptide from EGF repeat 32 was also identified (Figure 28A). Using EIC searches of the different glycoforms, EGFs 2, 3 and 32 were all found predominantly in the Glcβ1-O-Ser monosaccharide form (Figures 6B, 8B, 9B, 11B, 28B). It should be noted that the *O*-glucose monosaccharide has only been previously observed at a few sites in mouse Notch3, where the *O*-glucose monosaccharide form has been observed on EGF repeats 2, 12, and 25 (Tan E and Haltiwanger RS, unpublished), and on several EGF repeats from *Drosophila* Notch (Rana N and Haltiwanger, unpublished). In contrast, *O*-glucose was found fully elongated to the trisaccharide at all 17 predicted sites on mouse Notch1 (8).

Human Notch3 EGF Repeats Modified with Both O-Fucose and O-Glucose Glycans

Since *O*-fucosylation and *O*-glucosylation on Notch EGF repeats have different consensus sequences on different regions (Cys2-Cys3 and Cys1-Cys2, respectively), it is possible that Notch EGF repeats can have both consensus sequences in the same EGF repeat. In addition, it has been previously observed that EGF repeats with both consensus sequences can be both *O*-fucosylated and *O*-glucosylated at the same time (8). As such, four EGF repeats on hN3 are predicted to be both *O*-fucosylated and *O*-glucosylated at the same time: EGFs 11, 17, 19,

and 29 (Figure 5, both red and blue). Of these, a glycopeptide modified with both *O*-fucose monosaccharide and *O*-glucose trisaccharide was detected from EGF repeat 11 (Figure 15A). The specific order of mass losses during MS/MS fragmentation of this glycopeptide (*i.e.* either the fucose or a xylose can be lost first) demonstrate that the *O*-glucose trisaccharide is independent from the *O*-fucose monosaccharide, further suggesting that both *O*-fucosylation and *O*-glucosylation is present on EGF repeat 11 (Figure 15C). In addition, a glycopeptide from EGF repeat 19 modified with *O*-fucose monosaccharide was also observed (Figures 22A, 23A). Unfortunately, the peptide sequence begins after the Ser residue that is supposed to be *O*-glucosylated, thus suggesting that it is still possible that *O*-glucosylation occurs on EGF repeat 19, but was not observed at the time of this study (Table 1).

Non-Canonical Forms of O-Glycosylation on Human Notch3

In addition to the predicted forms of *O*-glycosylation observed above, several unusual forms of *O*-glycosylation were also observed. For instance, as mentioned above, a glycopeptide from EGF repeat 28 appeared to be modified with an *O*-fucose disaccharide based on the EIC search (Figure 26B). However, in the MS/MS fragmentation of the glycopeptide, losses of masses corresponding to the GlcNAc and fucose were observed to be independent of each other, suggesting that the two carbohydrates are not linked together as a disaccharide (Figure 26A). Furthermore, upon closer observation of the identified peptide, a large portion of EGF repeat 27's amino acid sequence was seen. These data suggest that an *O*-GlcNAc modification exists on this peptide since *O*-GlcNAc modifications on Notch EGF repeats have been observed on Ser and Thr residues between Cys5 and Cys6 (12), which are present in this peptide's sequence

(Figure 26A). Although not analyzed in detail, the trisaccharide glycoform of this peptide detected in the EIC (Figure 26B) is most likely elongation of the *O*-GlcNAc by galactose plus *O*-fucose.

As mentioned above, another unusual form of *O*-glycosylation was observed on EGF repeat 23 (Figure 24A). EGF 23 does not contain a predicted O-fucose site based on the $C^2XXXX(\underline{S}/\underline{T})C^3$ consensus (Figure 2), but it does have a similar sequence: $C^2GPG\underline{T}C^3$, with only three amino acids between C^2 and the modified threonine. The glycopeptide from EGF repeat 23, which contains an O-glucose consensus sequence and is modified with *O*-glucose trisaccharide, is also modified with *O*-fucose monosaccharide (Figure 24A, C). The EIC suggests that the fully *O*-fucosylated and *O*-glucosylated glycoform is the major species (Figure 24B). These results suggests that three variable amino acids, except for proline, can be used for *O*-fucosylation in addition to the known four variable amino acids of the consensus sequence (Figure 24A). This is an interesting observation because the old consensus sequence for *O*-fucosylation was $C^2X_{3-5}(\underline{S}/\underline{T})C^3$ before it was later revised to the current consensus sequence because a similar site on EGF15 of mouse Notch1 is not *O*-fucosylated (15).

Finally, on EGF repeat 30, the unmodified peptide was observed instead of the predicted modification with *O*-fucosylation (Figure 5). This was interesting to observe because, based on previous work, all EGF repeats predicted for *O*-glycosylation should be modified. However, this peptide was observed as one of the peptides found during a MASCOT database search that was not set to detect glycosylated peptides. Based on MASCOT's identification of the peptide's compound number in the appropriate mass spectra, the b- and y- ions were found to match the peptide's sequence, without the presence of a fucose (Figure 27A). An EIC search of the possible forms of *O*-fucosylation of the peptide further suggests that this peptide was indeed found in its

unmodified form (Figure 27B). However, it is interesting to note that the sequence also contains a consensus sequence for *N*-linked glycosylation of NX(S/T) (10). This suggests that *N*-linked glycosylation is responsible for blocking *O*-fucosylation at this EGF repeat. Furthermore, this peptide was only observed when the protein was first digested with PNGase F, which removes *N*-linked glycans from peptides, strongly suggesting that *N*-linked glycosylation may be responsible for blocking *O*-fucosylation at this EGF repeat (Table 1).

Chapter 4: Discussion

The results from this study are largely consistent with previous findings; the identified EGF repeats were *O*-fucosylated and/or *O*-glucosylated in accordance with the consensus sequences used to predict the *O*-glycosylation modifications. As summarized in Figure 5, six out of the ten EGF repeats predicted to be only *O*-fucosylation were identified, seven out of the ten EGF repeats predicted to be only *O*-glucosylated were also identified, and finally, two of the three EGF repeats predicted to be both *O*-fucosylated and *O*-glucosylated were also identified, although *O*-glucosylation was only observed on one of the two EGF repeats. Furthermore, it was shown that, in contrast with previous observations (8), the elongation of *O*-glucose was variable among different EGF repeats. As observed in EGF repeats 2, 3 and 32, instead of the fully modified trisaccharide, the *O*-glucose monosaccharide was found to be the predominant form (Figure 5). However, further work needs to be done to determine the reason for variability in elongation of *O*-glucosylated EGF repeats in hN3.

Furthermore, Figure 5 also shows the presence of non-canonical *O*-linked modifications for two of the identified EGF repeats in addition to another EGF repeat that was previously predicted to not have an *O*-linked modification. EGF repeat 23 was observed to be *O*-fucosylated in the absence of the current *O*-fucosylation consensus sequence $C^2XXXX(S/T)C^3$. At one point, *O*-fucosylation was believed to use the consensus sequence, $C^2X_{3-5}(S/T)C^3$. This consensus sequence was refined based on the observation showing the absence of *O*-fucosylation at a site in EGF repeat 15 in mouse Notch1 with the sequence, C^2HYGSC^3 (15). The fact that EGF repeat 23 is O-fucosylated in hN3 is efficiently O-fucosylated suggests that sites with just three amino

acids between C2 and the modified residue need to be re-examined, and the consensus sequence needs to be expanded back to C^2X_{3-5}(S/T)C3. In addition, it suggests that there is something about EGF repeat 15 from mouse Notch1 other than the three amino acids that is preventing *O*-fucosylation.

The data in Figure 25 suggests that EGF repeat 27 may contain an *O*-GlcNAc modification. Although the Ser and Thr residues between the fifth and sixth cysteine residues in an EGF repeat are known targets (12), a consensus sequence for *O*-GlcNAc modifications is still unknown. Finally, EGF repeat 30 contains an *O*-fucosylation consensus sequence, but was found unmodified. However, the consensus sequence for *O*-fucosylation in EGF 30 was found to also contain the consensus sequence for *N*-linked glycosylation (Table 1). The fact that the peptide from EGF repeat 30 was only detected after PNGase F digestion strongly suggests that this site is modified with an *N*-glycan and suggests that *N*-glycosylation, normally a co-translational process, occurs before O-fucosylation, and blocks addition of fucose.

Based on the observations made in this study, it is clear that more work needs to be done to determine the *O*-glycosylation pattern of the remaining 37% of EGF repeats predicted to be modified that were not mapped during this study's analysis. By determining the *O*-glycosylation pattern of the remaining 37%, more insight can be made of the existence of further "non-canonical" glycosylation sites, as well as to confirm whether the remaining unmapped sites are indeed glycosylated in accordance to their consensus sequences. At the time of this study, the major proteases used were chymotrypsin, trypsin, Asp-N, V8 (i.e. Glu-C) and combinations of chymotrypsin + trypsin and chymotrypsin + PNGase F. Unfortunately, the V8 used in this study resulted in aberrant peptide cleavages, which could not be identified even with the method described to identify peptides with aberrant cleavages using FindPept. However, the discovery of

an additional identified EGF repeat, EGF repeat 6, using a combination of chymotrypsin and trypsin, suggests that other combinations of proteases may be promising in identifying the *O*-glycosylation patterns of the remaining EGF repeats of human Notch3.

In addition, in contrast to previously used methods where truncated groups of Notch EGF repeats were used for *O*-glycosylation analysis (8, 13), a full-length protein of human Notch3's ECD was used for this thesis. The advantage of using a full-length protein, as opposed to truncated versions, is that the full-length protein might more closely reflect the glycosylation patterns that would be observed in the protein *in vivo.* However, the ion trap mass spectrometer that was used was set to fragment the three most prevalent ions in each MS spectra for MS/MS analysis. As a result, it is possible that some peptides containing EGF repeats were ignored due to the prevalence of other peptides.

In addition to increasing the coverage of identified peptides, possible factors for variability in elongation of *O*-glucosylated peptides in hN3 should be studied. As done by Takeuchi et al. 2012 (11), analysis of enzyme activity of Gxylt1, Gxylt2 and Xxylt1 on single EGF repeat constructs of hN3 may provide some insight to the efficiency of these elongation enzymes on specific EGF repeats. Preliminary studies suggest that amino acids in the X position of the *O*-glucose consensus sequence of EGF repeats from mouse Notch1 and mouse Notch2 may contribute to the extent of elongation of *O*-glucose (11). Comparison of EGF repeats 2, 3, and 32 (all of which are only elongated to the *O*-glucose monosaccharide form) with other EGF repeats from hN3 in *in vitro* assays with purified Gxylt1 may provide a useful platform to test this hypothesis.

Finally, a notable characteristic of hN3 is its relation to the pathogenesis of CADASIL. Here we have laid the groundwork for these studies by analyzing the glycosylation of wild-type hN3. However, in common with how *O*-fucosylation and elongation by Fringe was studied previously (13), *O*-glucosylation and elongation by Gxylt1, Gxylt2, and Xxylt1 could now be studied in CADASIL mutants of hN3 and compared to the findings made in this thesis. As described before, a diagnostic characteristic of CADASIL is the accumulation of GOMs and Notch3 ECD around vascular smooth muscle cells, and it is possible that changes in the *O*-glycosylation of hN3 may contribute to the aggregation of hN3 ECD. Since *O*-glucosylation and elongation are also major components of Notch, a comparison between the results of this study and of CADASIL mutants should be made to determine key differences that may be responsible for aggregation of hN3 ECD in CADASIL patients.

Chapter 5: Conclusion

In comparison to previous studies (8), the results presented in this thesis show that hN3's ECD, in general, follows the pattern of *O*-fucosylation and/or *O*-glucosyation and elongation among the predicted sites of *O*-linked glycosylation. As such, hN3 EGF repeats were seen to be *O*-fucosylated in the presence of the consensus sequence $C^2XXXX(\underline{S}/\underline{T})C^3$ (8), but not elongated with GlcNAc by Fringe to the disaccharide form because of lack of or little Fringe activity in HEK 293T cells (Kakuda S and Haltiwanger RS, unpublished). However, hN3 EGF repeats were found to *O*-glucosylated and, usually, elongated to their trisaccharide form by Gxylt1, Gxylt2 and Xxylt1 in the presence of the consensus sequence $C^1X\underline{S}X(P/A)C^2$ (11). However, it is important to note that this is not always the case as EGF repeats 2, 3, and 32 in hN3 have been found to be only *O*-glucosylated, but not predominantly elongated to the trisaccharide form (Figure 5). In addition, several "non-canonical" forms of *O*-linked glycosylation were found. This suggests that more work is needed on determining the various mechanisms of *O*-linked glycosylation.

In addition, because mutations exclusive to human Notch3 contribute to the pathogenesis of CADASIL, it is important to determine the *O*-linked glycosylation pattern of wild-type hN3 in order to compare it to the *O*-linked glycosylation pattern of CADASIL mutants of hN3. Wild-type hN3 was done for this thesis, but future work needs to be done for CADASIL mutants of hN3 in order to determine how *O*-glucosylation and elongation is affected in CADASIL. *O*-fucosylation was found to be unaffected on mouse Notch1 EGF repeats, but evidence exists suggesting that elongation of *O*-fucose by Fringe is affected (13).

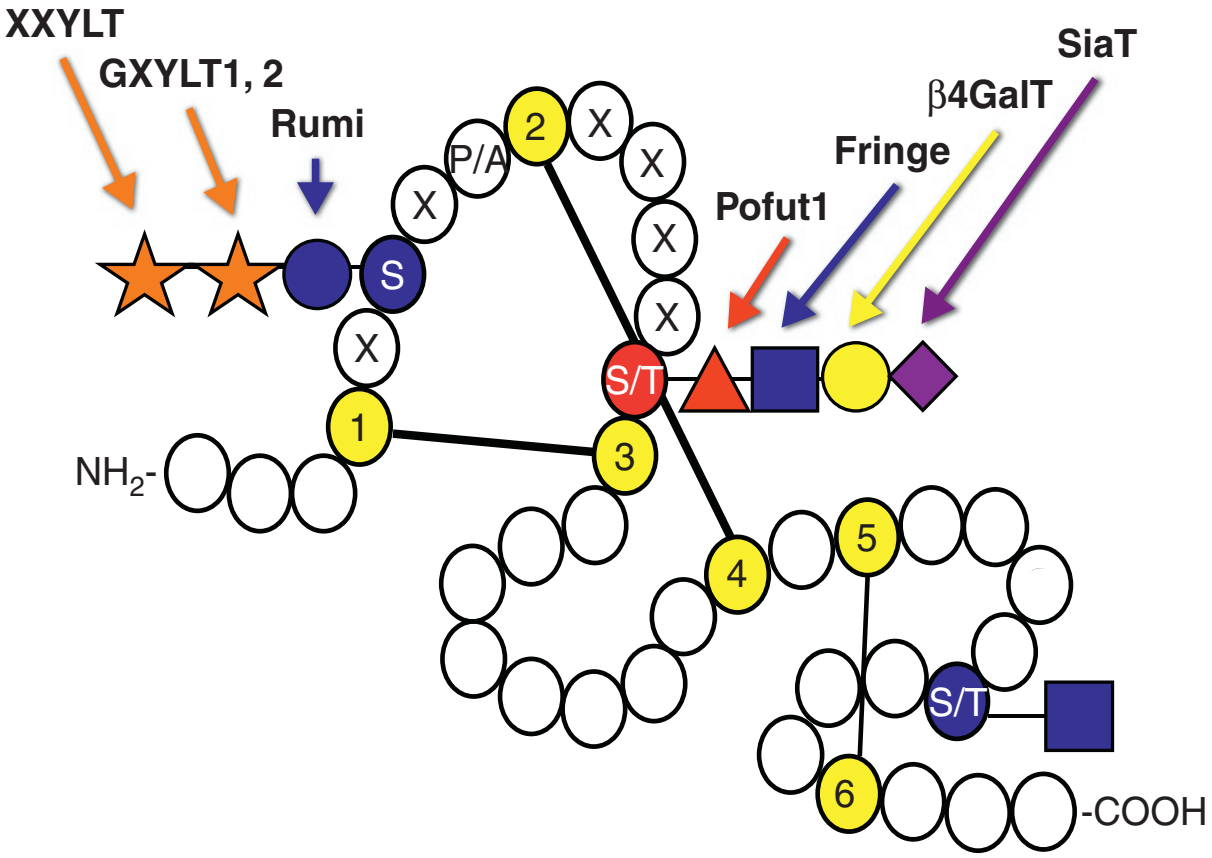


Figure 1: Graphical Representation of an EGF Repeat, Modified from (2): Modified from (2), a 40 amino acid motif of a Notch EGF repeat with canonical *O*-Glucosylation and *O*-Fucosylation consensus sequences with maximum possible elongation modification and elongation of these sites are shown above. Enzymes responsible for addition of each sugar are indicated. Cysteine residues of the EGF repeat have been colored yellow and numbered. Cysteine residues have also been paired with their corresponding cysteine residues that they form disulfide bridges with in a properly folded EGF repeat. A consensus sequence for *O*-GlcNAc addition is unknown, but serine and threonine residues between the fifth and sixth cysteine residues have been observed to exhibit this modification.

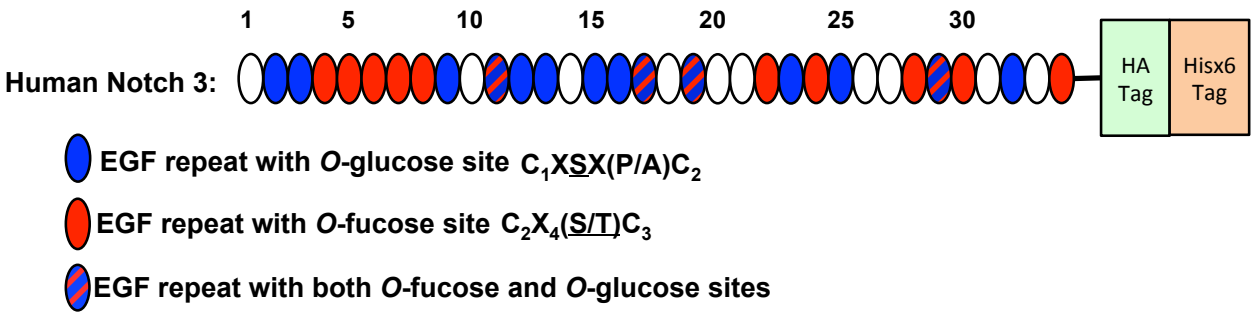


Figure 2: Graphical Representation of hN3 ECD + HA-His$_6$ Tag Used in this Study. hN3 ECD contains 34 EGF repeats with 24 EGF repeats predicted to be modified by *O*-linked glycosylation. Provided hN3 ECD contains HA tag and His$_6$ tag at C-terminus. EGF repeats (ovals) are colored red if they contain an *O*-Fucose consensus, blue if they contain an *O*-glucose consensus, and both if they contain both.

Modification Type	Charge State (+)	Mass	Modification Type	Charge State (+)	Mass
Fuc	1	146	Glc	1	162
	2	73		2	81
	3	48.7		3	54
	4	36.5		4	40.5
GlcNAc-Fuc	1	349	Xyl-Glc	1	294
	2	174.5		2	147
	3	116.3		3	98.45
	4	87.25		4	73.5
Gal-GlcNAc-Fuc	1	511	Xyl-Xyl-Glc	1	426
	2	255.5		2	213
	3	170.3		3	142
	4	127.75		4	106.5
GlcNAc	1	203	Xyl	1	132
	2	101.5		2	66
	3	67.67		3	44
	4	50.75		4	33
Gal	1	162	NANA	1	290
	2	81		2	145
	3	54		3	96.7
	4	40.5		4	72.5

Figure 3: Constant Neutral Loss Mass Numbers of Carbohydrates. Listed above are the expected mass losses of specific carbohydrates and carbohydrates according to specific charge state of the peptide in the MS/MS spectra during Constant Neutral Loss searches.

APPCLDGSPCANGGRCTQLPSREAACLCPPGWVGERCQLEDPCHSGPCAGR
GVCQSSVVAGTARFSCRCPRGFRGPDCSLPDPCLSSPCAHGARCSVGPDGR
FLCSCPPGYQGRSCRSDVDECRVGEPCRHGGTCLNTPGSFRCQCPAGYTGP
LCENPAVPCAPSPCRNGGTCRQSGDLTYDCACLPGFEGQNCEVNVDDCPGH
RCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVDECQLQPNACHNGGTCFNT
LGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASFYCACPMGKT
GLLCHLDDACVSNPCHEDAICDTNPVNGRAICTCPPGFTGGACDQDVDECSIG
ANPCEHLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRI
GQFTCICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGS
TCQLDVDECASTPCRNGAKCVDQPDGYECRCAEGFEGTLCDRNVDDCSPDP
CHHGRCVDGIASFSCACAPGYTGTRCESQVDECRSQPCRHGGKCLDLVDKYL
CRCPSGTTGVNCEVNIDDCASNPCTFGVCRDGINRYDCVCQPGFTGPLCNVEI
NECASSPCGEGGSCVDGENGFRCLCPPGSLPPLCLPPSHPCAHEPCSHGICY
DAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAGGTCSSDGMGFHCTC
PPGVQGRQCELLSPCTPNPCEHGGRCESAPGQLPVCSCPQGWQGPRCQQD
VDECAGPAPCGPHGICTNLAGSFSCTCHGGYTGPSCDQDINDCDPNPCLNGG
SCQDGVGSFSCSCLPGFAGPRCARDVDECLSNPCGPGTCTDHVASFTCTCPP
GYGGFHCEQDLPDCSPSSCFNGGTCVDGVNSFSCLCRPGYTGAHCQHEADP
CLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQTLVDWCSRQPCQNGGRCV
QTGAYCLCPPGWSGRLCDIRSLPCREAAAQIGVRLEQLCQAGGQCVDEDSSH
YCVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGYMGGYMCECLPGYNGDN
CEDDVDECASQPCQHGGSCIDLVARYLCSCPPGTLGVLCEINEDDCGPGPPLD
SGPRCLHNGTCVDLVGGFRCTCPPGYTGLRCEADINECRSGACHAAHTRDCL
QDPGGGFRCLCHAGFSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGGLTFT
CHCAQPFWGPRCERVARSCRELQCPVGVPCQQTPRGPRCACPPGLSGPSCR
SFPGSPPGASNASCAAAPCLHGGSCRPAPLAPFFRCACAQGWTGPRCEAPAA
APEVSEEPRCPRAACQAKRGDQRCDRECNSPGCGWDGGDCSLSVGDPWRQ
CEALQCWRLFNNSRCDPACSSPACLYDNFDCHAGGRERTCNPVYEKYCADHF
ADGRCDQGCNTEECGWDGLDCASEVPALLARGVLVLTVLLPPEELLRSSADFL
QRLSAILRTSLRFRLDAHGQAMVFPYHRPSPGSEPRARRELAPEVIGSVVMLEI
DNRLCLQSPENDHCFPDAQSAADYLGALSAVERLDFPYPLRDVRGEPLEPPEP
SVPLYPYDVPDYAHHHHHH

Figure 4: Total Amino Acid Sequence Coverage of hN3. Total amino acid sequence of hN3 ECD. Highlighted amino acids denote amino acids identified by GPM, MASCOT or manual searches (e.g. identified glycopeptides identified using CNL searches). Green amino acids denote amino acids that code for EGF repeats. Blue amino acids denote those predicted *O*-glucosylated. Red amino acids denote those predicted to be *O*-fucosylated. Red Highlighted amino acids denote HA + His x6 Tags. 1042 out 1619 AAs were identified.

hN3 *O*-Glycosylation Sites (WT)

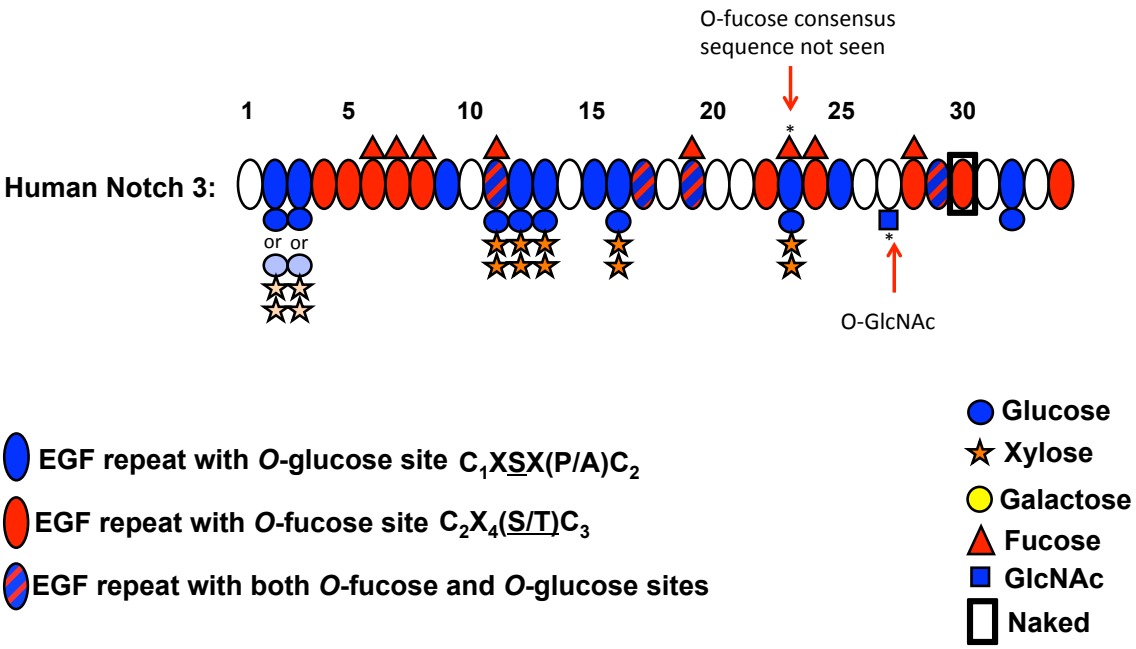


Figure 5: Graphical Summary of Predicted and Identified Sites of hN3. Identified sites of *O*-linked glycosylation sites of EGF repeats on wild-type hN3 were marked according to the legend key shown above. Faded colors on EGF repeats 2 and 3 signify that the trisaccharide form is not the predominant form.

EGF	AA Sequence	Parent Ion (M+H)	Naked Peptide (M+H)	Mass Δ	Naked Pred. Mass (M+H)	Enzyme Used	Figure #
		1907.5	1744.6	162.9	1743.6	Tryp	5
		2172.4	1744.6	427.8	1743.6	Tryp	6
2	^{76}CQLEDPCHSGPCAGR^{90}	1890	1727.2	162.8	1743.6*	Tryp	7
		2317.9	2155	162.9	2155.3	Tryp	8
	^{114}GPDCSLPDPCLSSPCAHGAR^{133}	2581	2155.6	425.4	2155.4	Tryp	9
3	^{113}RGPDCSLPDPCLSSPCAHGARCSVGPDGRFL^{143}	3563.4	3400.2	163.2	3400.8	Chym	10
6	^{245}CLNGGTCVDGVNTY^{258}	1677	1530.2	146.8	1529.6	Chym + Tryp	11
7	^{259}NCQCPPEWTGQFCTEDVDECQLQPNACHNGGTCF^{292}	4267	4120.2	146.8	4119.4	Chym	12
8	^{325}HGATCHDRVASFY^{338}	1668.4	1520.8	147.6	1520.6	Chym	13
	^{437}SGPCRNQATCL^{447}	1837.6	1263.8	573.8	1263.5	Chym	14
11	^{428}CETDVNECLSGPCR^{441}	2124.6	1697.2	427.4	1696.6	Tryp	15
12	^{471}DECQSSPCVNGGVCK^{485}	2124.6	1697.6	427	1696.6	Asp-N	16
	^{488}VNGFSCTCPSGFSGSTCQLDVDECASTPCR^{517}	3786.2	3359	427.2	3358.6	Tryp	17
	^{488}VNGFSCTCPSGFSGSTCQLDVDECASTPCRNGAK^{521}	4157.8	3730.6	427.2	3729	Tryp	18
13	^{500}SGSTCQLDVDECASTPCRNGAKCVDQPDGY^{529}	3793.3	3366.7	426.6	3349.5*	Chym	19
16	^{605}LCRCPSGTTGVNCEVNIDDCASNPCTF^{631}	3530.4	3106.2	424.2	3106.4	Chym	20
		2543.8	2397.4	146.4	2397.6	Tryp	21
19	^{745}AGGTCSSDGMGFHCTCPPGVQGR^{767}	2543.2	2397.1	146.1	2397.6	Chym + Tryp	22
23	^{872}SCSCLPGFAGPRCARDVDECLSNPCGPGTCTDHVASF^{908}*	4690.2	4117.8	572.4	4117.5	Chym	23
24	^{909}TCTCPPGYGGFHCEQDLPDCSPSSCFNGGTCVDGVNSF^{946}	4393.4	4247	146.4	4246.5	Chym	24
27-28	^{1070}CVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGY^{1102}*	4213.8	3865.4	348.4	3865.2	Chym	25
30	^{1178}HNGTCVDLVGGF^{1189}	1277.4	1276.6	0.8	1275.6	Chym + PNGase F	26
32	^{1248}SPCESQPCQHGGQCRPSPGPGGGLTF^{1273}	2919.7	2755.9	163.8	2757	Chym	27

Table 1: Summary of Identified Glycopeptides from Human Notch3. Amino acid sequences of identified glycopeptides are paired with their corresponding EGF repeat number, parent ion mass, naked peptide mass, mass difference between parent ion and naked peptide, predicted mass of naked peptide, enzyme used during in-gel digestion and corresponding figure number (Figures 5-27). Observed masses were all converted to their +1 charge state. Predicted mass numbers with asterisk (*) denote loss or gain of H_2O between actual and predicted naked peptide (~18). Mass differences are compared to masses in Figure 3, and corresponding masses indicate specific *O*-linked glycosylation exhibited such as 572 corresponding with peptide exhibiting *O*-glucose trisaccharide and *O*-fucose monosaccharide. Blue, red, and green respectively indicate *O*-glucosylated, *O*-fucosylated and *O*-GlcNAc modified amino acids.

Figures 6A-28A: MS, MS/MS Spectra of Identified EGFs.

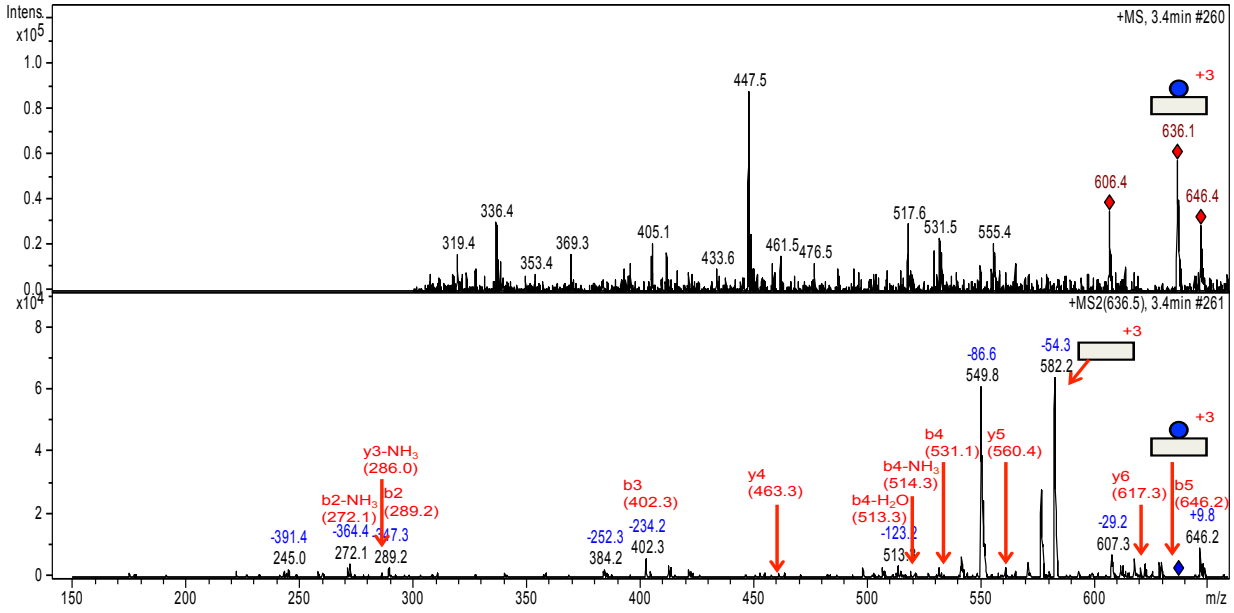
Figures 6B-28B: EICs of Identified EGFs.

Figures 15C, 24C: Zoomed in MS/MS Spectra of Identified EGFs.

MS, MS/MS Spectra, EICs, and Zoomed in MS/MS spectra shown in the following pages correlate with their Figure ID number shown in Table 1. Top half spectra in "A's" denote MS spectra while bottom half spectra in "A's" denote MS/MS spectra with identified b- and y- ions marked in the MS/MS spectra. Glycopeptide forms found were marked using the following key: "naked" glycopeptide = grey rectangle, +glucose = blue circle, +xylose = orange star, +fucose = red triangle, +GlcNAc = blue square, +galactose = yellow circle, +NANA = purple diamond. Zoomed in MS/MS (Figures 14C, 23C) show independent fragmentation of *O*-fucose from *O*-glucose trisaccharide, indicating that the two carbohydrates are independent modifications on the glycopeptide.

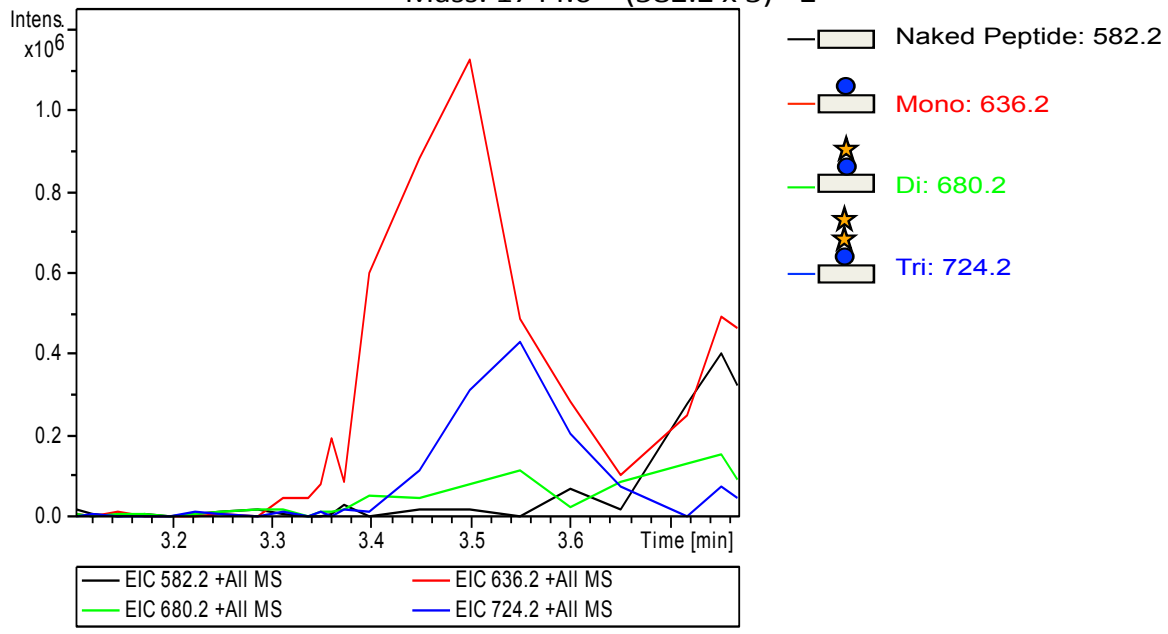
6A

CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Mono)
Mass: 1744.6 = (582.2 x 3) - 2



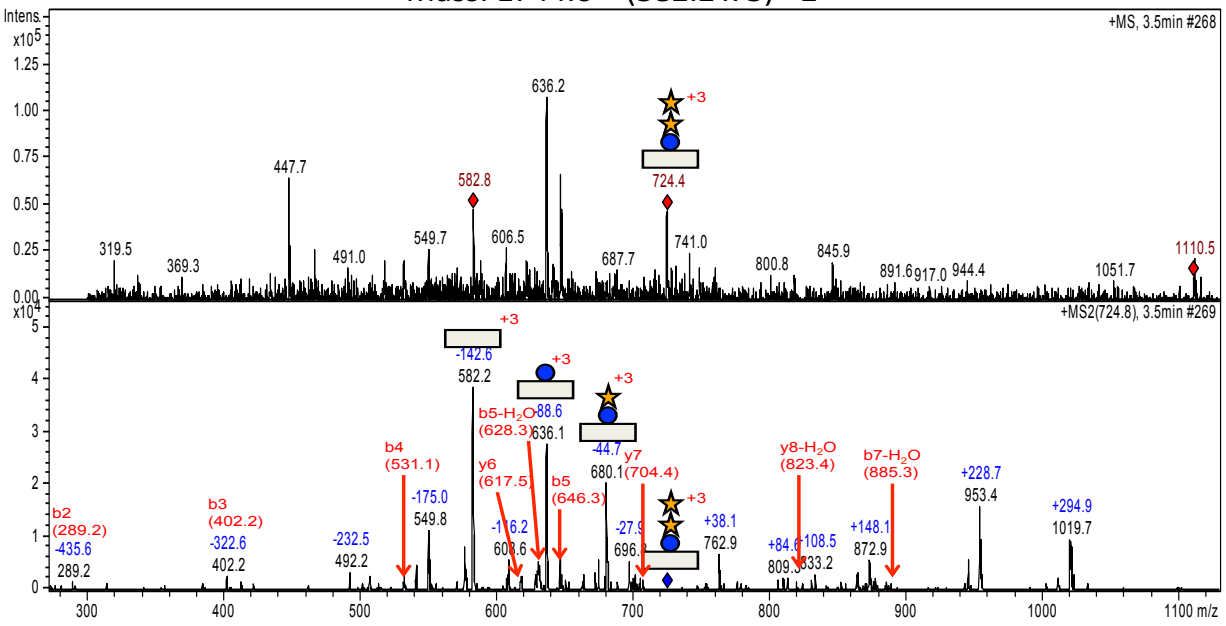
6B

CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Mono)
Mass: 1744.6 = (582.2 x 3) - 2



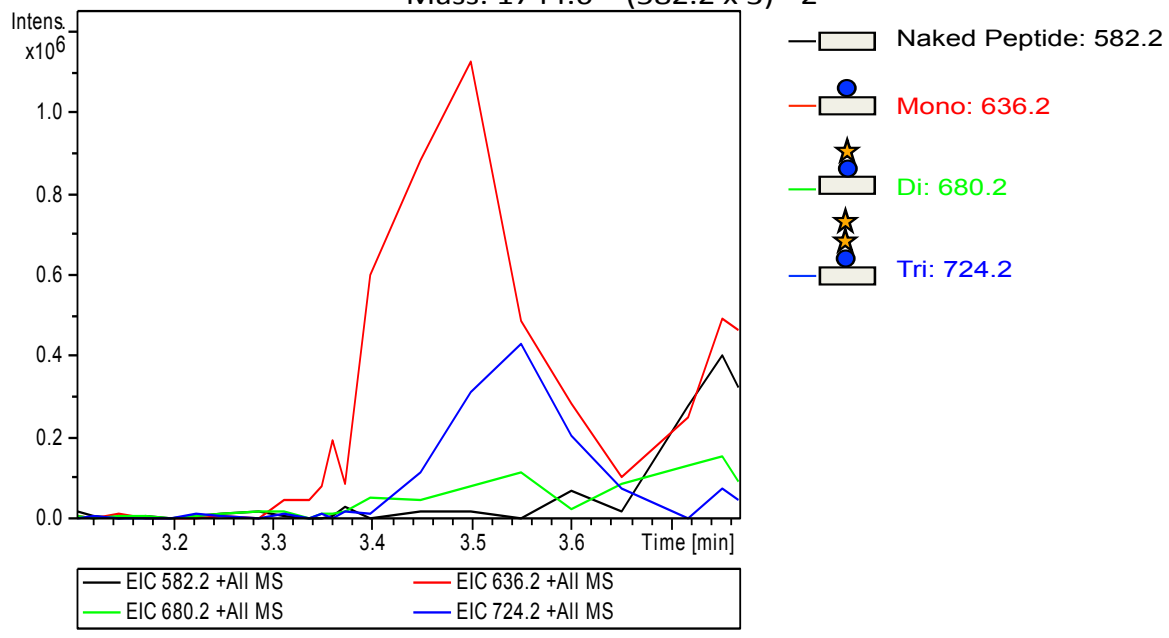
7A

CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Tri)
Mass: 1744.6 = (582.2 x 3) - 2

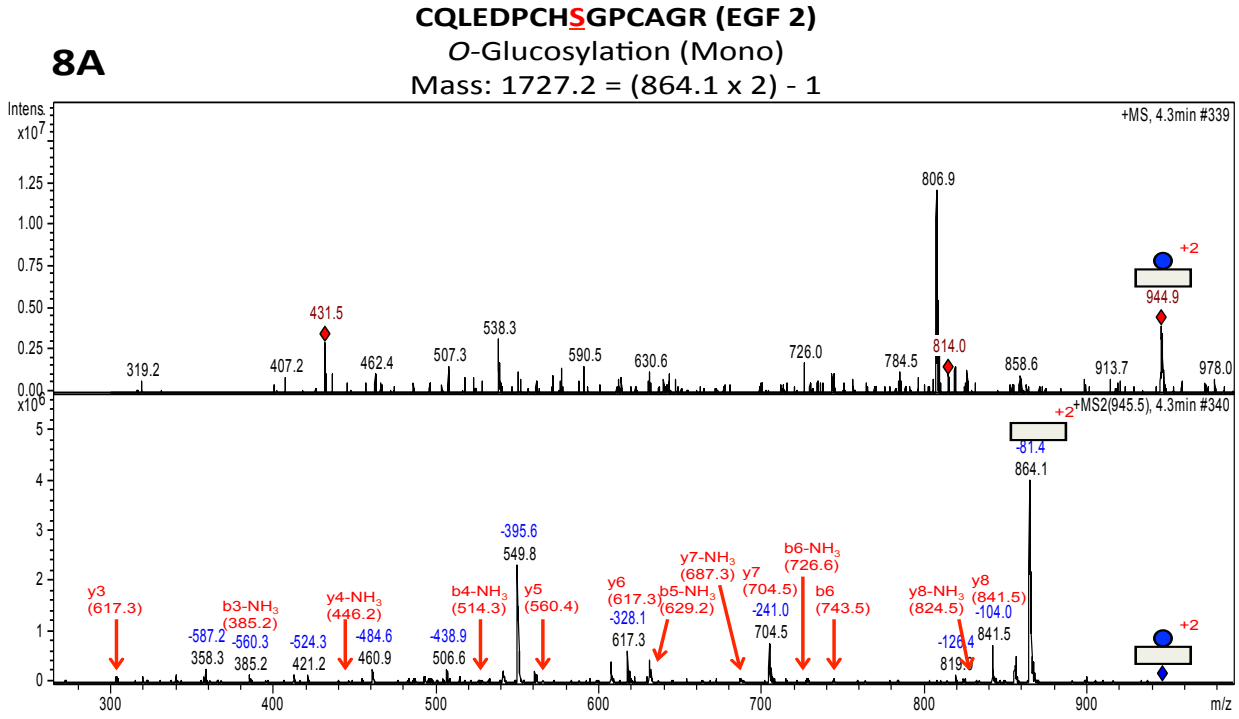


7B

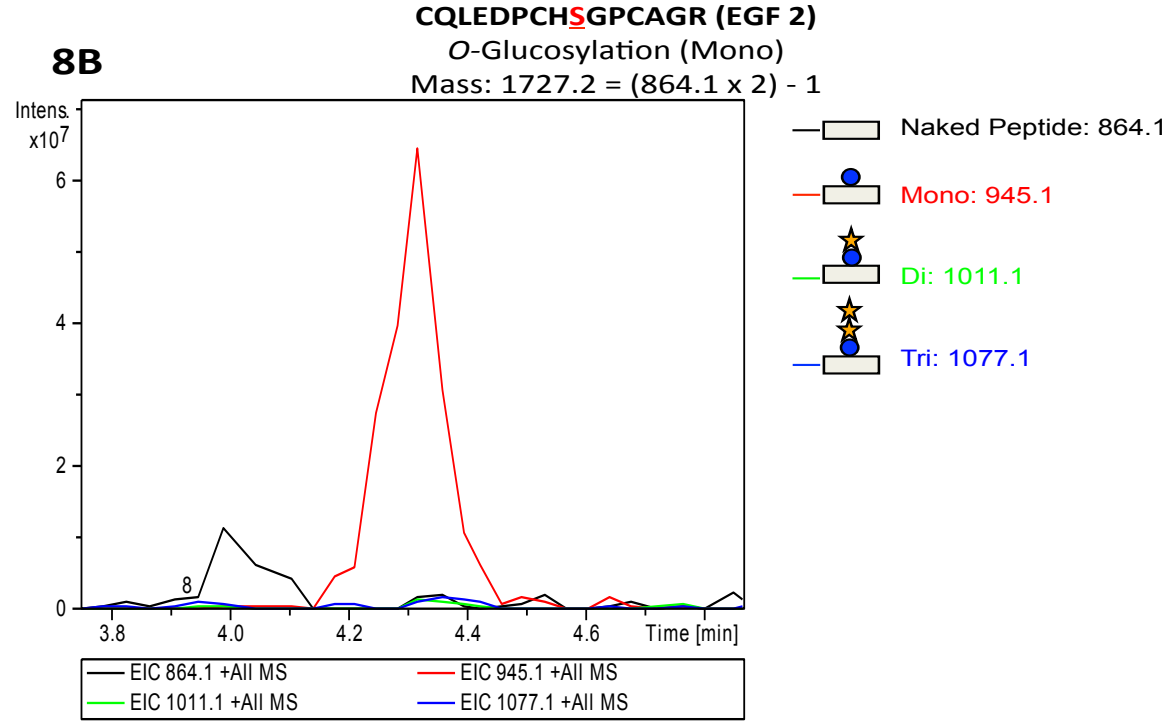
CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Tri)
Mass: 1744.6 = (582.2 x 3) - 2



8A

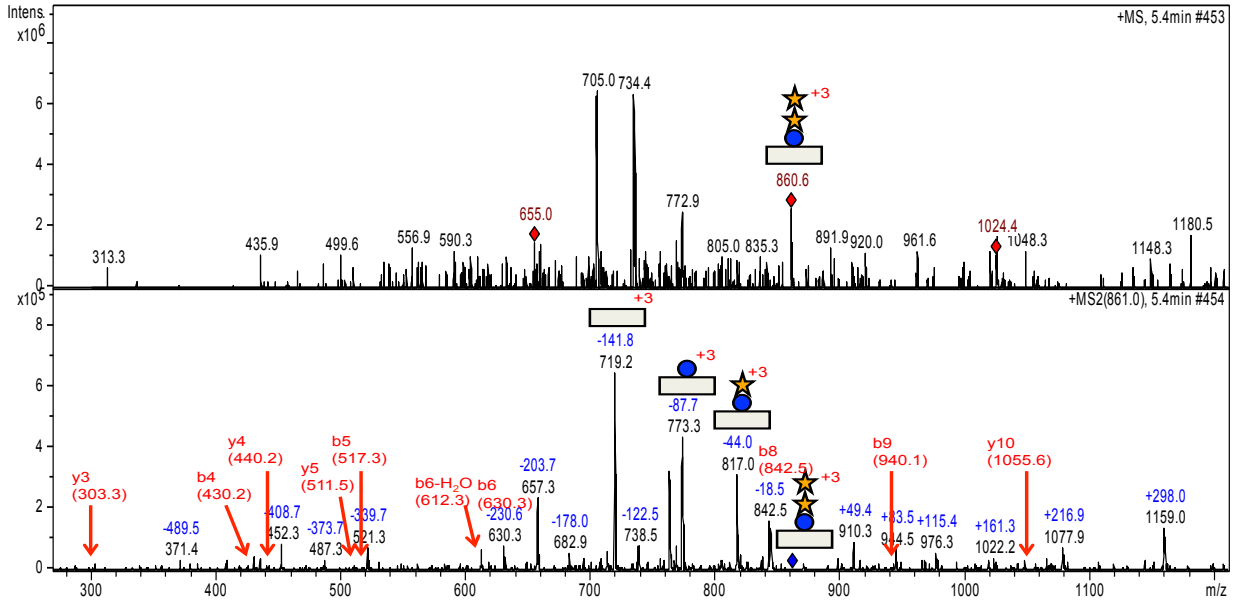

CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Mono)
Mass: 1727.2 = (864.1 x 2) - 1
+MS, 4.3min #339
+MS2(945.5), 4.3min #340
Intens.
m/z

8B


CQLEDPCHSGPCAGR (EGF 2)
O-Glucosylation (Mono)
Mass: 1727.2 = (864.1 x 2) - 1
Intens.
Time [min]
Naked Peptide: 864.1
Mono: 945.1
Di: 1011.1
Tri: 1077.1
EIC 864.1 +All MS
EIC 945.1 +All MS
EIC 1011.1 +All MS
EIC 1077.1 +All MS

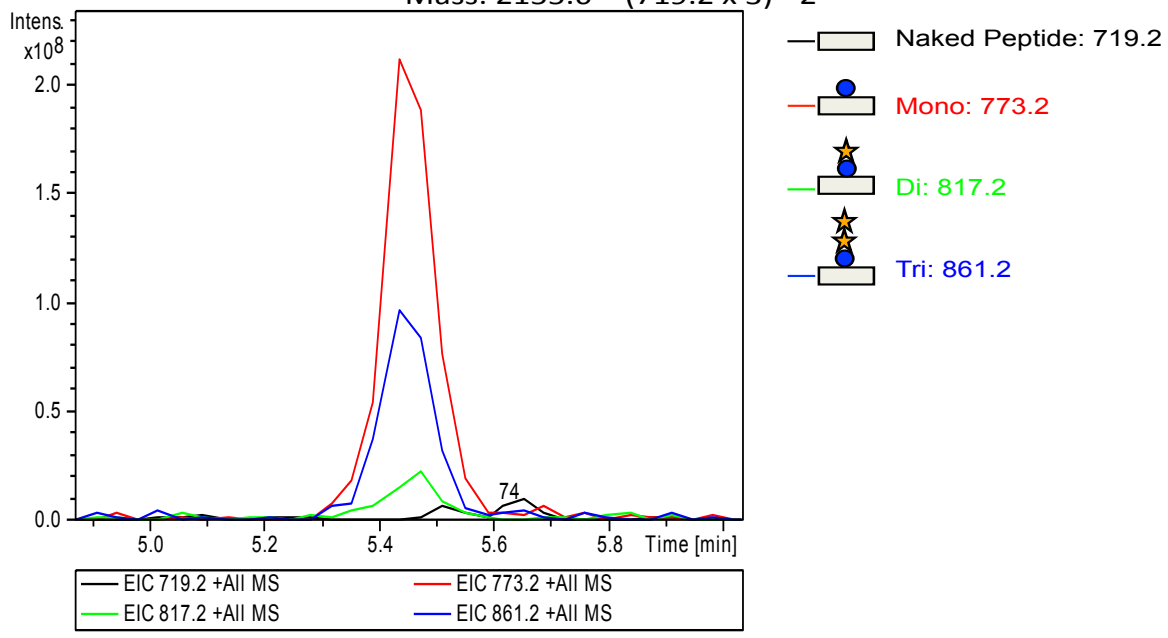
9A

GPDCSLPDPCLSSPCAHGAR (EGF 3)
O-Glucosylation (Tri)
Mass: 2155.6 = (719.2 x 3) - 2



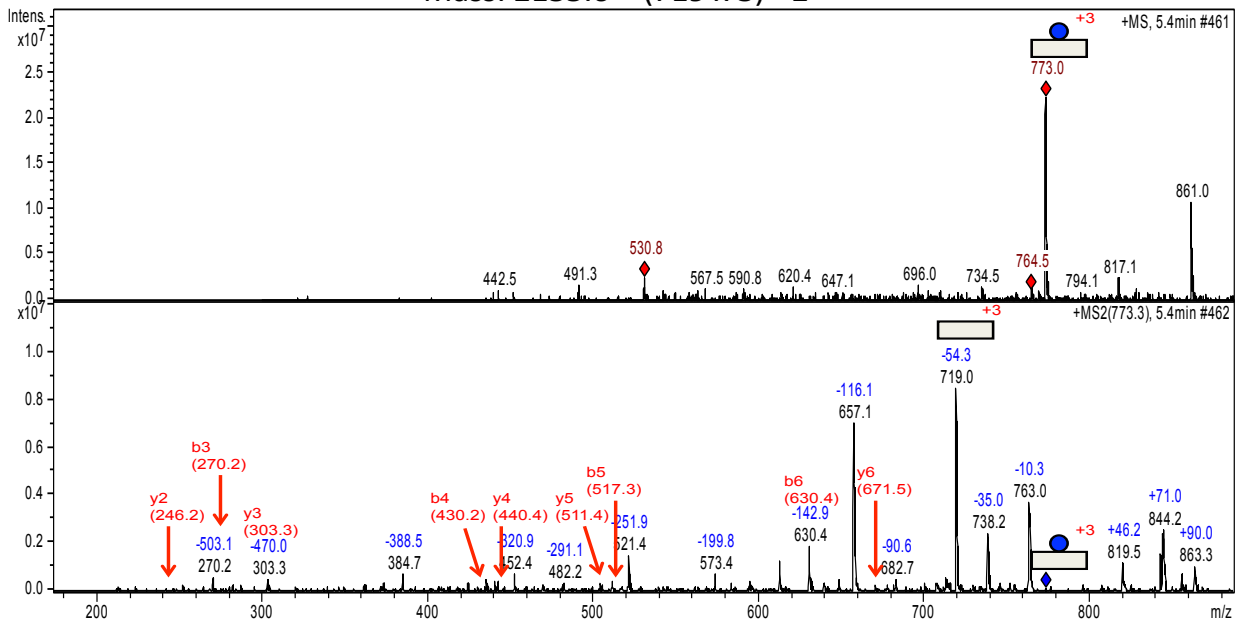
9B

GPDCSLPDPCLSSPCAHGAR (EGF 3)
O-Glucosylation (Tri)
Mass: 2155.6 = (719.2 x 3) - 2



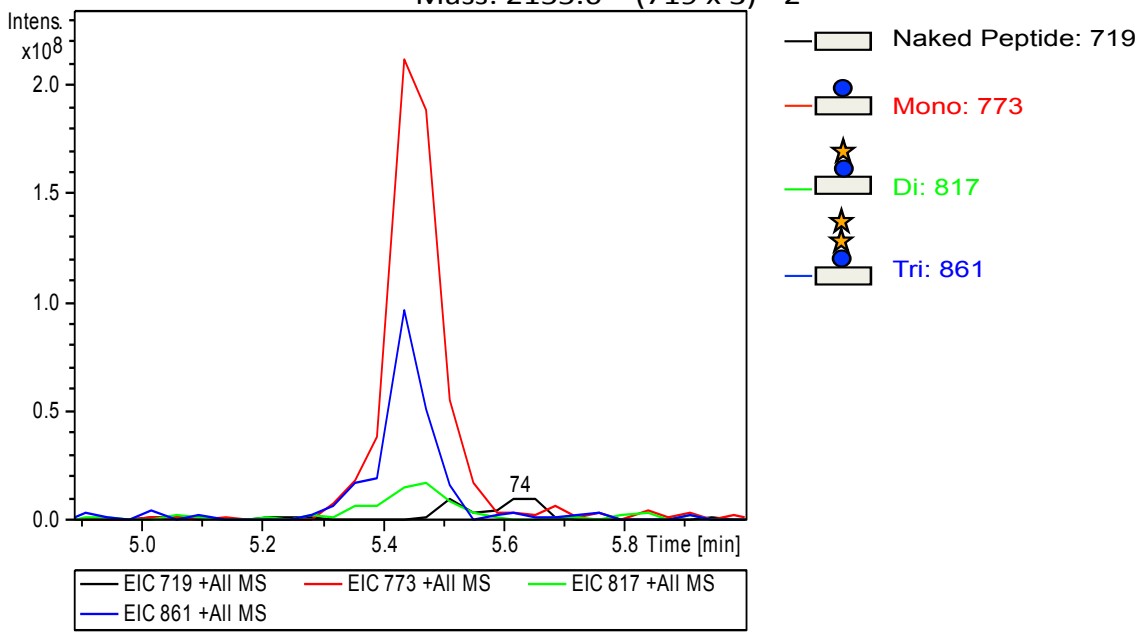
10A

GPDCSLPDPCLSSPCAHGAR (EGF 3)
O-Glucosylation (Mono)
Mass: 2155.6 = (719 x 3) - 2



10B

GPDCSLPDPCLSSPCAHGAR (EGF 3)
O-Glucosylation (Mono)
Mass: 2155.6 = (719 x 3) - 2

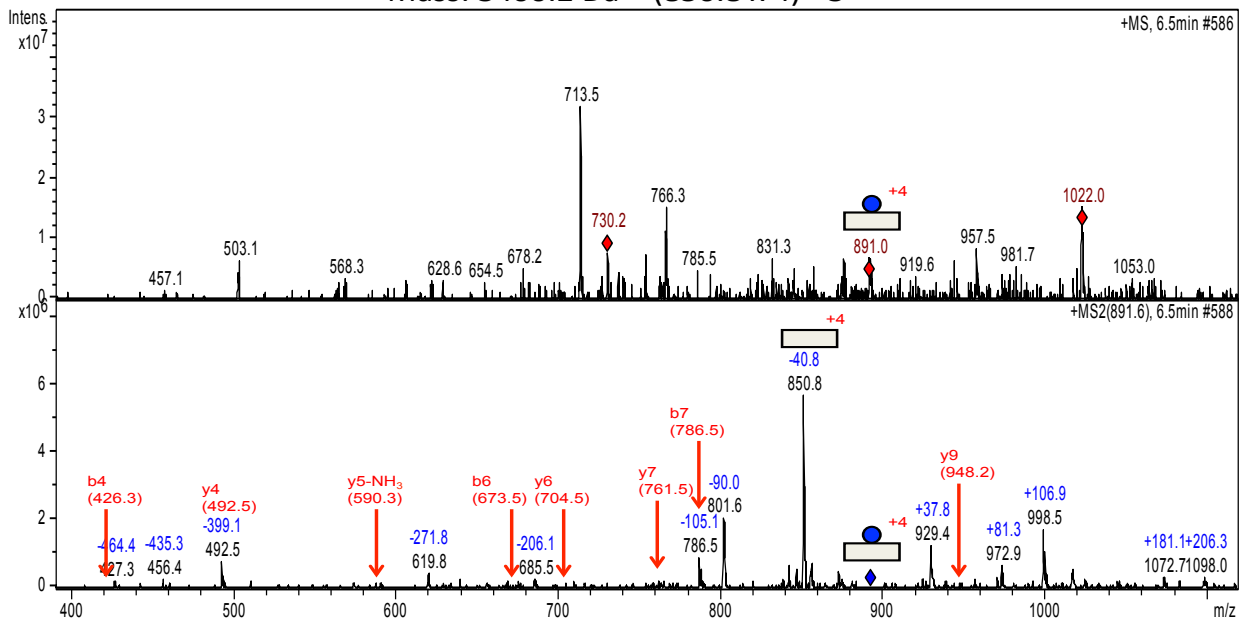


11A

RGPDCSLPDPCLSSPCAHGARCSVGPDGRFL (EGF 3)

O-Glucosylation (Mono)

Mass: 3400.2 Da = (850.8 x 4) - 3

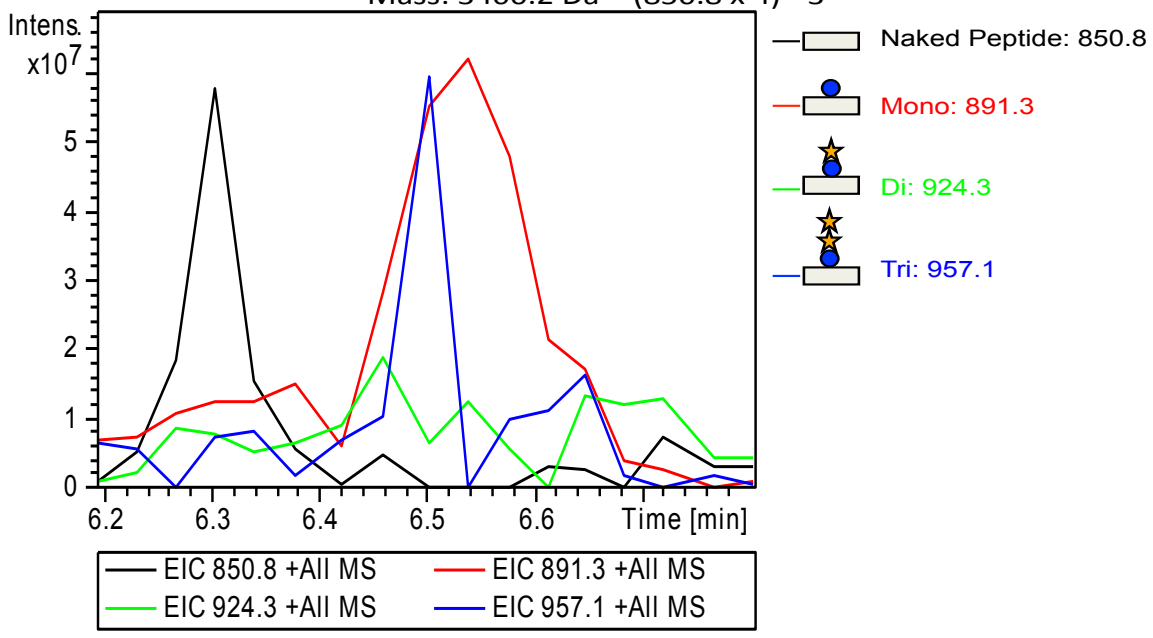


11B

RGPDCSLDPCLSSPCAHGARCSVGPDGRFL (EGF 3)

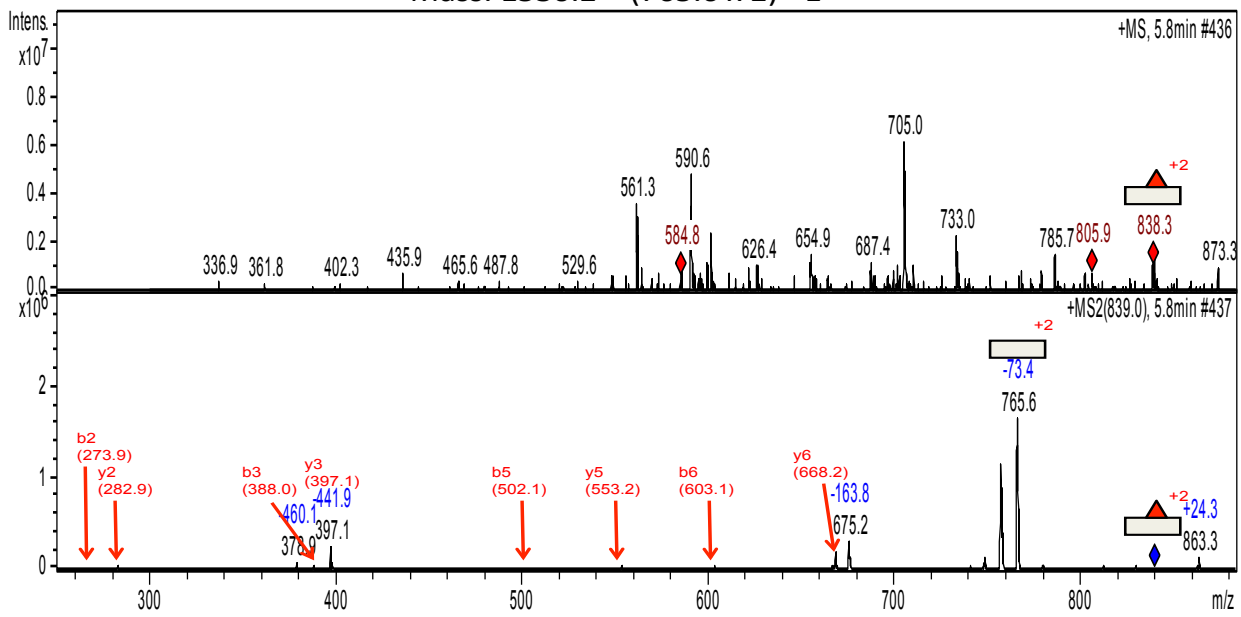
O-Glucosylation (Mono)

Mass: 3400.2 Da = (850.8 x 4) - 3



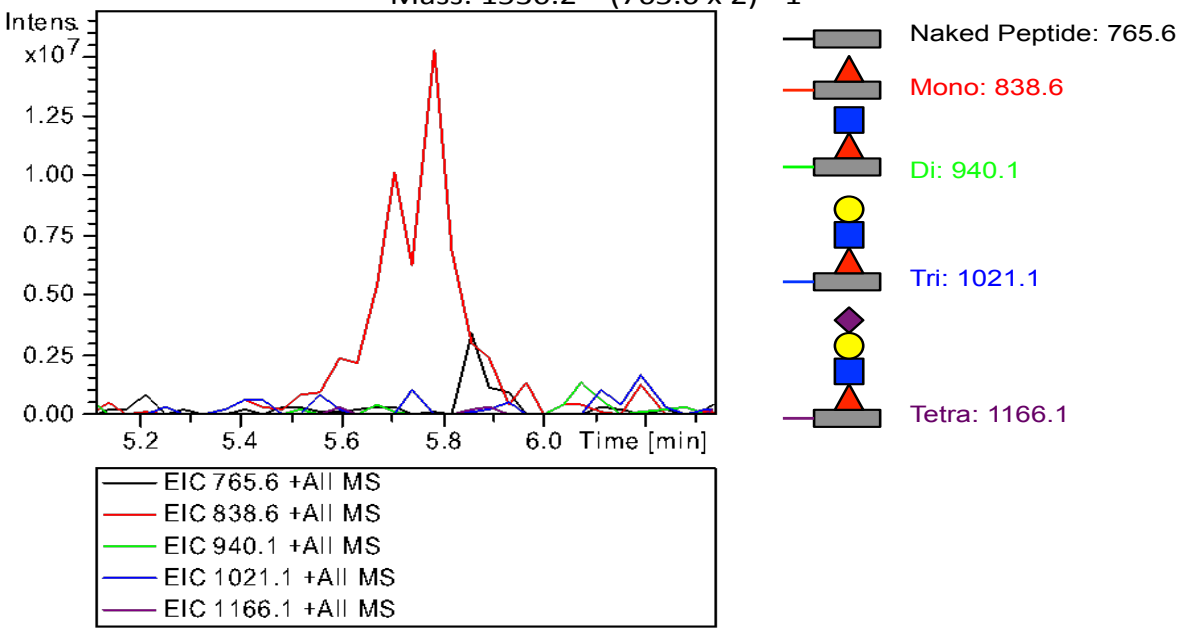
12A

CLNGGTCVDGVNTY (EGF 6)
O-Fucosylation (Mono)
Mass: 1530.2 = (765.6 x 2) - 1



12B

CLNGGTCVDGVNTY (EGF 6)
O-Fucosylation (Mono)
Mass: 1530.2 = (765.6 x 2) - 1

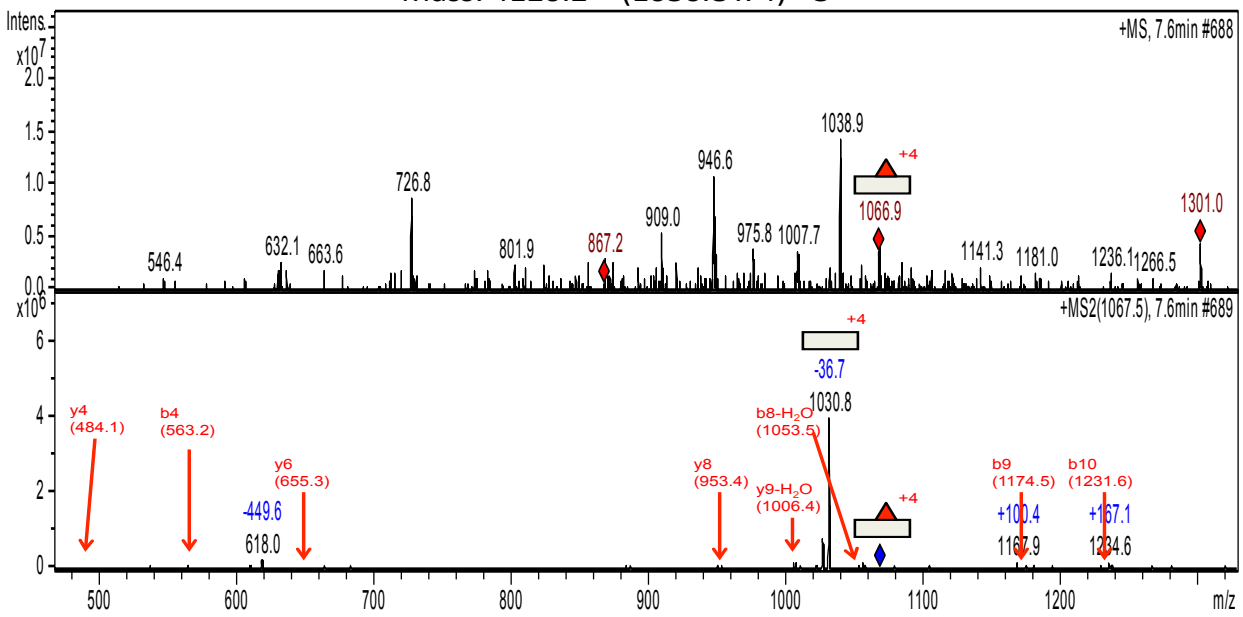


NCQCPPEWTGQFCTEDVDECQLQPNACHNGGTCF (EGF 7)

13A

O-Fucosylation (Mono)

Mass: 4120.2 = (1030.8 x 4) - 3

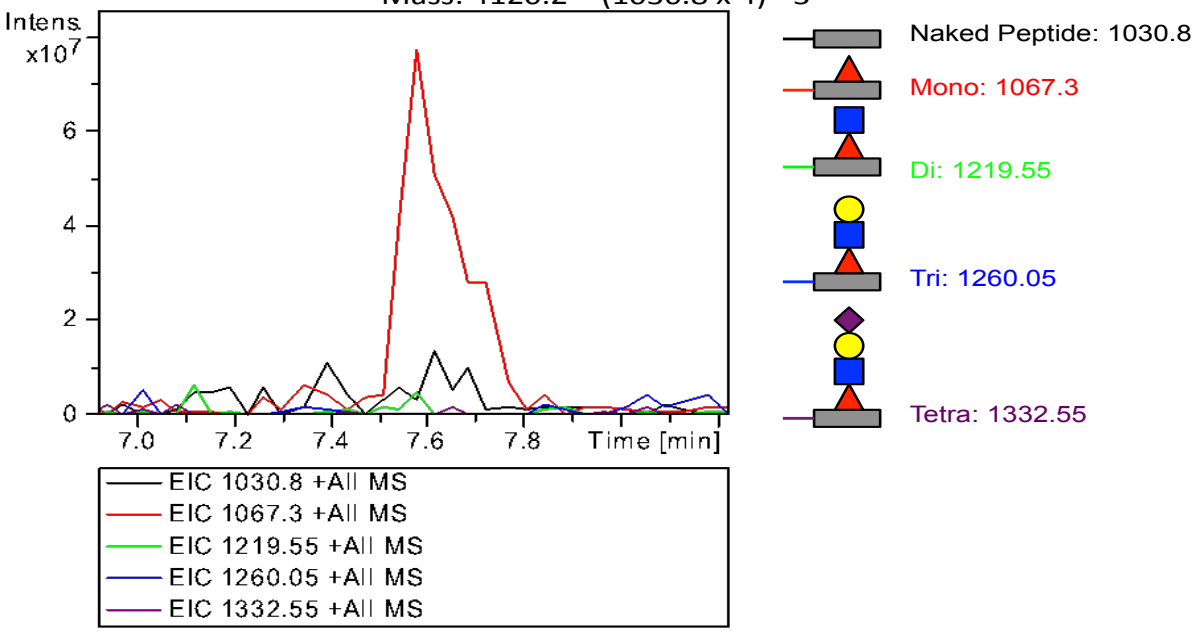


NCQCPPEWTGQFCTEDVDECQLQPNACHNGGTCF (EGF 7)

13B

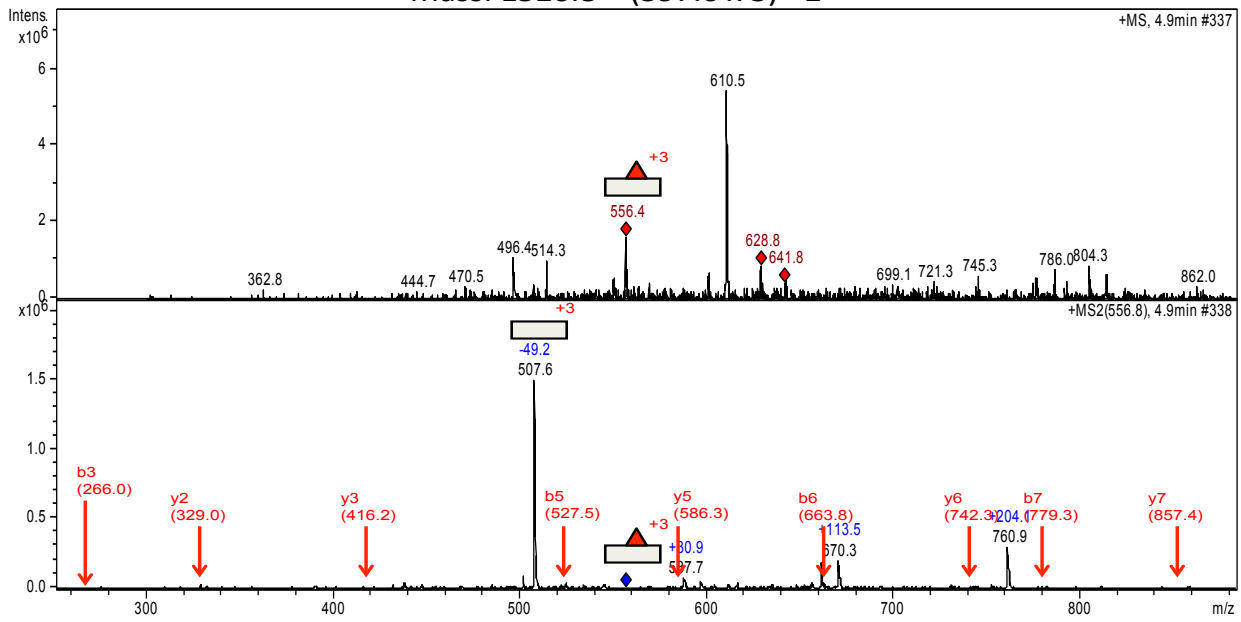
O-Fucosylation (Mono)

Mass: 4120.2 = (1030.8 x 4) - 3



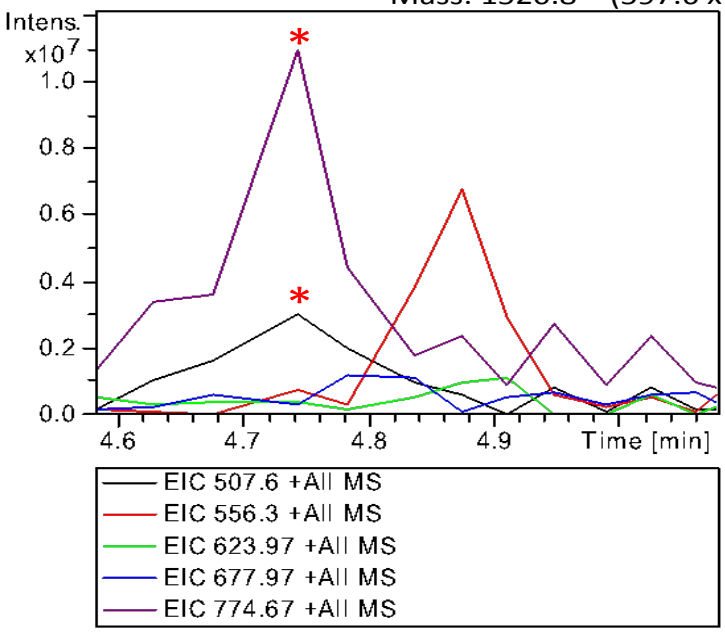
14A

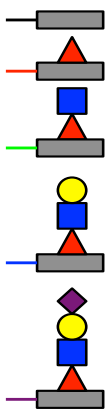
HGATCHDRVASFY (EGF 8)
O-Fucosylation (Mono)
Mass: 1520.8 = (597.6 x 3) - 2



14B

HGATCHDRVASFY (EGF 8)
O-Fucosylation (Mono)
Mass: 1520.8 = (597.6 x 3) - 2





Naked Peptide: 507.6

Mono: 556.3

Di: 623.97

Tri: 677.97

Tetra: 774.67

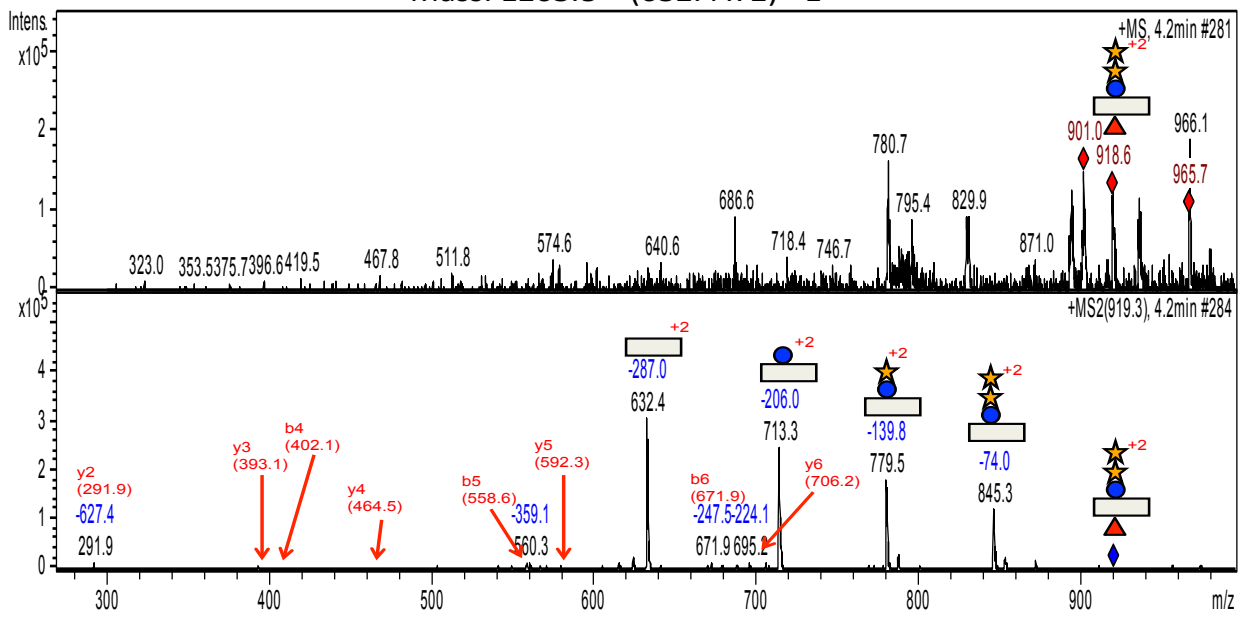
*Peaks are not part of the corresponding peptide

15A

SGPCRNQATCL (EGF 11)

O-Fucosylation (Mono) & *O*-Glucosylation (Tri)

Mass: 1263.8 = (632.4 x 2) - 1

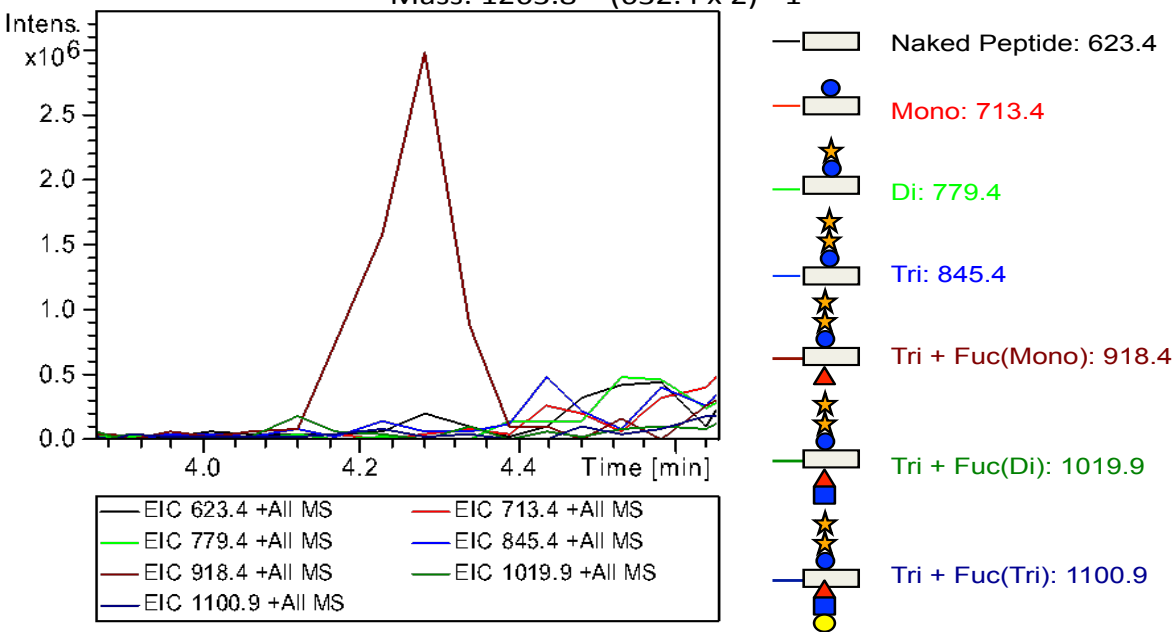


15B

SGPCRNQATCL (EGF 11)

O-Fucosylation (Mono) & *O*-Glucosylation (Tri)

Mass: 1263.8 = (632.4 x 2) - 1

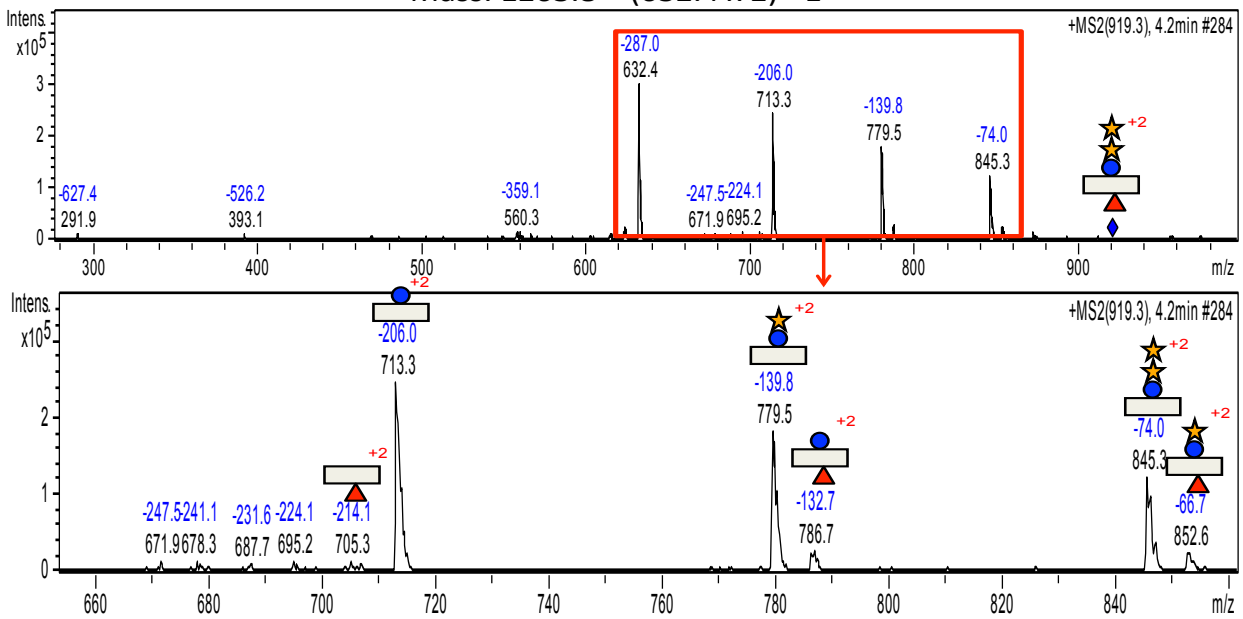


15C

SGPCRNQATCL (EGF 11)

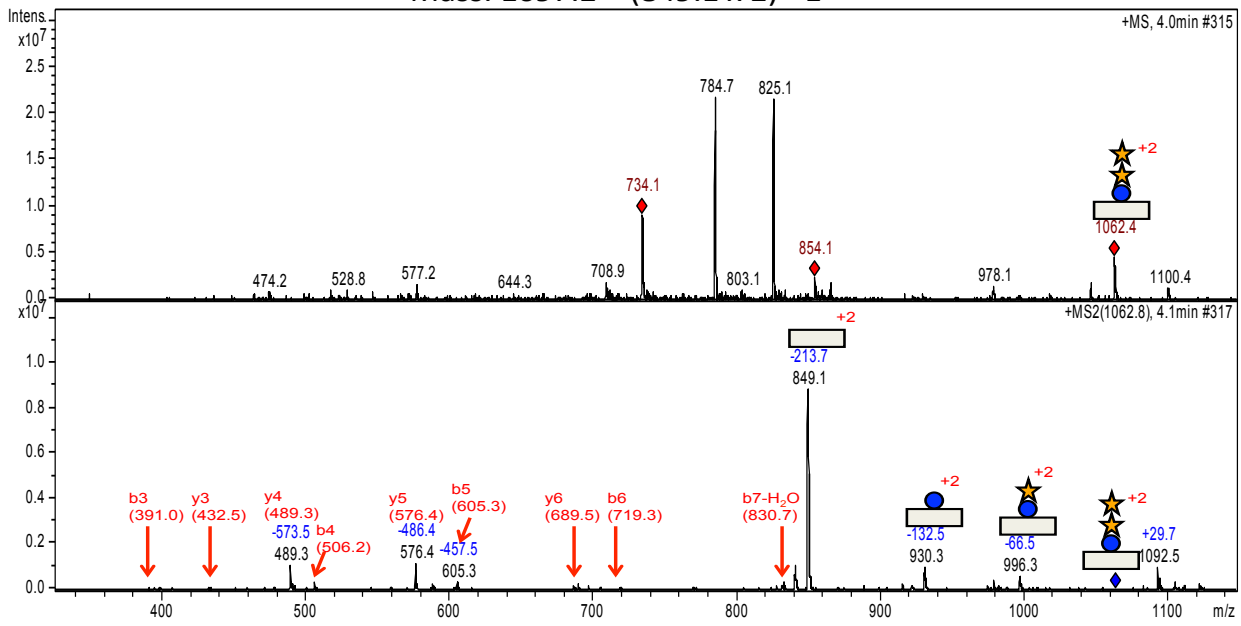
O-Fucosylation (Mono) & *O*-Glucosylation (Tri)

Mass: 1263.8 = (632.4 x 2) - 1



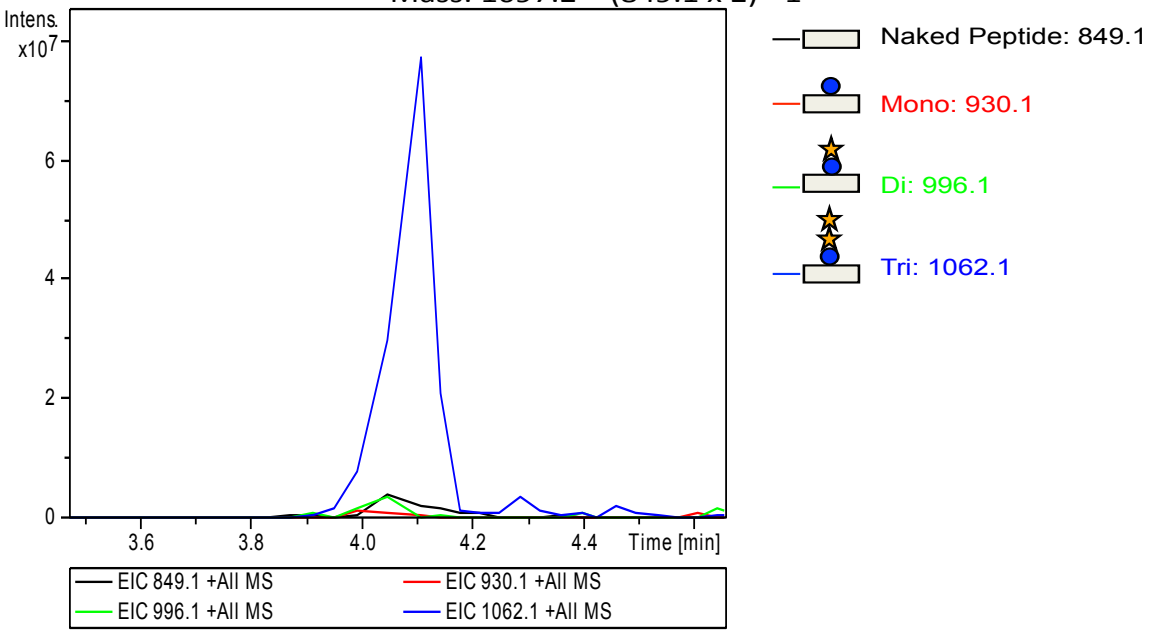
16A

CETDVNECLSGPCR (EGF 11)
O-Glucosylation (Tri)
Mass: 1697.2 = (849.1 x 2) - 1



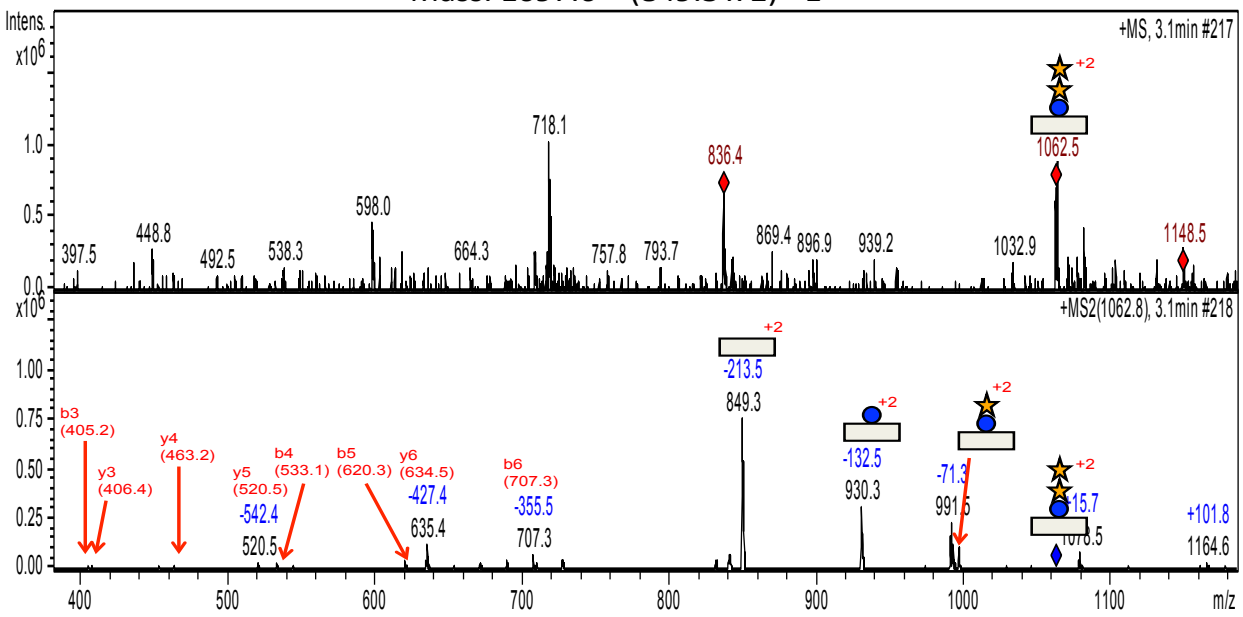
16B

CETDVNECLSGPCR (EGF 11)
O-Glucosylation (Tri)
Mass: 1697.2 = (849.1 x 2) - 1



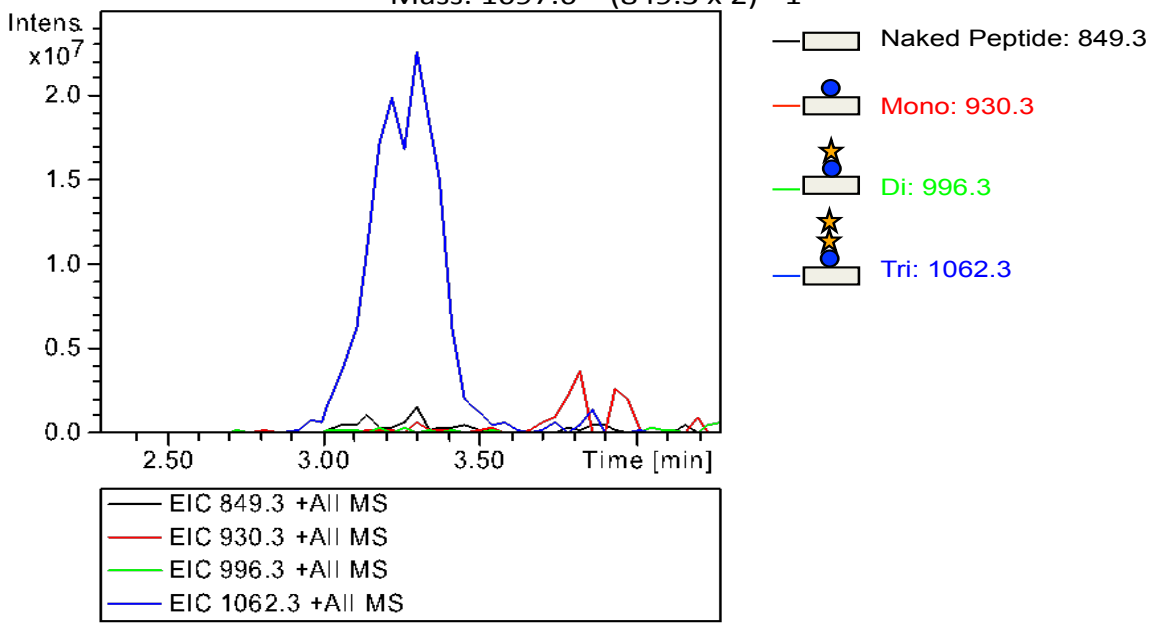
17A

DECQSSPCVNGGVCK (EGF 12)
O-Glucosylation (Tri)
Mass: 1697.6 = (849.3 x 2) - 1



17B

DECQSSPCVNGGVCK (EGF 12)
O-Glucosylation (Tri)
Mass: 1697.6 = (849.3 x 2) - 1

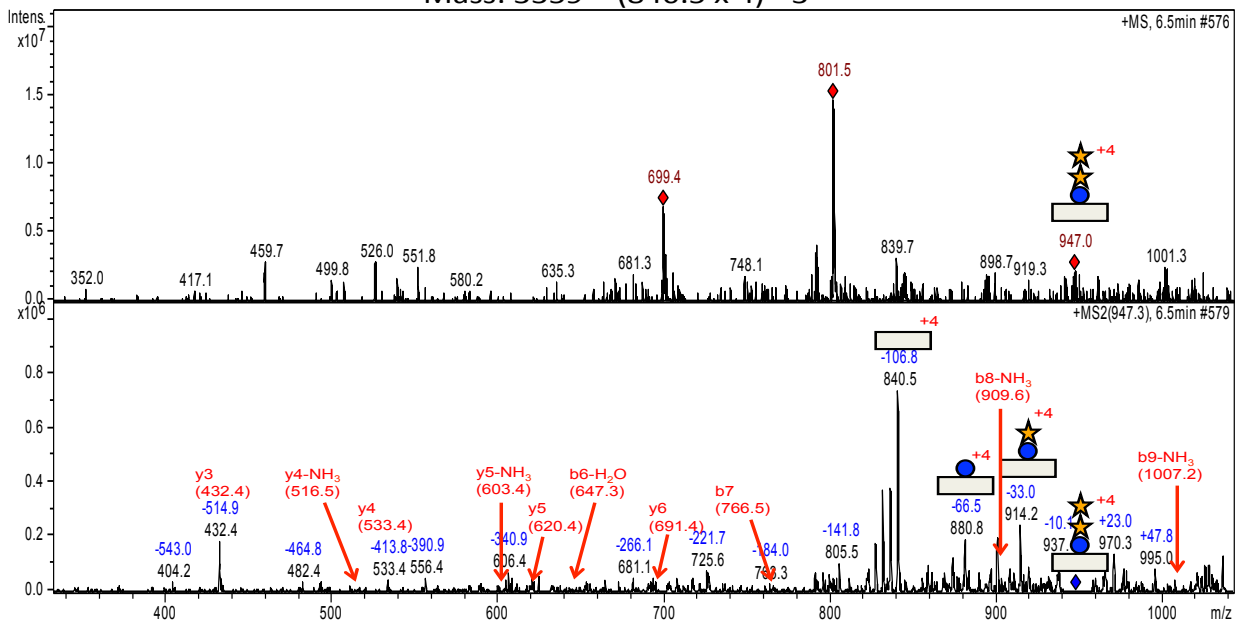


18A

VNGFSCTCPSGFSGSTCQLDVDECASTPCR (EGF 13)

O-Glucosylation (Tri)

Mass: 3359 = (840.5 x 4) - 3

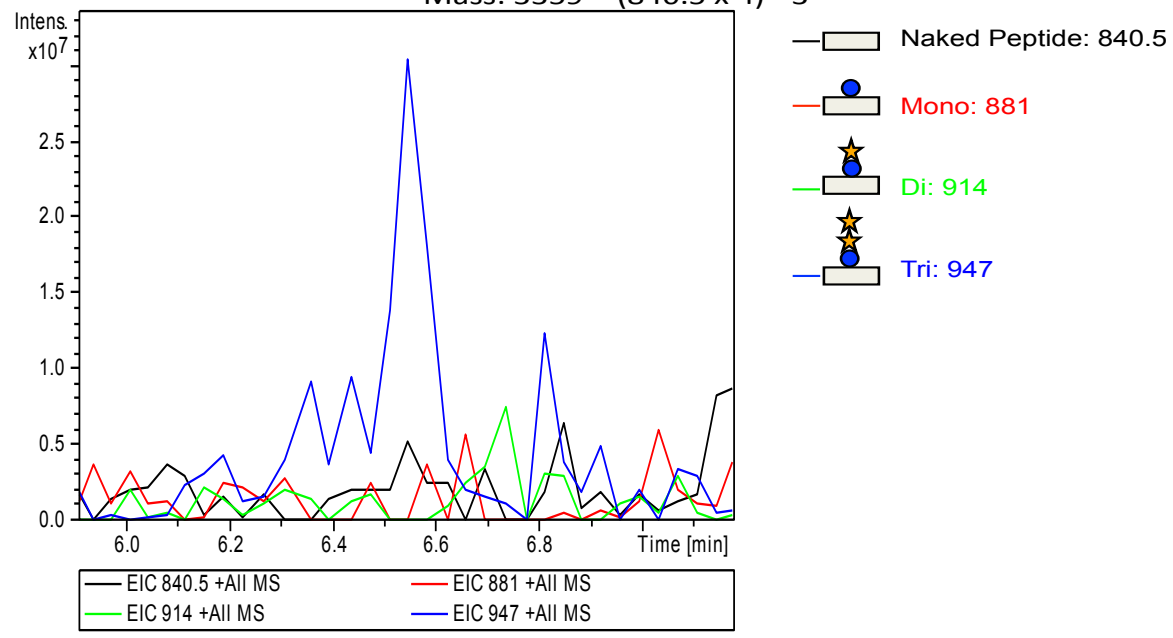


VNGFSCTCPSGFSGSTCQLDVDECASTPCR (EGF 13)

18B

O-Glucosylation (Tri)

Mass: 3359 = (840.5 x 4) - 3

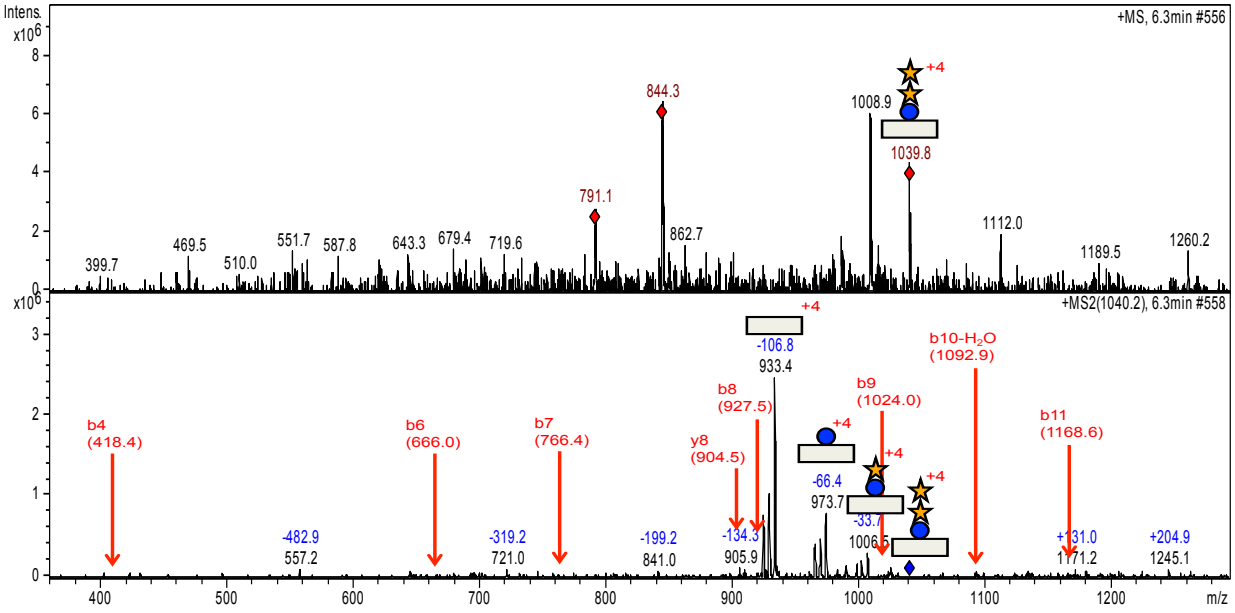


19A

VNGFSCTCPSGFSGSTCQLDVDECASTPCRNGAK (EGF 13)

O-Glucosylation (Tri)

Mass: 3730.6 Da = (933.4 x 4) - 3

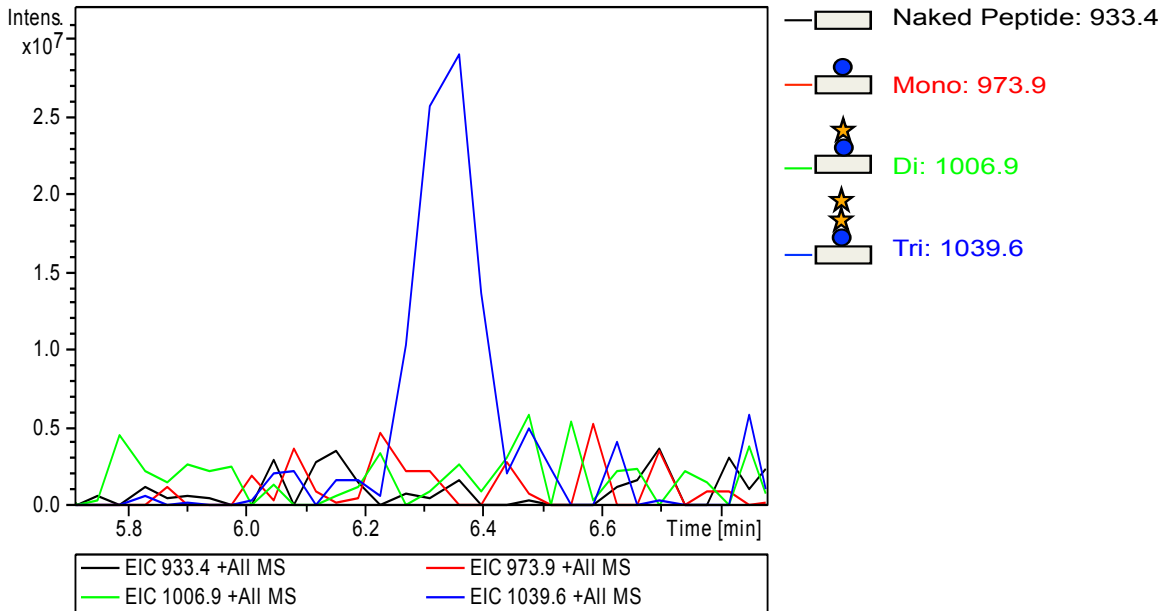


19B

VNGFSCTCPSGFSGSTCQLDVDECASTPCRNGAK (EGF 13)

O-Glucosylation (Tri)

Mass: 3730.6 Da = (933.4 x 4) - 3

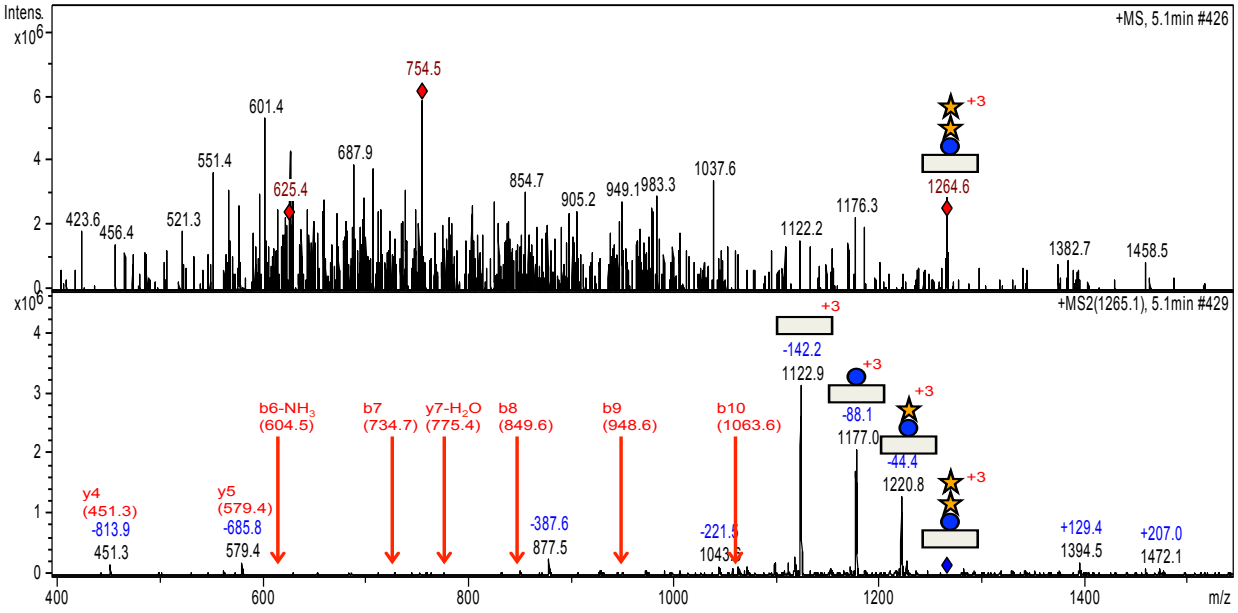


20A

SGSTCQLDVDECASTPCRNGAKCVDQPDGY (EGF 13)

O-Glucosylation (Tri)

Mass: 3366.7 Da = (1122.9 x 3) - 2

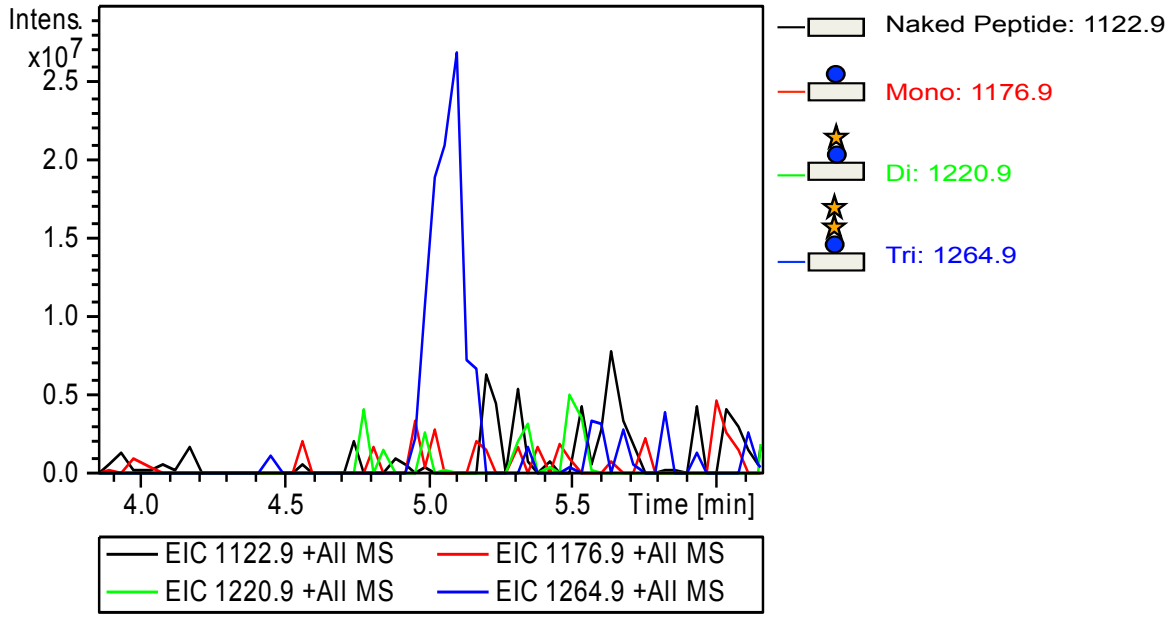


20B

SGSTCQLDVDECASTPCRNGAKCVDQPDGY (EGF 13)

O-Glucosylation (Tri)

Mass: 3366.7 Da = (1122.9 x 3) - 2

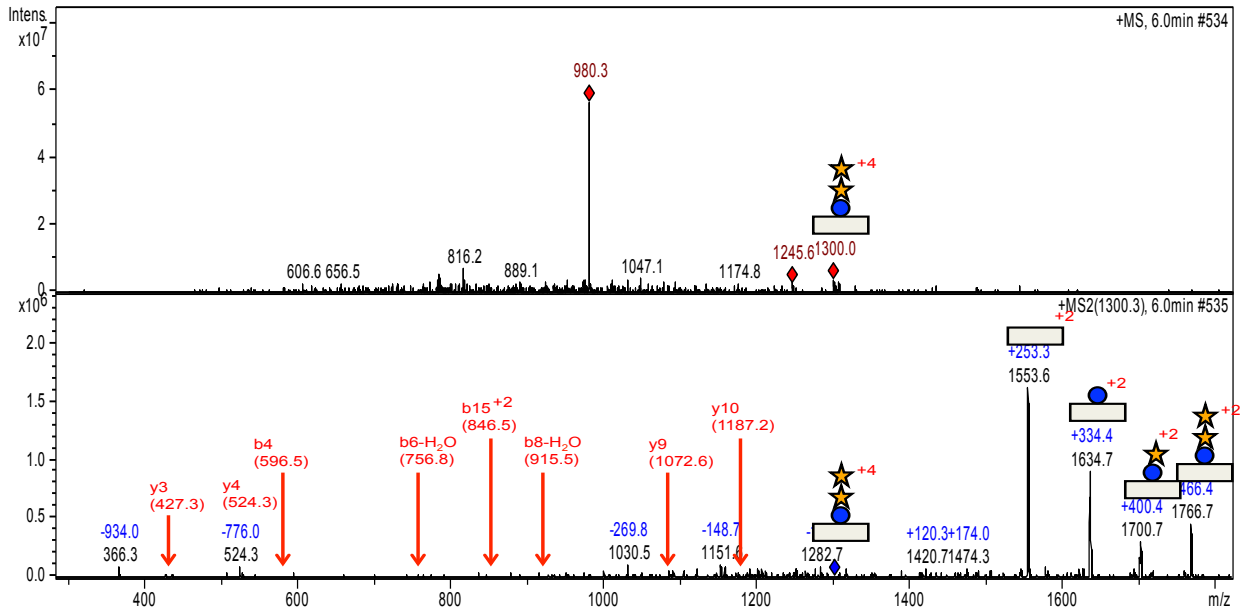


21A

LCRCPSGTTGVNCEVNIDDCASNPCTF (EGF 16)

O-Glucosylation (Tri)

Mass: 3106.2 Da = (1553.6 x 2) – 1

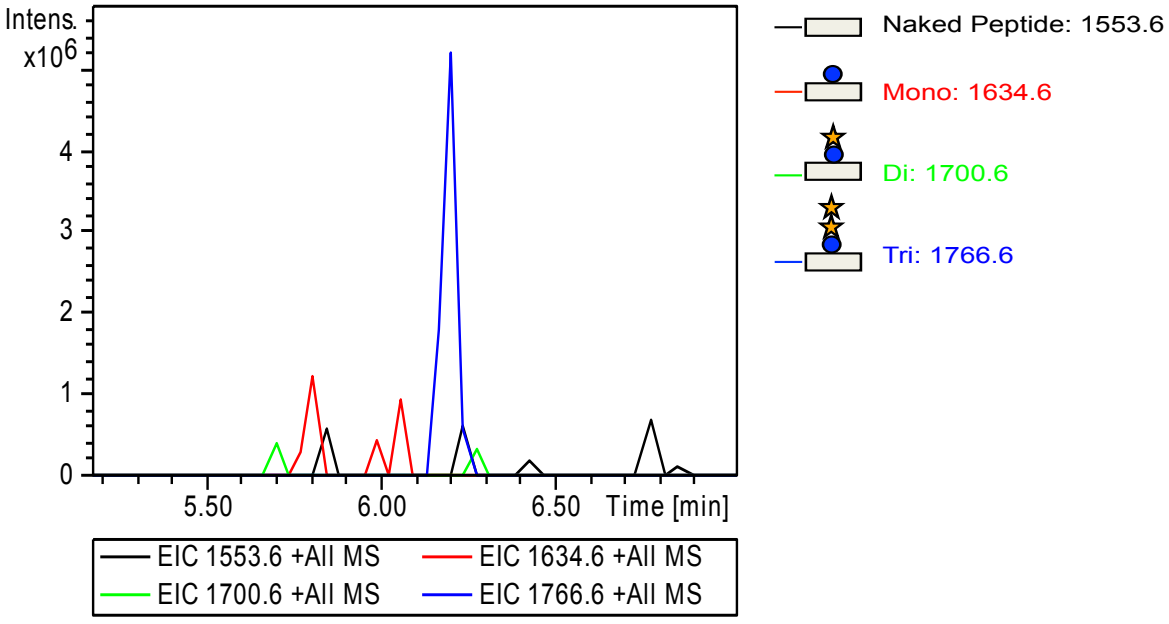


21B

LCRCPSGTTGVNCEVNIDDCASNPCTF (EGF 16)

O-Glucosylation (Tri)

Mass: 3106.2 Da = (1553.6 x 2) – 1

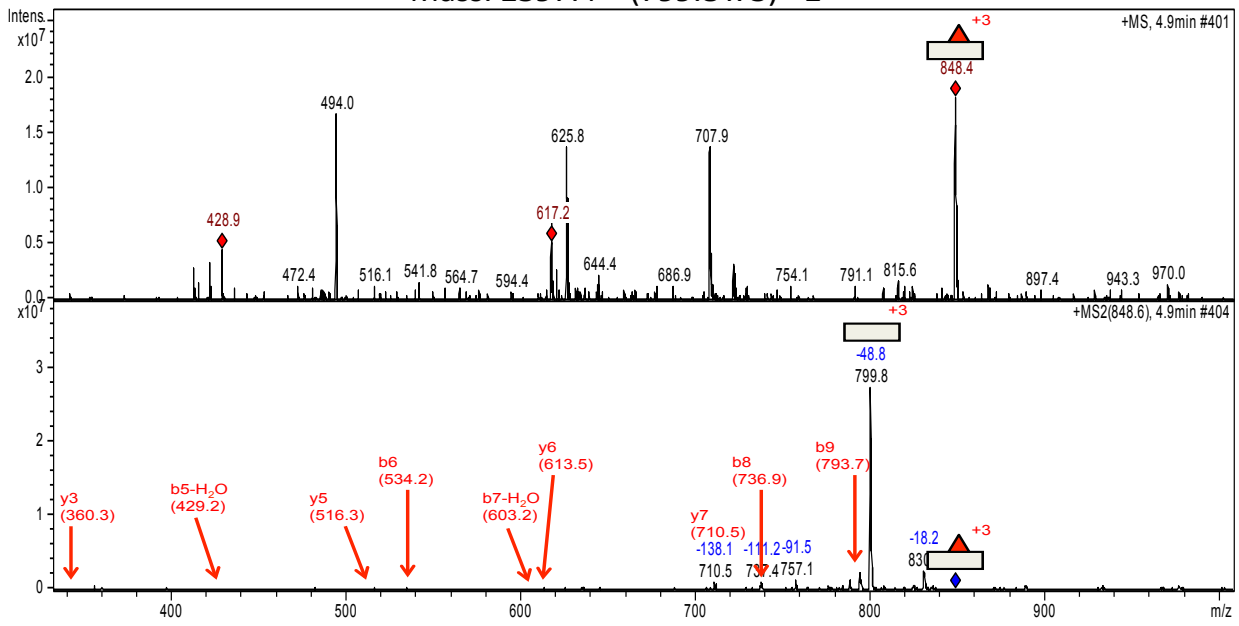


22A

AGGTCSSDGMGFHCTCPPGVQGR (EGF 19)

O-Fucosylation (Mono)

Mass: 2397.4 = (799.8 x 3) - 2

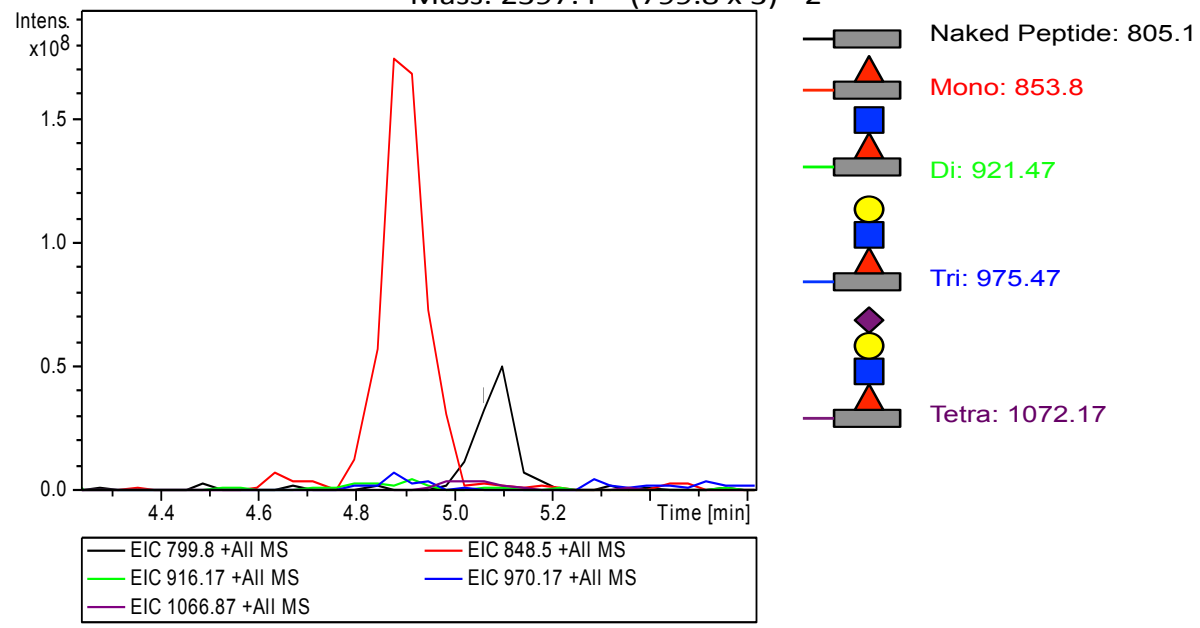


22B

AGGTCSSDGMGFHCTCPPGVQGR (EGF 19)

O-Fucosylation (Mono)

Mass: 2397.4 = (799.8 x 3) - 2

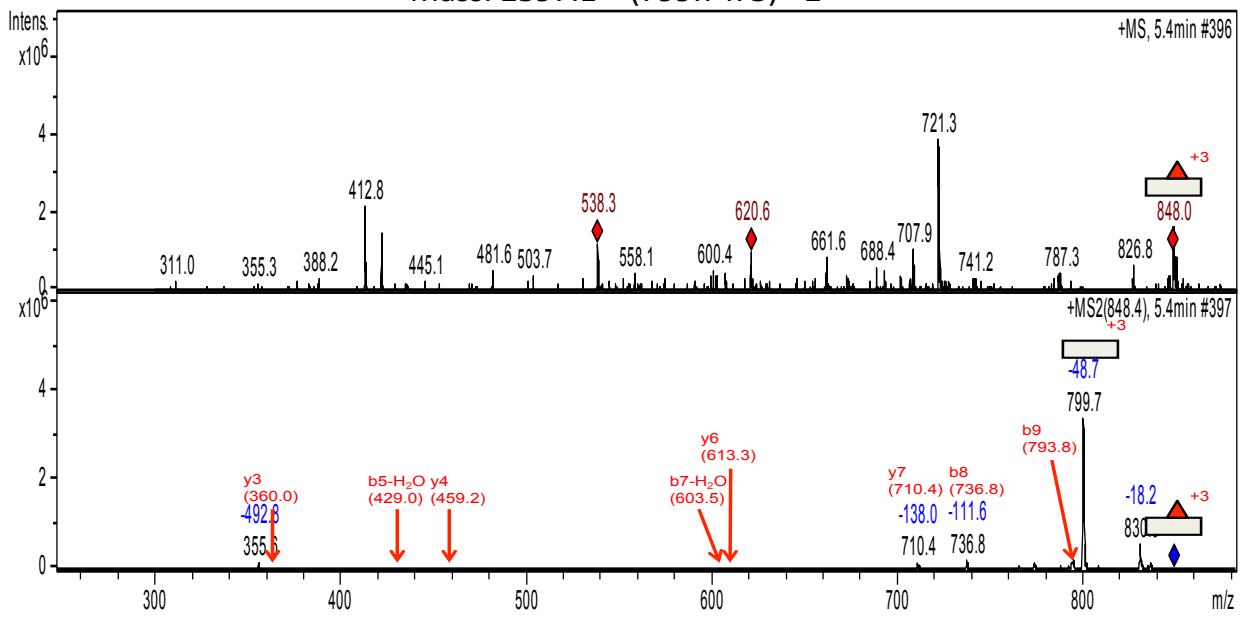


AGGTCSSDGMGFHCTCPPGVQGR (EGF 19)

23A

O-Fucosylation (Mono)

Mass: 2397.1 = (799.7 x 3) - 2

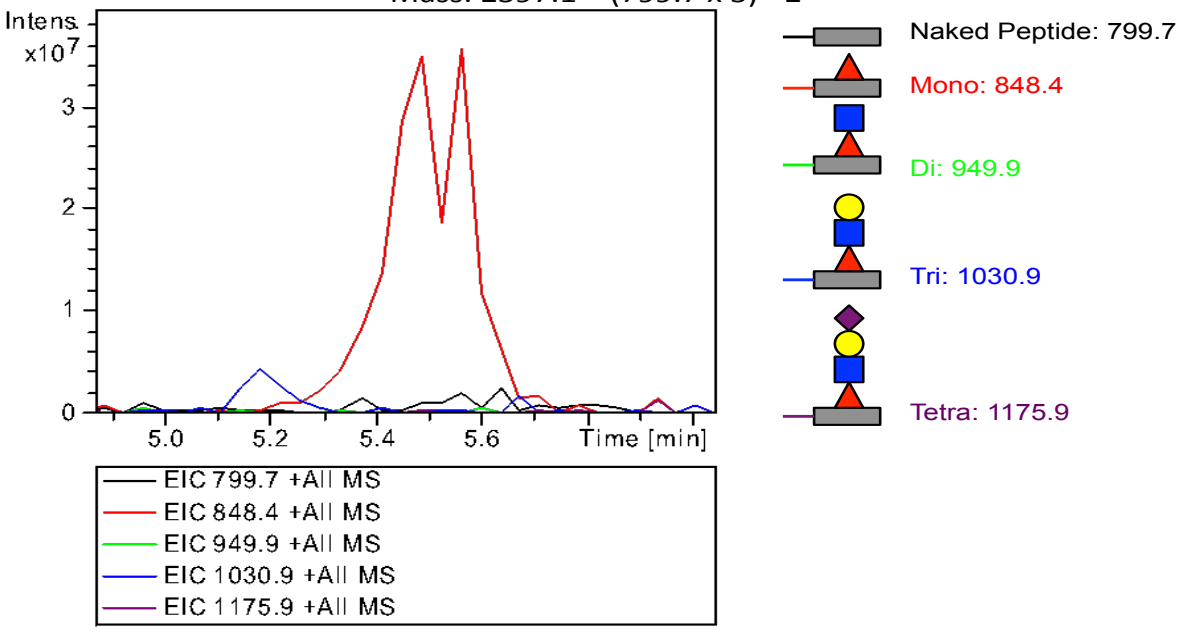


AGGTCSSDGMGFHCTCPPGVQGR (EGF 19)

23B

O-Fucosylation (Mono)

Mass: 2397.1 = (799.7 x 3) - 2

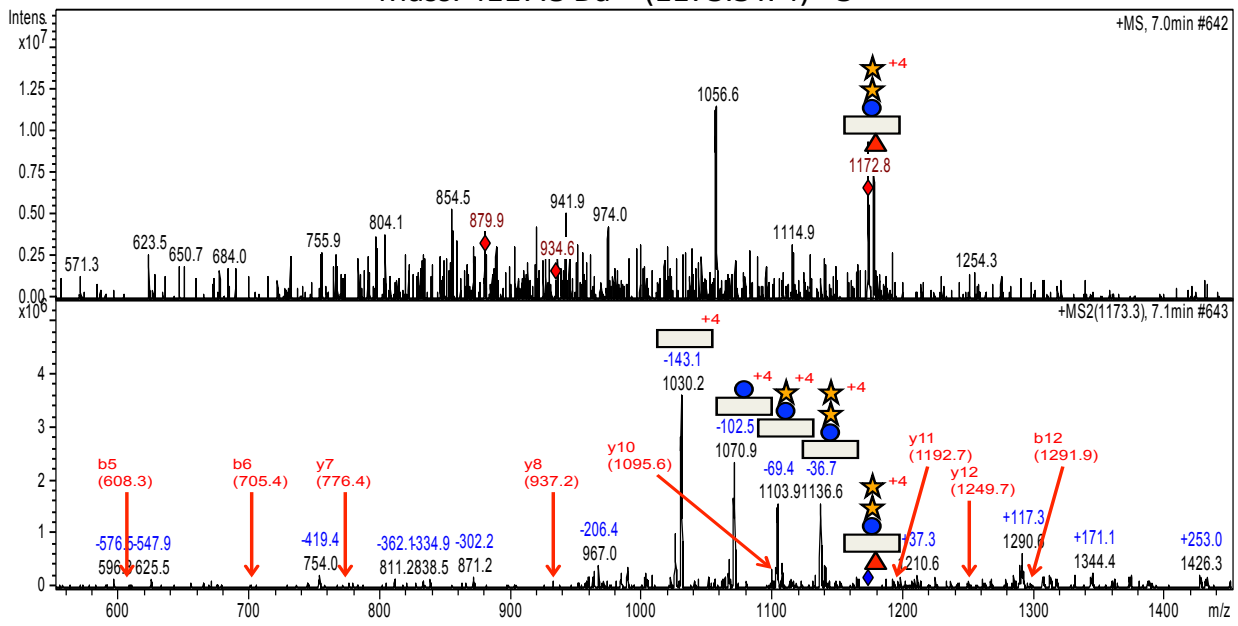


SCSCLPGFAGPRCARDVDECLSNPCGPGTCTDHVASF (EGF 23)

24A

O-Glucosylation (Tri) & *O*-Fucosylation (Mono)

Mass: 4117.8 Da = (1173.3 x 4) - 3

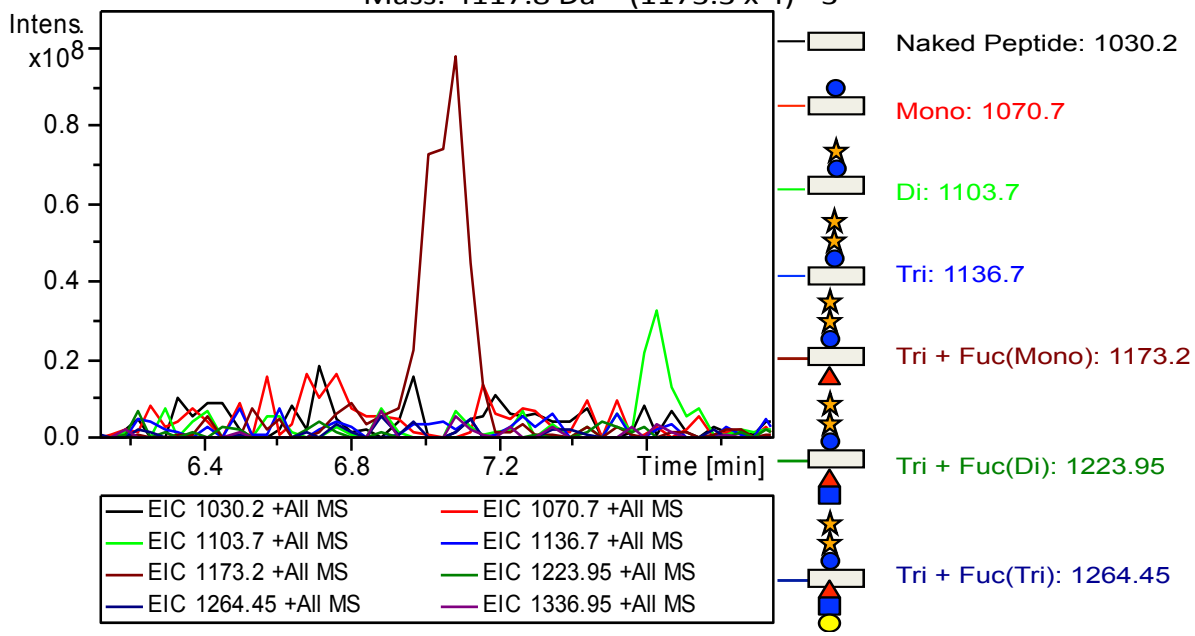


SCSCLPGFAGPRCARDVDECLSNPCGPGTCTDHVASF (EGF 23)

24B

O-Glucosylation (Tri) & *O*-Fucosylation (Mono)

Mass: 4117.8 Da = (1173.3 x 4) - 3

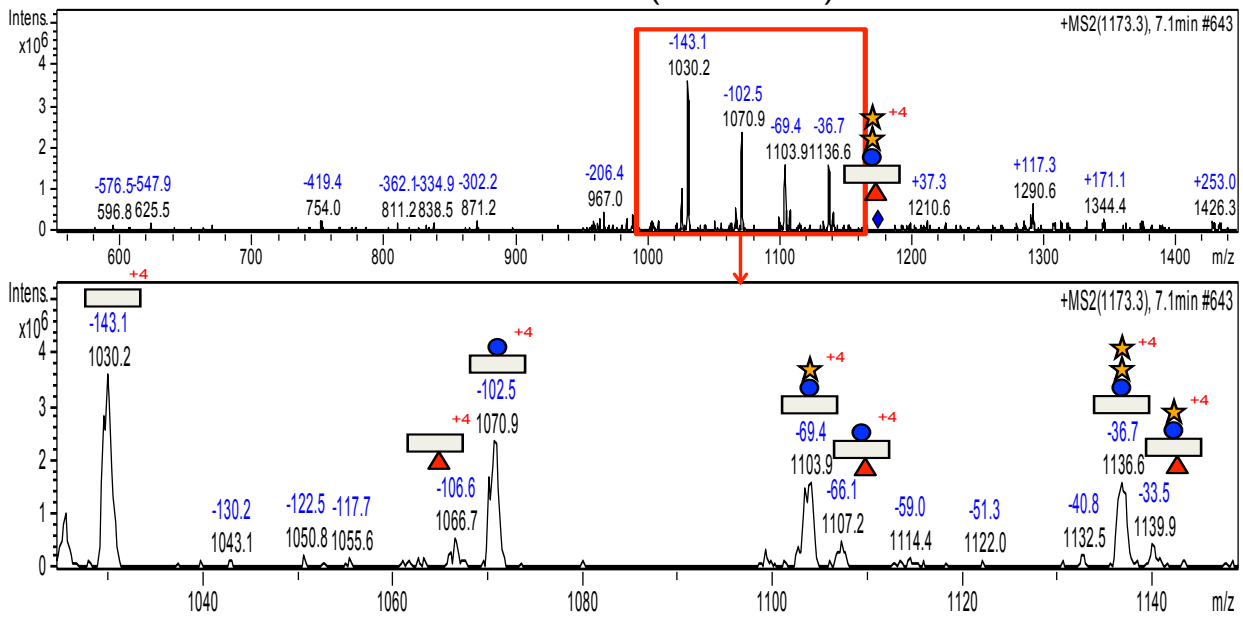


SCSCLPGFAGPRCARDVDECLSNPCGPGTCTDHVASF (EGF 23)

24C

O-Glucosylation (Tri) & *O*-Fucosylation (Mono)

Mass: 4117.8 Da = (1173.3 x 4) - 3

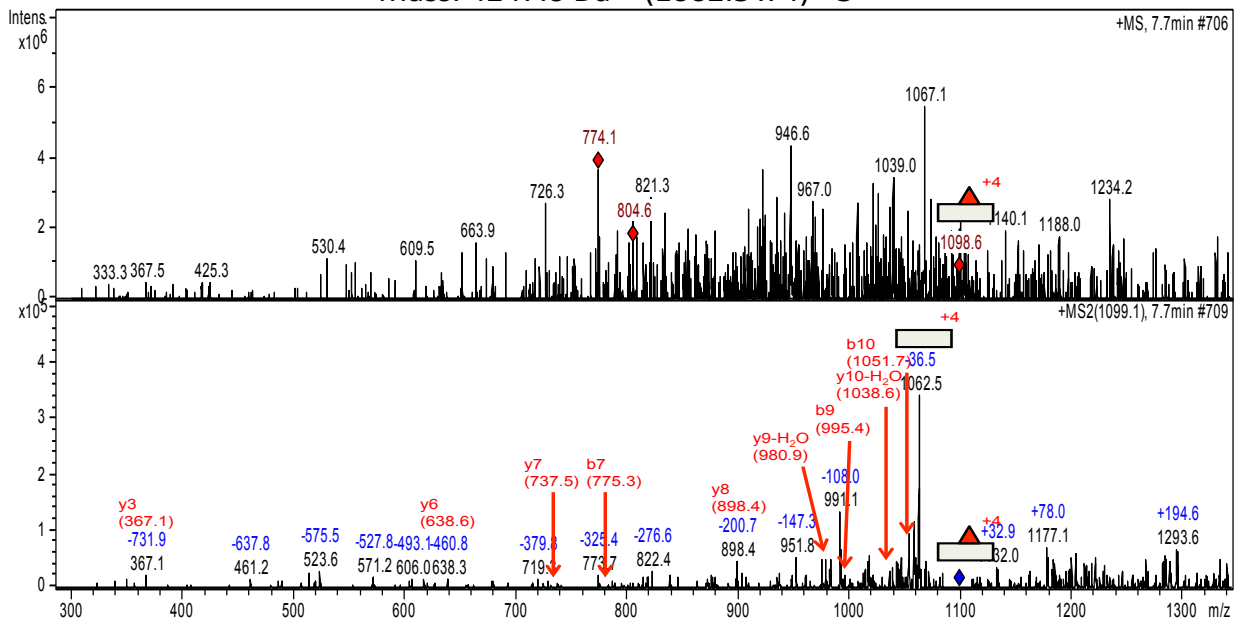


TCTCPPGYGGFHCEQDLPDCSPSSCFNGGTCVDGVNSF (EGF 24)

25A

O-Fucosylation (Mono)

Mass: 4247.0 Da = (1062.5 x 4) - 3

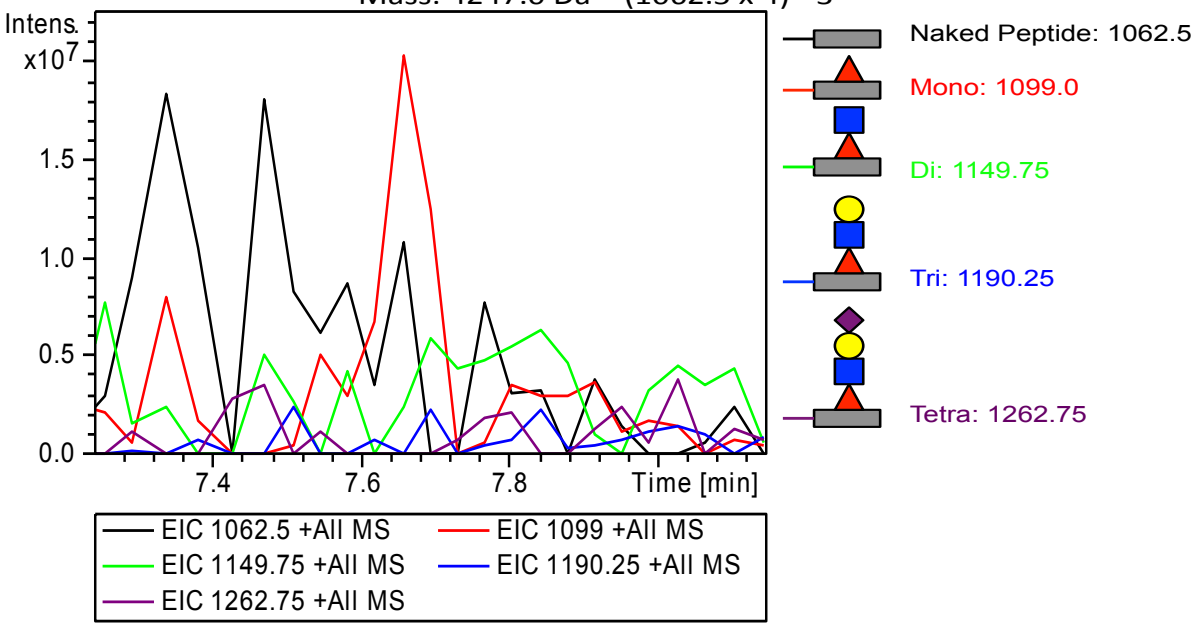


TCTCPPGYGGFHCEQDLPDCSPSSCFNGGTCVDGVNSG (EGF 24)

25B

O-Fucosylation (Mono)

Mass: 4247.0 Da = (1062.5 x 4) - 3

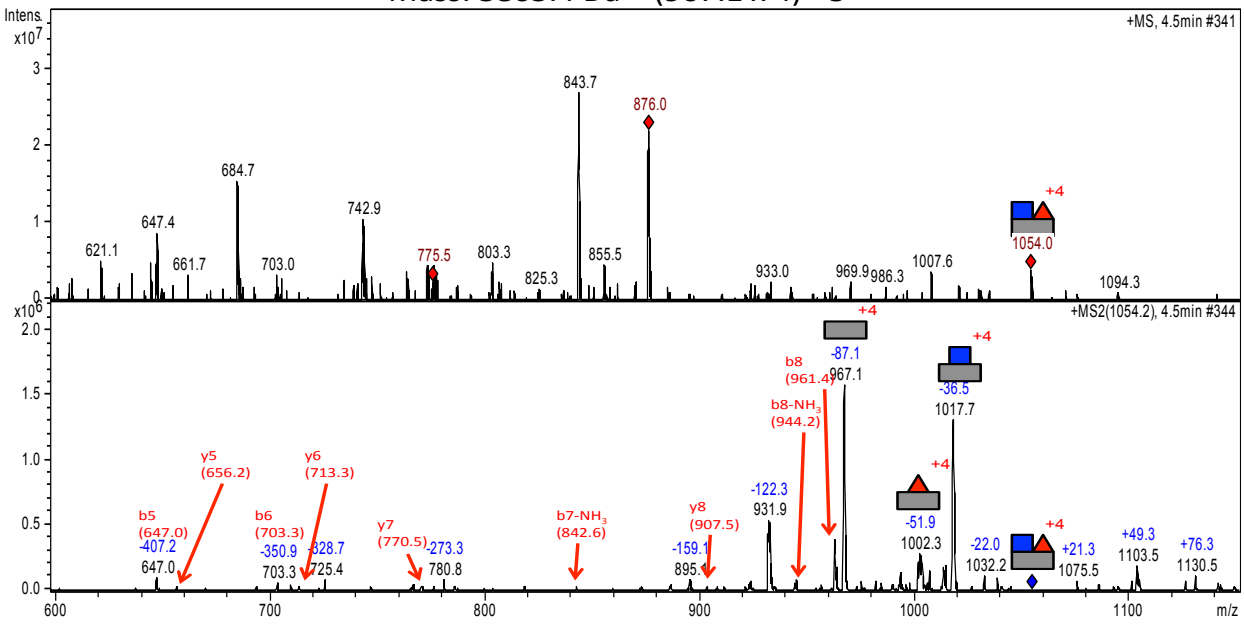


CVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGY (EGF 27 - 28)

26A

O-Fucosylation & O-GlcNAc

Mass: 3865.4 Da = (967.1 x 4) - 3

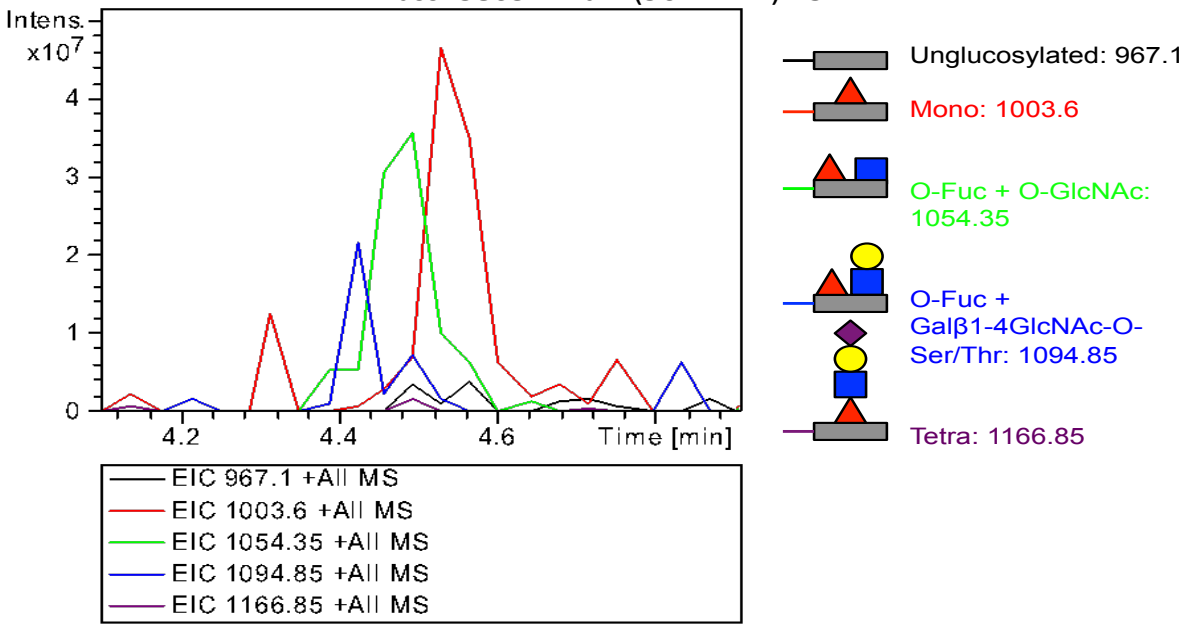


CVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGY (EGF 27 - 28)

26B

O-Fucosylation & O-GlcNAc

Mass: 3865.4 Da = (967.1 x 4) - 3

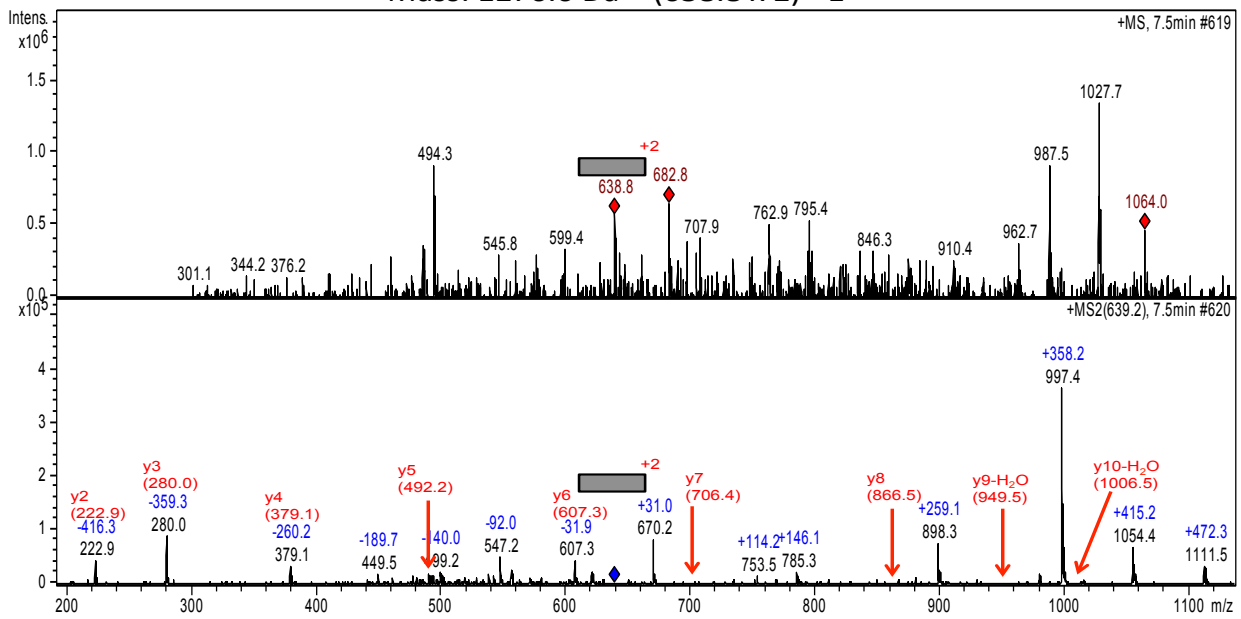


27A

HNGTCVDLVGGF (EGF 30)

O-Fucosylation not found (naked)

Mass: 1276.6 Da = (638.8 x 2) - 1

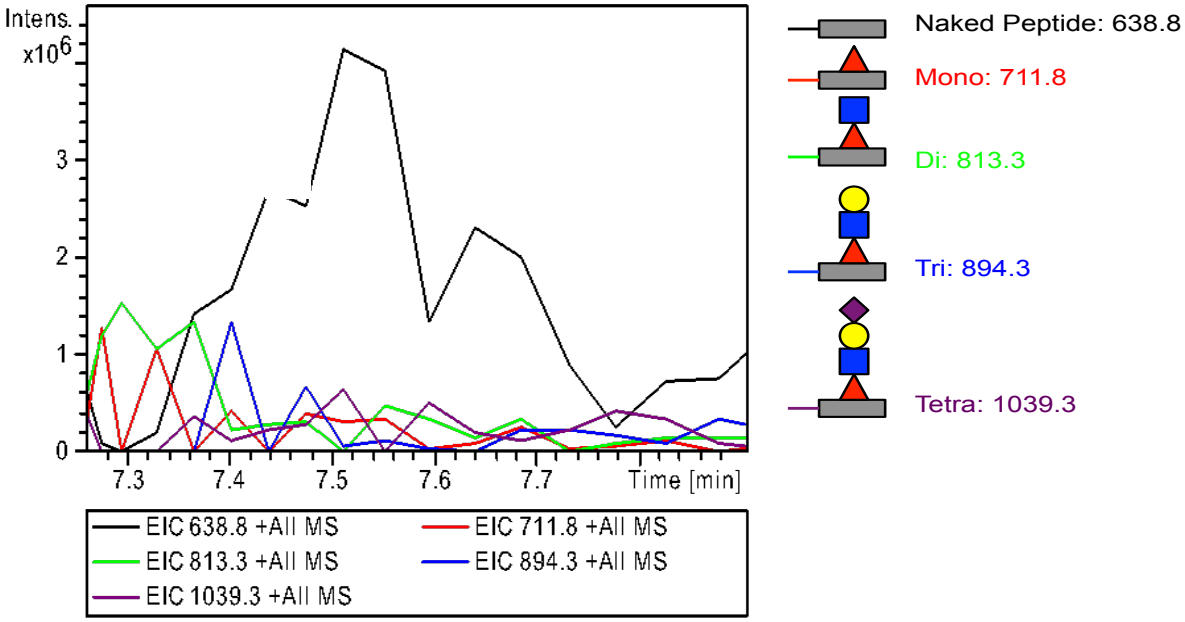


27B

HNGTCVDLVGGF (EGF 30)

O-Fucosylation not found (naked)

Mass: 1276.6 Da = (638.8 x 2) - 1

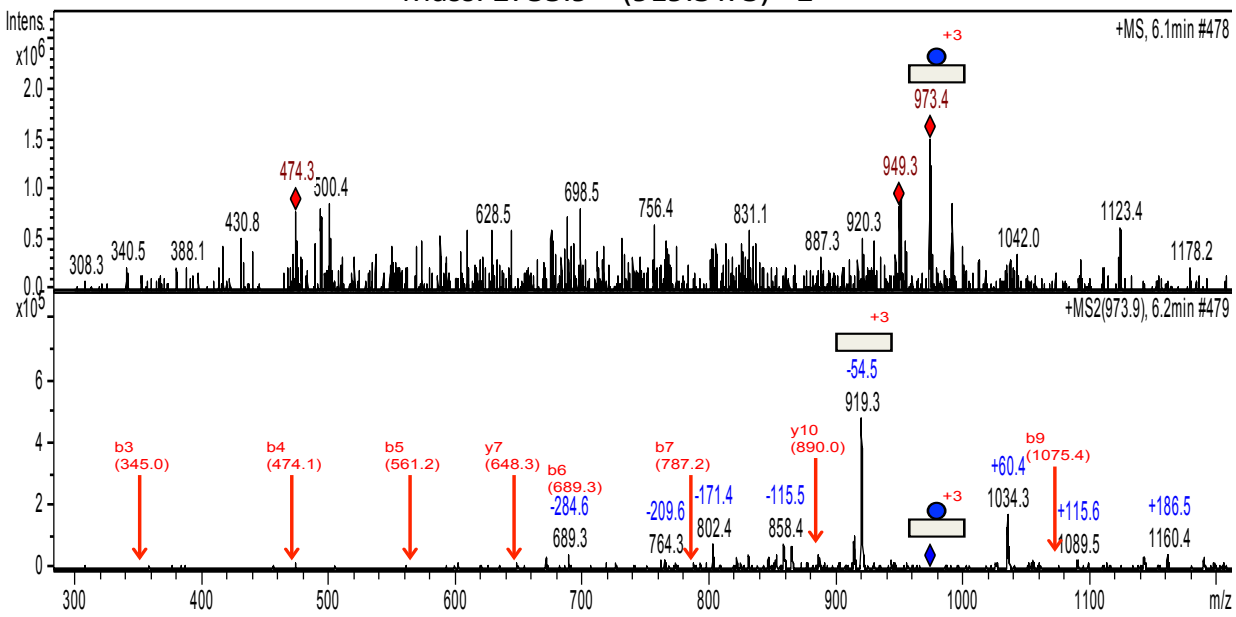


28A

SPCESQPCQHGGQCRPSPGPGGGLTF (EGF 32)

O-Glucosylation (Mono)

Mass: 2755.9 = (919.3 x 3) - 2

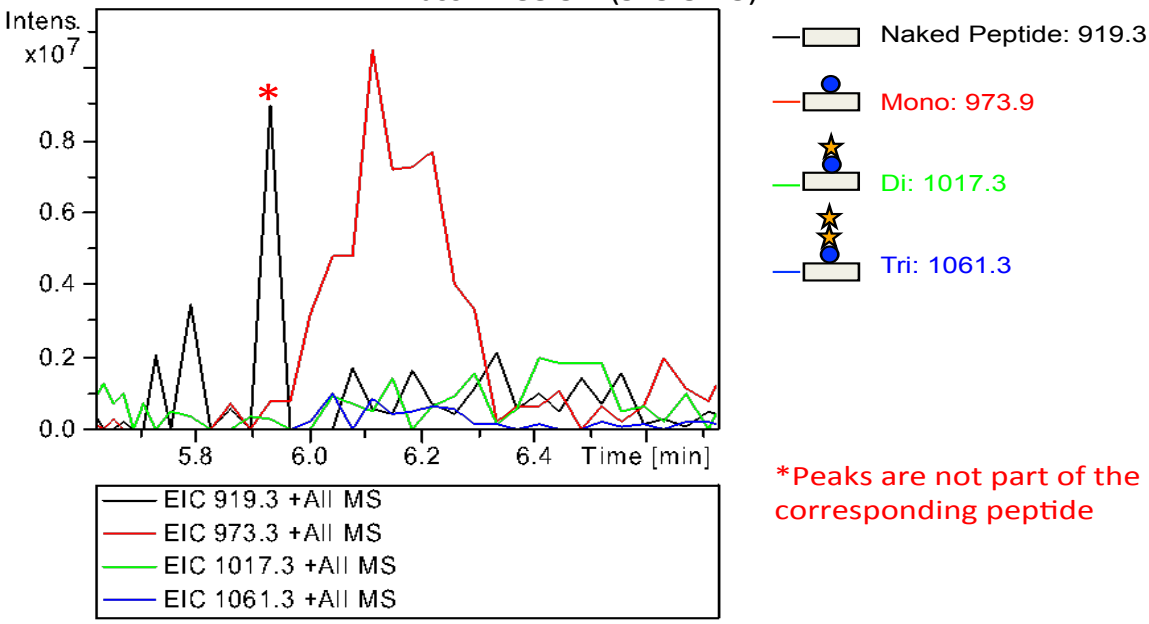


28B

SPCESQPCQHGGQCRPSPGPGGGLTF (EGF 32)

O-Glucosylation (Mono)

Mass: 2755.9 = (919.3 x 3) - 2



*Peaks are not part of the corresponding peptide

References

1) Kopan R: **Notch Signaling**. *Cold Spring Harb. Perspect. Biol.* 2012, doi: 10.1101/cshperspect.a011213

2) Rana NA, Haltiwanger RS: **Fringe benefits: Functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors**. *Curr. Opin. in Struct. Biol.* 2011, **21**:1-7

3) Moloney DJ, Shair LH, Lu FM, Xia J, Locke R, Matta KL, Haltiwanger RS: **Mammalian Notch1 Is Modfied with Two Unusual Forms of *O*-Linked Glycosylation Found on Epidermal Growth Factor-like Modules.** *J. Biol. Chem.* 2000, **275**:9604-9611

4) Rampal R, Luther KB, Haltiwanger RS: **Notch signaling in normal and disease States: possible therapies related to glycosylation**. *Curr. Mol. Med.* 2007, **7**:427-445.

5) Federico A, Di Donato I, Bianchi S, Di Palma C, Taglia I, Dotti MT: **Hereditary cerebral small vessel diseases: A review**, *J. Neurol. Sci.* 2012, doi:10.1016/ j.jns.2012.07.041

6) Lewandowska E, Dziewulska D, Parys M, Pasennik E: **Ultrastructure of granular osmiophilic material deposits (GOM) in arterioles of CADASIL patients**. *Folia Neuropathol*. 2011, **49**:174-180

7) Fouillade C, Monet-Leprêtre M, Baron-Menguy C, Joutel A: **Notch signaling in smooth muscle cells during development and disease**. *Cardiovasc. Res.* 2012, **95**:138-146.

8) Rana NA, Nita-Lazar A, Takeuchi H, Kakuda S, Luther KB, Haltiwanger RS: ***O*-Glucose Trisaccharide Is Present at High but Variable Stoichiometry at Multiple Sites on Mouse Notch1**. *J. Biol. Chem.* 2011, **286**:31623-31637

9) Brooks SA: **Strategies for Analysis of the Glycosylation of Proteins: Current Status and Future Perspectives**. *Mol. Biotechnol.* 2009, **43**:76-88

10) An HJ, Froehlich JW, Lebrilla CB: **Determination of glycosylation sites and site-specific heterogeneity in glycoproteins**. *Curr. Opin. in Chem. Biol.* 2009, **13**:421-426.

11) Takeuchi H, Kantharia J, Sethi MK, Bakker H, Haltiwanger RS: **Site-specific *O*-Glucosylation of the Epidermal Growth Factor-like (EGF) Repeats of Notch: EFFICIENCY OF GLYCOSYLATION IS AFFECTED BY PROPER FOLDING AND AMINO ACID SEQUENCE OF INDIVIDUAL REPEATS**. *J. Biol. Chem.* 2012, **287**:33934-33944

12) Sakaidani Y, Furukawa K, Okajima T: ***O*-GlcNAc Modification of the Extracellular Domain of Notch Receptors**. *Methods in Enzym.* 2010, **480**:355-373

13) Arboleda-Velasquez JF, Rampal R, Fung E, Darland DC, Liu M, Martinez MC, Donahue

CP, Navarro-Gonzalez MF, Libby P, D'Amore PA, Haltiwanger RS, Kosik KS: **CADASIL mutations impair Notch3 glycosylation by Fringe**. *Hum. Mol. Gen.* 2005, **14**:1631-1639

14) Al-Shareffi E, Chaubard JL, Leonhard-Melief C, Wang SK, Wong CH, Haltiwanger RS: **6-Alkynyl Fucose is a bioorthogonal analogue for *O*-fucosylation of Epidermal Growth Factor-like repeats and Thrombospondin Type-1 repeats by Protein *O*-Fucosyltransferases 1 and 2**. *Glycobiol*. 2012, doi:10.1093/glycob/cws140

15) Shao Li, Moloney DJ, Haltiwanger RS: **Fringe Modifies *O*-Fucose on Mouse Notch1 at Epidermal Growth Factor-like Repeats within the Ligand-binding Site and the Abruptex Region**. *J. Biol. Chem.* 2003. **278**:7775-7782