Stony Brook University

On the "Fringe" of Notch: Mapping *O*-Fucose and *O*-Glucose Glycans at Specific Sites on the Extracellular Domains of Notch Receptors Using Nano-LC-ESI-MS/MS

A Dissertation Presented

by

Nadia A. Rana

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

May 2011

Stony Brook University

The Graduate School

Nadia A. Rana

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

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

Dr. Erwin London – Chairperson of Defense
Professor, Department of Biochemistry and Cell Biology

Dr. William J. Lennarz
Distinguished Professor, Department of Biochemistry and Cell Bilogy

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

Dr. Kenneth D. Irvine
Professor, Department of Microbiology
Rutgers University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

On the "Fringe" of Notch: Mapping *O*-Fucose and *O*-Glucose Glycans at Specific Sites on the Extracellular Domains of Notch Receptors Using Nano-LC-ESI-MS/MS

By

Nadia A. Rana

In

Molecular and Cellular Biology

Stony Brook University

2011

Notch is an essential cell surface receptor whose proper functioning is critical during development in metazoans. Notch can be modified with two unique forms of O-linked glycosylation: *O*-fucose and *O*-glucose. Protein *O*-fucosyltransferase1 (Pofut1, OFUT1 in flies) and Protein *O*-glucosyltransferase (Poglut, Rumi in flies) are the glycosyltransferases that add *O*-fucose and O-glucose monosaccharides to Notch, respectively. Loss of *Pofut1/Ofut1* or *Poglut/rumi* in mice or flies phenocopies Notch-null phenotypes, suggesting both forms of *O*-glycosylation are essential for Notch function. To gain insights into how these essential modifications impart functionality to the receptor, it is vital to identify sites of *O*-glycosylation, and structures of glycans at each of these sites.

The extracellular domain (ECD) of mouse Notch1 contains 36-tandem epidermal growth factor-like (EGF) repeats, 16 of which contain the consensus sequence for *O*-glucosylation (C^1xSxPC^2). Using nano-LC-ESI-MS/MS to determine occupancy at these 16 sites, I found most predicted sites were efficiently modified, though efficiency of modifying *O*-glucose sites was cell type-dependent, and some sites were under-glucosylated. *O*-Glucose was elongated to Xyl-α1,3-Xyl-α1,3-Glc at all sites. In addition, a novel non-traditional consensus site was found to be efficiently *O*-glucosylated at EGF 9, leading to a revision of the consensus sequence for *O*-glucosylation to allow Alanine N-terminal to Cysteine 2: $C^1xSx(P/A)C^2$.

The ECD of *Drosophila* Notch (dN) also contains 36 EGF repeats, the majority of which contain consensus sequences for *O*-fucosylation ($C^2xxxx(S/T)C^3$) and/or *O*-glucosylation ($C^1xSx(P/A)C^2$). *O*-Fucose moieties can be further extended to a disaccharide by Fringe, a β3N-acetylglucosaminyltransferase and known modulator of Notch. *O*-Glucose moieties can be elongated by a xylosyltransferase to a disaccharide. I have mapped both *O*-fucose and *O*-glucose glycans on the ECD of dN over-expressed in S2 cells co-transfected with or without Fringe using nano-LC-ESI-MS/MS. I have also compared relative levels of *O*-fucosylation and Fringe elongation, as well as *O*-glucosylation and subsequent xylosylation, at specific sites using Multiple Reaction Monitoring (MRM), a semi-quantitative method that allows relative quantitation of glycoforms at a specific site. My results suggest the predicted glycosylation sites are very efficiently modified with the *O*-glycan monosaccharides and that extent of elongation of either *O*-linked modification is substoichiometric and site-specific.

Table of Contents

Abbreviations

ADAM	A Disintegrin And Metalloproteinase
Ala	Alanine
Arg	Arginine
BPC	Base Peak Chromatogram
CbEGF	Calcium binding Epidermal Growth Factor-like repeat
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
Cys	Cysteine
dHex	deoxy-Hexose
ECD	Extracellular domain
EGF	Epidermal Growth Factor-like
EIC	Extracted Ion Chromatogram
ER	Endoplasmic reticulum
ETD	Electron Transfer Dissociation
Fng	Fringe
Fuc	Fucose
Gal	Galactose
Glc	Glucose
Glc	Glucuronic Acid
GlcNAc	N-Acetylglucosamine
GXYLT	Glucoside xylosyltransferase
HD	Heterodimerization Domain
Hex	Hexose
HexNAc	N-Acetylhexosamine
ICD	Intracellular Domain
LNR	Lin-12/Notch repeats
mN1	mouse Notch1
MRM	Multiple Reaction Monitoring
N-EGF:FLAG	*Drosophila* Notch EGF 1-36, 3X-FLAG
Nano-LC-ESI-MS/MS	Nano-Liquid Chromatography-ElectroSpray Ionization-Mass Spectrometry/Mass Spectrometry
NRR	Negative Regulatory Region
Ofut1	*O*-fucosyltransferase 1
Pen	Pentose
PEST	Proline Glutamate Serine Threonine
Pofut1	Protein *O*-fucosyltransferase 1
Poglut	Protein *O*-glucosyltransferase
Pro	Proline
Ser	Serine
Sia	Sialic Acid
Thr	Threonine
XXYLT	Xyloside xylosyltransferase
Xyl	Xylose
β4GalT-1	β4-Galactosyltransferase-1

List of Figures

List of Tables

Acknowledgements

The synthesis of this dissertation, from conception of the experiments through the writing itself, has been the most significant challenge of my academic career. It would not have been possible without the support, patience, guidance, and compassion of the following individuals. It is to these people that I owe my deepest gratitude and thanks.

I especially would like to thank my advisor, Dr. Robert S. Haltiwanger, for his guidance throughout my years of study at Stony Brook University, and for giving me the opportunity as a sophomore to get my first introductions to biochemistry, molecular biology, and the vast world of Glycobiology in his lab, and for giving me a space to flourish there ever since. After MacGyver, you have been the most influential scientist in my life. To my Dissertation Committee—I would like to thank Dr. Erwin London for acting as Chairperson of my Dissertation Committee and for helpful discussions throughout my graduate studies. I would also like to thank Dr. William Lennarz and Dr. Deborah Brown for endless encouragement and helpful discussions. To Dr. Ken Irvine, I give my thanks for hosting me in his laboratory, where I received my introduction to S2 cell culture—without which Chapter 3 of this thesis would not have been possible, and also for serving as my external committee member. I would like to acknowledge Dr. Aleksandra Nita-Lazar, who initiated studies on mouse Notch1, as well as Dr. Shinako Kakuda and Dr. Hideyuki Takeuchi for their efforts on this project. I would like to thank my students Min Li Xu, Hillary Moss, and Ahmed Rab for their hard work and camaraderie, and for making teaching such a joy. Furthermore, I have to thank the members of the Haltiwanger lab who have made pulling 16 hour days fun and memory-filled. From the "Golden Era" when I first began as an undergrad, I must thank Li Shao, Yi Luo, Raajit Rampal, Ola Nita-Lazar, and Kelvin Luther for their encouragements and support. From the "Renaissance" period of my graduate studies, I thank Hideyuki Takeuchi, Shinako Kakuda, Devin Caswell, Eugene Tan, Deepika Vasudevan, and Esam Al-Shareffi for their friendships.

Last but not least, I would like to thank mi familia, who in every stage of my education have given me nothing but the greatest encouragement. To Moma—when I first embarked on this winding path to the Ph.D., you guided me in times of darkness, cheered me on to brighter days of success, and helped me to truly appreciate all of the twists and turns that science (and life) have to offer. You have helped me to foster a love of science and questioning the six W's (especially "why"). And thank you for helpful discussions without which Table 3 would not be possible! To Shazia (Dr. Rana part 1)—for paving the way to the road of success—you are and always will be my role model. Doctors squared! To Omar— thank you for reminding me to have fun when I'm too serious, for making me a proud big sister with your hard work in lab in the summers, and for your help with manual mass spectrometry data searches! Remember: when in doubt, look it up!

Curriculum Vitae

Education:

2004—May 2011	Stony Brook University, Molecular & Cellular Biology Doctoral Program
2009	Fundamentals of the Bioscience Industry Certification
2000—2004	Stony Brook University, (B.S. Biochemistry, Minor in Art History) (2004)

Honors and Awards:

Fellowships:

2008—present	National Institutes of Health Predoctoral Fellow
2004—present	Stony Brook University Presidential Fellowship
2003	Pfizer Summer Undergraduate Research Fellow
2003	Undergraduate Research & Creative Activities (URECA) Summer Fellow
2002	URECA Summer Fellow
2002	Howard Hughes Medical Institute Summer Fellow

Honors:

2004	Departmental Honors in Biochemistry
2004	Magna Cum Laude
2004—present	Phi Beta Kappa
2000—2004	Dean's List
2000—2004	Women In Science & Engineering (WISE) Program Graduate
2002—2004	Honors College Graduate
2002—present	Sigma Beta Honor Society
2002—present	National Society of Collegiate Scholars (NSCS)
2001—present	Golden Key International Honor Society
2000—2001	Chairman's List (College of Engineering & Applied Sciences)

Scholarships:

2010	ASMS Student Travel Award Recipient
2006—2009	Society for Glycobiology Travel Award Recipient
2004	Undergraduate Recognition Award in Expanded Learning for Excellence and Outstanding Achievement
2003	Anne Sayre Award Recipient *The Anne Sayre Prize is competitively awarded annually to a WISE woman who reflects the spirit of Anne Sayre, a scholar, a biographer, fiction writer, and judge. Anne Sayre was the author of the well-known biography Rosalind Franklin and DNA, which established the critical role Rosalind Franklin played in the discovery of the structure of DNA.*
2002—2003	Sam and Rose Berezin Endowment Scholar
2002—2003	Honors College Scholarship
2002	UNCF-Merck Scholarship
2000—2001	Women In Science & Engineering (WISE) Scholarship

Memberships:

2010—present	American Society of Mass Spectrometry (ASMS)
2009—present	New York Academy of Sciences (NYAS)
2006—present	Society for Glycobiology

Research Experience:

Doctoral Research (August 2005—May 2011)
Thesis: On the "Fringe" of Notch: Mapping *O*-fucose and *O*-glucose glycans at specific sites on the extracellular domains of Notch receptors using nano-LC-ESI-MS/MS
Advisor: Dr. Robert S. Haltiwanger

Undergraduate Research for Departmental Honors (January 2002—May 2004)
Thesis: Identification of *O*-Fucosylation Sites on Epidermal Growth Factor (EGF)-like Repeats of Mouse Delta-like1
Advisor: Dr. Robert S. Haltiwanger

Current Responsibilities:

- Planning and execution of three independent research projects and 4 collaborative projects involving identification of sites of *O*-glycosylation on Notch and related proteins.
- Managing role directing and teaching undergraduates and graduate rotation students working on related sub-projects. Also responsible for training new lab and inter-departmental personnel in ESI mass spectrometry as well as sample preparation and analysis.

Skills:

- Technical skills include nano-LC-ESI-MS/MS, SDS-PAGE, Western blotting, protein over-expression/ purification, Superdex/Sephadex column chromatography, immunoprecipitation, PCR mutagenesis, transformations, mammalian and insect cell culture, insect microdissections.
- Computer skills include mastery of Agilent DataAnalysis Software, MASCOT and GPM database searches, Pubmed database searches, BLAST searches, CANVAS, and Microsoft Word, Powerpoint, Excel.
- Language skills include intermediate reading, writing, speaking abilities in French.

Teaching & Mentoring Experience:

Supervised each student below in three or more of the following technical areas: molecular biology techniques, mass spectrometry, tissue culture, protein purification, SDS-PAGE, Western Blotting, chromatography, power point presentations, and research paper writing.

5 graduate rotation students
2 undergraduate Beckman Research Scholars
3 undergraduate HHMI Summer Research Fellows
1 MD/PhD rotation student
2 Medical students completing Honors in Research
5 Honors College Undergraduates
4 high school researchers

Teaching Assistant for Undergraduate Biochemistry I, II (Fall 2005—Spring 2006)
Women In Science and Engineering (WISE) Junior Mentor (2003-2004)

Poster Sessions, Talks:

2011	Glycobiology Gordon Research Conference
2010—2011	Biochemistry and Cell Biology Retreat
2008—2009	Institute of Chemical Biology and Drug Discovery Annual Symposium ICB&DD Annual Symposium Poster Winner (2009)
2008	Biochemistry and Structural Biology Graduate Program Brookhaven National Laboratory Annual Symposium
2006—2010	Society for Glycobiology Conference
2008	Invited Speaker: Glycans and Glycan Binding Proteins in Human Disease Symposium, held in conjunction with the Society for Glycobiology conference
2002—2004	Stony Brook University Celebration of Undergraduate Achievements
2003	Pfizer SURF Awards/Poster Session

Published Abstracts:

Analysis of *O*-glucosylation on mouse Notch1. N.A. Rana, A. Nita-Lazar, S. Kakuda, R.S. Haltiwanger. Glycobiology (2010) 20(11): 1467

Fringe Benefits: Analysis of *O*-glycosylation and Fringe elongation on the extracellular domain of *Drosophila* Notch. N.A. Rana, M.L. Xu, H. Moss, A. Rab, J. Leonardi, H. Jafar-Nejad, R.S. Haltiwanger. Glycobiology (2010) 20(11): 1486

Regulation of the mammalian Notch signaling by the protein *O*-glucosyltransferase Rumi. R. Fernandez-Valdivia, H. Takeuchi, A. Samarghandi, M. Lopez, J.Leonardi, N.A. Rana, R.S. Haltiwanger, H. Jafar-Nejad. Glycobiology (2010) 20(11): 1456

O-Glucosylation of multiple EGF repeats of Notch by Rumi is required for optimal Notch signaling. J. Leonardi, R. Fernandez-Valdivia, N.A. Rana, Y. Li, A. Simcox, R.S. Haltiwanger, H. Jafar-Nejad. Glycobiology (2010) 20(11): 1483

Examining how *O*-fucose and *O*-glucose glycans affect Notch-ligand binding. H. Takeuchi, P. Taylor, S. Kakuda, N.A. Rana, P.A. Hanford, R.S. Haltiwanger. Glycobiology (2010) 20(11): 1484

On the "Fringe" of Notch: mapping sites of *O*-fucosylation and Fringe elongation on *Drosophila* Notch extracellular domain using mass spectrometry. N.A. Rana, M.L. Xu, H. Moss, A. Rab, A. Nita-Lazar, D. Caswell, E. Tan, R.S. Haltiwanger. Glycobiology (2009) 19(11): 1334

O-Glycosylation of ligand-binding and hinge regions of Notch extracellular domain are pivotal to full Notch activation. N.A. Rana, A. Nita-Lazar, H. Takeuchi, E. Tan, R.S. Haltiwanger. Glycobiology (2008) 18(11): 987

Mapping Notch *O*-glycosylation patterns for mechanistic elucidation. N.A. Rana, J. Chaubard, D. Caswell, E. Tan, M.L. Xu, K. Kao, R.S. Haltiwanger. Glycobiology (2008) 18(11): 990

O-Fucose and *O*-glucose: essential modifications for Notch function. R.S. Haltiwanger, H. Takeuchi, N.A. Rana, K. Luther, J. Chaubard, D. Caswell, E. Tan, M.L. Xu. Glycobiology (2008) 18(11): 947

Development of mass spectral methods for mapping *O*-fucose and *O*-glucose modifications on Notch receptors. N.A. Rana, K. Kao, D. Caswell, E. Tan, E. Al-Shareffi, R.S. Haltiwanger. Glycobiology (2007) 17(11): 1246

Identification of a protein *O*-glucosyltransferase required for Notch signaling in *Drosophila*. H. Takeuchi, M. Acar, H. Jafar-Nejad, A. Rajan, D. Ibrani, N.A. Rana, H. Pan, H.J. Bellen, R.S. Haltiwanger. Glycobiology (2007) 17(11): 1212

Role of *O*-glycosylation in quality control of Notch folding. N.A. Rana, A. Nita-Lazar, Y. Luo, R.S. Haltiwanger. Glycobiology (2006) 16(11): 1132

O-glycosylation of Cysteine-knot motifs. R.S. Haltiwanger, M. Dlugosz, Y. Luo, K. Luther, A. Nita-Lazar, N.A. Rana, H. Takeuchi, B.C. Holdener. Glycobiology (2006) 16(11): 1107

O-Linked glycosylation of EGF repeats in development: effects of *O*-fucose on Notch and Nodal signaling. R.S. Haltiwanger, L. Shao, Y. Luo, K. Luther, R. Rampal, N.A. Rana, A. Li. Glycobiology (2002)

Publications:

1. Nadia A. Rana*, Aleksandra Nita-Lazar*, Michael P. Myers, Hideyuki Takeuchi, Shinako Kakuda, Kelvin Luther, and Robert S. Haltiwanger. *O*-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. (In preparation)
   *These authors contributed equally to this work.

2. Nadia A. Rana and Robert S. Haltiwanger. On the "Fringe" of Notch: Mapping *O*-fucose and *O*-glucose glycans at specific sites on the extracellular domains of *Drosophila* Notch receptors using nano-LC-ESI-MS/MS. (In preparation)

3. Nadia A. Rana and Robert S. Haltiwanger. Fringe Benefits: Analysis of *O*-glycosylation and Fringe elongation on the extracellular domain of Notch receptors. Current Opinions In Structural Biology. (In preparation)

4. Hideyuki Takeuchi, Rodrigo C. Fernandez-Valdivia, Devin S. Caswell, Aleksandra Nita-Lazar, Shinako Kakuda, Nadia A. Rana, Megan Macnaughtan, Hamed Jafar-Nejad, and Robert S. Haltiwanger. Rumi functions as both a protein *O*-glucosyltransferase and protein *O*-xylosyltransferse. (In preparation)

5. Bakker, H., T. Oka, A. Ashikov, A. Yadav, M. Berger, N.A. Rana, X. Bai, Y. Jigami, R.S. Haltiwanger, J.D. Esko, and R. Gerardy-Schahn (2009) Functional UDP-xylose transport across the endoplasmic reticulum/Golgi membrane in a Chinese hamster ovary cell mutant defective in UDP-xylose Synthase. *J Biol Chem* **284**, 2576-83.

6. Acar, M., H. Jafar-Nejad, H. Takeuchi, A. Rajan, D. Ibrani, N.A. Rana, H. Pan, R.S. Haltiwanger, and H.J. Bellen (2008) Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **132**, 247-58.

7. Xu, A., N. Haines, M. Dlugosz, N.A. Rana, H. Takeuchi, R.S. Haltiwanger, and K.D. Irvine (2007) In vitro reconstitution of the modulation of Drosophila notch-ligand binding by fringe. *J Biol Chem* **282**, 35153-62.

Chapter I: Introduction and Background

Glycosylation is one of the most complex forms of co- or post-translational modifications to occur on proteins. Over the last several decades, work by numerous laboratories dedicated to studying a broad array of signaling pathways have provided evidence that glycosylation is in fact required for numerous and varied biological events, including inter-cellular communication, blood group type variations, post-implantation embryonic development, and cell migration and morphogenesis (1). Targeted abrogation of glycosylation through knockdown or knockout of genetic loci involved in glycosylation in several model systems, including *Drosophila* and mice, reveal development is severely impaired by glycosylation deficits. A class of naturally occurring mutations in glycosylation pathways has even received special categorization due to their presentation in numerous clinical pathologies, known as Congenital Disorders of Glycosylation (CDG) (2). The amalgamation of these studies and pathologies demonstrates the critical role that glycoconjugates play in both cell-cell communication and throughout development and across species.

Glycans can be categorized into two major classes: *N*-linked and *O*-linked. The former can be found on the amide *N* of Asn residues at the consensus sequence Asn-X-Ser/Thr on a multitude of proteins, and consist of a wide variety of highly branched structures (3,4). Similarly, *O*-linked glycans are found on the hydroxyl *O* of Ser and Thr, and they occur in several subclasses. The most abundant subclass is *O*-GalNAc, which forms the basis of mucin-type *O*-glycosylation. *O*-Xylosylation forms a second common subclass, and serves as the core glycan for glycosaminoglycan formation. *O*-Mannose was originally described in yeast, but

recent work has shown that defects in *O*-mannose glycans result in congenital muscular dystrophies (5). Three other subclasses of *O*-glycosylation include *O*-glucose, *O*-fucose, and *O*-GlcNAc, all of which have been reported on the Notch receptor. It has been long-established that *O*-fucosylation is essential for the Notch signaling cascade, an important pathway in metazoan development (6,7). More recently, *O*-glucosylation has also been shown to be a major player in proper Notch function (8-10). Although its function has yet to be established, *O*-GlcNAc has also been observed on Notch. The biological and structural relevance of these *O*-glycans to Notch function will be the focus of this work.

_O_-Fucose, _O_-Glucose, and _O_-GlcNAc modification of EGF repeats

The first reports of *O*-glucosylation were on bovine blood coagulation factors VII and IX (11,12). An *O*-glucose trisaccharide (Xyl-α1,3,-Xyl-α1,3-Glc-β1-*O*-Ser) was later reported on these and additional proteins, including the homologous human factors VII and IX, as well as protein Z (13), thrombospondin (14) and murine fetal antigen-1/delta-like protein (FA-1-DLK) (15). Shortly after, *O*-fucosylation was also discovered on a small group of secreted proteins including factors VII, IX, and XII, as well as tissue-type and urinary-type plasminogen activators (16). Most recently, *O*-GlcNAcylation was reported on *Drosophila* Notch (17). All of these modifications were localized to specific sites on Epidermal Growth Factor-like (EGF) repeats in the modified proteins.

EGF repeats are a widely occurring short structural motif of approximately 40 amino acids that is defined by the inclusion of six conserved Cysteines that pair to form three disulfide bonds in a specific pattern: Cys1-Cys3, Cys2-Cys4, and Cys5-Cys6 (Figure 1A). *O*-Glucose modifications occur on Serines only between the first and second conserved Cysteines of the

consensus sequence C^1xSxPC^2, which was first proposed by Nishimura *et al.* in 1989, based on the identification of five *O*-glucosylated proteins. Proteins identified by database searches as potentially modified based on containment of this consensus sequence have been demonstrated by our lab and others to bear the modification, indicating its utility as a strong and accurate prediction tool for *O*-glucosylation (8). The *O*-fucose modification is observed on Serines or Threonines between the second and third conserved Cysteines, at the consensus sequence $C^2xxxx(S/T)C^3$ (18,19). Recently, *O*-GlcNAcylation, a third form of *O*-glycosylation was identified on EGF repeats as well, on hydroxy amino acids between C5 and C6 (17). Databases searches for additional proteins predicted to be *O*-glycosylated using the consensus sequences described reveal the Notch family of receptors contain the highest incidence of each consensus sequence, and are predicted to be the most widely modified by *O*-glycosylation (2,9).

In *Drosophila*, *O*-fucose modifications have been shown to exist as trisaccharides. Recently, Aoki *et al.* carried out studies to define the glycome of *Drosophila melanogaster* by mass spectrometry, using β-elimination to remove and derivatize the *O*-linked sugar modifications from total cell lysates (20). They discovered a novel *O*-linked fucose trisaccharide consisting of Fucose, GlcNAc, and Glucuronic Acid. Given the composition of the structure, the glycan was proposed to be derived from Notch, an EGF repeat-containing protein know to bear *O*-Fucose-GlcNAc disaccharide, the core structure of the novel glycan. This modification remains to be confirmed on Notch. *O*-Glucosylation has also been detected on EGF repeats in monosaccharide form, but is capable of being elongated by a xylose to a disaccharide form of the structure: Xyl-α1,3-Glc-β1-*O*-Ser (8,20). Further elongated forms of both *O*-glycans are observed in mammals, where *O*-fucose can be elongated to the tetrasaccharide structure Sia-α2,6-Gal-β1,4-GlcNAc-β1,3-Fuc-α1-*O*-Ser/Thr (14,16), and *O*-glucose is capable of being

elongated by an additional xylose to the trisaccharide Xyl-α1,3-Xyl-α1,3-Glc-β1-*O*-Ser (9,13,21,22) (Figure 1A).

Several enzymes are responsible for the addition and subsequent elongation of *O*-fucose to EGF repeats. Protein *O*-fucosyltransferase 1 (Pofut1 in mice, OFUT1 in flies) adds fucose to Serines and Threonines (23,24). Fringe is a β1,3 N-acetylglucosaminyltransferase that extends *O*-fucose to a disaccharide, the most extended structure detected in flies (25). While three Fringe homologs exist in mammals (Lunatic, Manic, and Radical Fringe), a single Fringe exists in flies. The structure GlcNAc-β1,3-Fuc-α1-*O*-Ser/Thr can be further elongated in mammals to a trisaccharide by β1,4-Galactosyltransferase, which adds a Galactose to the GlcNAc. Sialyltransferase adds a terminal sialic acid on the galactose to generate the fully extended tetrasaccharide in mammals. Rumi is a protein *O*-glucosyltransferase (Poglut in mammals) that is responsible for addition of glucose specifically to Serines only (8). Two human glucoside xylosyltransferases have been recently identified: GXYLT1 and GXYLT2 are glucose: α1,3-xylosyltransferases that extend glucose to a disaccharide, the most extended structure observed in flies (26). Also recently discovered is XXYLT, the xylose: α1,3-xylosyltransferase that adds a second xylose to generate the Xyl-α1,3-Xyl-α1,3-Glc-β1-*O*-Ser trisaccharide in mammals (Sethi and Bakker, Personal Communication). *O*-GlcNAcylation has also been recently observed to occur on EGF repeats in *Drosophila* (17,27). A strict consensus sequence has yet to be defined, though the modification is known to occur on Serines and Threonines that appear immediately C-terminal to an aromatic residue, between Cys5 and Cys6 of EGF repeats. No known biological function for *O*-GlcNAc on EGF repeats has been established to date.

Biological roles of *O*-Fucose, *O*-Glucose, and *O*-GlcNAc on Notch

The Notch protein plays an important role as a transmembrane signaling receptor in a wide variety of developmental pathways (28,29). The Notch genetic locus was first identified in *Drosophila* by T.H. Morgan in 1917 (30). *Notch* mutants display an X-linked lethal phenotype in flies, with female carriers bearing a characteristic small "notch" in their wings. Notch is conserved across several species, including humans. Importantly, there are four mammalian homologs of Notch (Notch1-4). Loss of individual mouse Notch homologs 1 or 2 results in embryonic lethality (31,32). Mutations of Notch or downstream signaling components within the pathway have been implicated in a multitude of disease states in humans, including T-cell Leukemia (33), CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy) (34), Alagille Syndrome (35), Spondylocostal Dysostosis (36), Multiple Sclerosis (37), several heart defects (7) and Breast Cancer (38). The Notch locus encodes a large (~300 kDa) single-pass Type I transmembrane receptor, comprised of a large extracellular domain (ECD) involved in ligand-binding interactions, a transmembrane region, followed by a short intracellular domain (ICD) involved in downstream signaling events (Figure 1B) (39). The ICD of Notch is rich in regulatory motifs, including a RAM domain, immediately followed by an NLS and seven tandem ankyrin repeats. Following these motifs are a second NLS, a Transcriptional Activation Domain (TAD), and a PEST degron sequence involved in turnover of the ICD at the C-terminus. The ECD of Notch is made up of 29-36 tandem Epidermal Growth Factor like (EGF) repeats, followed by a Negative Regulatory Region (NRR) comprised of 3 LIN-12/Notch repeats (LNR) and a heterodimerization domain (HD) (40). Before transport to the cell surface, full length Notch is initially processed in the trans-Golgi

network by a furin-like convertase, generating the mature heterodimeric receptor that can be presented on the cell surface. This first cleavage event is referred to as S1 processing.

The Notch signaling pathway is activated upon binding of the Notch receptor to one of its ligands presented on an apposing cell (Figure 2). The Notch ligands, collectively referred to as DSL (Delta, Serrate, Lag2), are, like the receptor itself, Type I transmembrane glycoproteins. There are two *Drosophila* ligands, Delta and Serrate, while mammals have five counterparts: Delta homologs Delta-like (Dll) 1, Dll3, Dll4, and Serrate homologs Jagged 1 and Jagged 2. Following ligand interaction, Notch undergoes a proteolytic cleavage by an extracellular protease (Kuzbanian in *Drosophila*, ADAM10 in mammals), resulting in cleavage of the ECD at the membrane interface (known as S2 processing). A final cleavage is catalyzed by the enzyme gamma-secretase just inside the membrane, resulting in release of the ICD of Notch (S3 processing). The ICD is then free to translocate to the nucleus, bind to a member of the CSL (CBF/SuH/LAG-1) family of transcription factors via its RAM domain, and this binary complex serves as a platform for MAML-family co-factors, which in turn can act as scaffolding for transcriptional complex formation. The assembled transcriptional activation complex can then activate expression of Notch target genes, including *Hairy-Enhancer of Split* (Hes) and *Hairy-Enhancer of Split related with YRPW motif* (Hey) family members (40).

O-Fucosylation of Notch

Genetic studies of murine *Pofut1*, and its *Drosophila* counterpart *Ofut1*, have indicated that this enzyme is universally required for all Notch signaling (24,41,42). In *Drosophila*, RNAi under the control of different promoters was used to abrogate the expression of Ofut1, both generally, and in specific tissues (41). Expression of *Ofut1* RNAi resulted in lethality, and

expression in specific tissues of *Drosophila*, such as the wing and eye, presented phenotypes that are consistent with those observed as a consequence of Notch signaling reductions. In addition, independent knock-out experiments with flies carrying a null allele of maternal *Ofut1*, known as *Neurotic* (*nti*), also showed that activity of nti/Ofut1 was necessary for proper wing formation, a phenotype consistent with that observed in Notch mutants (42). In these studies, nti/Ofut1 was also shown to be vital for binding between Notch and Delta. Similar genetic experiments were conducted in a murine system. Knockout experiments in which mice lacking the *Pofut1* gene were produced, resulted in embryonic lethality at day 9.5 (24). These embryos showed severe defects in several areas of development, including neurogenesis, somitogenesis, vasculogenesis, and cardiogenesis. The phenotypes displayed in these embryos are similar to those observed in mice that have global defects in the Notch signaling pathway, indicating lack of *O*-fucosylation is equivalent to non-functional Notch. More recently, Stanley and co-workers observed that while murine ES cells that were Pofut1-/- stably expressed Notch receptors at the cell surface, these receptors were not able to bind ligand or transduce Notch signals in the absence of *O*-fucosylation (43).

The effect of reduced *O*-fucosylation on Notch activity was explored by the Lowe lab in a pair of elegant studies using two different knockout mouse models. GDP-4-keto 6-deoxymannose epimerase-reductase (FX) is one of two key enzymes involved in the conversion of GDP-mannose to GDP-fucose, an essential step in the *de novo* synthesis of GDP-fucose, the donor sugar for Pofut1 and OFUT1. Taking advantage of a salvage pathway available in mammals, whereby exogenous fucose can be converted to GDP-fucose, conditionally deficient FX null mice can be effectively rescued providing a fucose-supplemented diet. The involvement of Notch in T lymphopoiesis having been established, the Lowe lab explored the role of *O*-

fucosylation of Notch in myeloid cell development using common myeloid progenitors (CMPs) isolated from FX null mice (44). Co-culture of these cells with ligand expressing cells resulted in impaired Notch-ligand binding and activation relative to wildtype derived CMPs, indicating *O*-fucosylation of Notch is required for myelopoiesis. In a second study, to further explore the physiological functions of *O*-fucosylation on Notch function, Pofut1 was inactivated in Mx-Cre mice by induction, which resulted in reduced T lymphopoiesis and increased marginal-zone B cell production (45). Consistent with the studies in Pofut1 null ES cells (43), Pofut1 deficiency did not appear to affect cell surface expression of Notch1 or Notch2 significantly, but did prevent binding with Delta-like ligands, resulting in markedly decreased Notch activation. This is significant because it delineates a role for *O*-fucosylation in hematopoiesis through governance of Notch-ligand binding interactions. These experiments, in conjunction with those carried out in *Drosophila*, demonstrate that modification of Notch by *O*-fucose is essential for Notch function.

Fringe is responsible for the transfer of *N*-Acetylglucosamine (GlcNAc) onto *O*-linked fucose in a ß1,3 linkage on EGF repeats (25). It was first identified due to its modulation of Notch signaling during *Drosophila* wing development, where it has opposing effects on Delta and Serrate ligand activation: Fringe serves to potentiate Delta activation of Notch, while signaling via Serrate is inhibited. In wing discs where both ligands are expressed, it is this opposing effect on cell-signaling that places a narrow stretch of Notch activation along the edge of *fringe* expression at the interface of dorsal and ventral cells. The Fringe mutant phenotype in flies is therefore characterized by loss of wing development, due to a lack of differential Notch activation at the dorsal-ventral boundary during development of the wing imaginal disc. In addition, loss of Fringe affects development of other fly tissues, including eyes, legs, and

ovaries, and is therefore essential in the regulation of the development of multiple tissues in the *Drosophila* system (7).

The complementary mammalian system is more complex when juxtaposed with the fly system.. In contrast to Pofut1 knockouts, mice with mutations targeted to one of the three mammalian Fringe homologs (Manic, Lunatic, and Radical Fringes) do not exhibit as severe a phenotype. Lunatic Fringe knockout mice exhibit perinatal lethality, marked by severe defects in somite formation, including malformation of ribs and truncated vertebral columns (46). Individual Radical and Manic Fringe knockout mice do not show an obvious developmental phenotype, and double knockout mice lacking both Lunatic and Radical Fringe exhibit the Lunatic Fringe phenotype only, and not a more severe one (47). Lunatic and Manic Fringe double knockout mice exhibit defects in splenic B-cell development (48). Lunatic Fringe has previously been distinguished for its ability to drive B-cell fate over T-cell fate in the thymus (49). Lunatic and Manic Fringe also enhance interactions of Notch1 medicated by Dll4 in T-cell progenitors in thymic epithelial cells (50,51). These GlcNAc transferases also modify Notch2 in concert in marginal zone B-cell progenitors by enhancing Notch2-Dll1 interactions on endothelial cells, thereby regulating MZ B-cell production. In addition, Notch2-Dll1 interactions may be highly Fringe-dependent. Specifically, Lunatic and Manic Fringes were found to be endogenously co-expressed with Notch2 in specific splenic B-cell subpopulations, including NF B cells and MZPs. Dll1 null mice were used to identify splenic cell subpopulations responsible for Dll1-driven marginal zone B-cell development (52). In combination with the studies in *Drosophila*, these results demonstrate that Fringe has a profound effect on the Notch pathway in a wide variety of biological contexts.

O-Glucosylation of Notch

Genetic studies on the role of *O*-glucosylation has not been as extensive as for *O*-fucosylation, but the recent identification of the *Drosophila* Poglut (Rumi) has allowed for parallel studies of this *O*-glycan. A Poglut activity capable of adding *O*-glucose to EGF repeats was detected and initially characterized in extracts of a variety of cell line and rat tissues (53). *Rumi*, the gene encoding Poglut in *Drosophila,* was subsequently identified during a mutant screen in for modifiers of Notch activity (8). Loss of *rumi* was shown to result in a temperature sensitive Notch-phenotype in flies. A mammalian homolog of Rumi has been identified (54), and recent work by Fernandez-Valdivia *et al.* demonstrated that *Rumi*-null mice phenocopy Notch-like defects, that also result in embryonic lethality (9). In addition, the genes encoding both mammalian homologs of the *O*-glucose-α1,3-xylosyltransferase, as well as the xyloside-α1,3-xylosyltransferase, have recently been identified (55), (Sethi and Bakker, Personal Communication), though the biological relevance of elongation by xyloses on Notch *O*-glucosylation sites is undefined as of yet.

O-GlcNAcylation of Notch

O-GlcNAc has also been recently observed to modify the EGF repeats of *Drosophila* Notch (17). This was a surprising finding, as this modification has previously and extensively been reported to occur only on cytoplasmic and nuclear proteins. *O*-GlcNAc Transferase (OGT) is the enzyme responsible for carrying out intracellular transfer of *O*-GlcNAc to proteins, but RNA*i* knockdown does not decrease levels of *O*-GlcNAcylation on *Drosophila* Notch, indicating a separate OGT is responsible for carrying out this modification along the secretory pathway

(17). A biological function for this modification remains to be elucidated for this form of glycosylation on Notch.

Molecular details for how *O*-glycans affect Notch function

Studies on O-fucose

O-Fucosylation is essential for proper Notch signaling. Initial reports from *in vitro* studies indicated that EGF 11-12, which define the ligand-binding region of Notch receptors, were both necessary and sufficient for ligand binding (Figure 1B). However, data from several studies in the *Drosophila* and murine systems indicate that *O*-fucose plays a role in ligand binding, and that multiple *O*-fucose sites may be involved in contributing to receptor-ligand binding. Deletion mapping of multiple EGF repeats of the ECD of *Drosophila* Notch in combination with site-directed mutagenesis of *O*-fucosylation sites allowed the Irvine lab to determine that multiple *O*-fucose sites along the length of the receptor contribute to Notch-ligand binding in the fruit fly, not just EGF 11-12 of the ligand-binding region (56). To investigate the degree of importance of EGF 12 of Notch1 in mice, which is both *O*-fucosylated and a part of the ligand binding region, Ge and Stanley generated a transgenic mouse line carrying a point mutation in the Threonine of the *O*-fucosylation site (57). Genetic ablation of the *O*-fucose site at EGF 12 results in only a mild Notch phenotype, strongly indicating additional *O*-fucose sites are likely contributing to Notch activity. Consistent with this idea, recent work from our laboratory examined the role of highly conserved *O*-fucosylation sites in mouse Notch1 in addition to EGF 12. Mutation of individual *O*-fucose sites and testing of their activities using cell-based signaling assays revealed elimination of *O*-fucose at EGF repeats 12 or 27 resulted in significant decreases in Dll1 and Jagged-1 mediated Notch1 signaling, while mutation at EGF

repeat 26 resulted in increased Notch1 activation. From these studies, it is apparent that additional sites along Notch, independent of the ligand-binding region, are contributing to Notch activation.

In flies, Ofut1 has been proposed to hold a dual role as both a glycosyltransferase and as a chaperone, with elimination of the enzyme resulting in retention of Notch in the ER (58,59). However, this dual functionality is not clearly observed in mice, as ES cells and thymocytes lacking Pofut1 are able to express Notch on the cell surface (43,57). The major effect of losing Pofut1 on Notch function in mice appears to be loss of ligand binding. Further work needs to be done to investigate the chaperone role of Pofut1/Ofut1 in more contexts before this issue is resolved.

Effects of Fringe on Notch activity

The majority of data suggests that the modifications mediated by Fringe cause a change in Notch-ligand binding. A compendium of studies have been aimed at determining whether this effect is directly mediated by transfer of GlcNAc, or if this modification serves as a platform for indirect regulation by allowing further downstream elongation to the tri- or tetrasaccharide forms of the *O*-linked glycan.

Fringe has been shown in the *Drosophila* system to exert its regulatory effects by altering Notch-ligand binding, and to be able to mediate this effect without requiring further extension past the disaccharide form of *O*-fucosylation. This was first suggested in experiments with *Drosophila* S2 cells in 2000, where assays involving Notch co-expressing Fringe with soluble Delta ligand showed increased receptor-ligand binding when compared to binding in the absence of Fringe (60). Subsequently, Okajima *et al.* showed in 2003 that Fringe influences binding of

both *Drosophila* ligands (Delta and Serrate) to Notch in cell-based and *in vitro* binding experiments (61). Importantly, a differential effect of Fringe on Delta and Serrate was demonstrated *in vitro*, where Notch-Delta binding was potentiated, and Notch-Serrate binding was inhibited, which paralleled data from *in vivo* studies.

In flies, data indicates that GlcNAc is the terminal sugar added to *O*-fucose residues on *Drosophila* Notch, and that the disaccharide is sufficient for observing a Fringe effect (62). More recently, *in vitro* extension of GlcNAc-Fucose-*O*-Ser/Thr to a trisaccharide using galactosyltransferase showed no difference in ligand binding, demonstrating that it is the GlcNAc added by Fringe that directly regulates Notch-ligand binding in *Drosophila*. Therefore, although in mammalian tissue culture lines it appears important, galactose is not required to see a Fringe effect in *Drosophila*. This is a very interesting result, as it showcases the simplicity of the *Drosophila* system, and makes it very appealing for experimental studies in comparison to the complexity of the homologous mammalian systems. This is a major issue in selecting a system for studying Fringe effects, because when studying the effects of Fringe on ligand-binding, it is essential to include any additional factors that may be necessary for Fringe to exert its effects, and in the case of *Drosophila*, no additional glycosyltransferases are necessary past Fringe elongation to see its effect. The simplicity of the *Drosophila* system, combined with the conclusion that Fringe exerts its effects on Notch signaling due to its ability to alter ligand binding, and the fact that *in vivo* observations can be recapitulated *in vitro*, makes further studies in this system very promising.

In stark contrast, the complexity of the mammalian system can be defined by the fact that in the canonical Notch signaling cascade, there are 4 receptors, 5 ligands, and 3 Fringe homologs to contend with. Because of the existence of multiple homologs, the effects of the different

Fringes on various Notch-ligand pairs becomes somewhat more complicated to decipher than in *Drosophila*. This can be seen in cell-based signaling assays carried out by Hicks *et al.* that showed Lunatic Fringe potentiates Dll1-mediated Notch1 signaling but inhibits Jagged-1 mediated Notch1 signaling (63). This study also demonstrated Lunatic Fringe potentiates both Dll1 and Jagged-1 mediated Notch2 signaling. Contradictory to these complex findings were results from a study conducted by Shimizu *et al.*, which showed Lunatic Fringe has no effect on Dll1 and Notch2 signaling, and that Lunatic Fringe does not potentiate, but instead inhibits Jagged-1 mediated Notch2 signaling (64).

It has been observed that mammalian Fringes (Manic or Lunatic) are capable of decreasing binding of Jagged1 (equivalent to *Drosophila* Serrate) to mammalian Notch2 (64). To address whether the disaccharide generated by Fringe action (GlcNAc-Fucose-*O*-Ser/Thr) is sufficient to observe a Fringe effect, or if additional elongation is necessary to see this effect, Notch expressing CHO cells deficient in tri- or tetrasaccharide biosynthesis, and co-expressing Manic or Lunatic Fringe were co-cultured with Jagged1 expressing cells to determine the minimal extent of elongation necessary to observe a Fringe effect (65). Elongation beyond GlcNAc to the trisaccharide (Gal-ß1,4-GlcNAc-ß1,3-Fucose-α1-*O*-Ser/Thr) was necessary to see a Fringe effect this mammalian cell system, an important departure from requirements in the fly system. Consistent with these findings, ß4GalT1 knockout mice exhibit a very minor Notch phenotype (66).

Effects of O-glucose on Notch function

To examine what roles Rumi and *O*-glucose play in proper Notch function, Acar *et al*. decided to study the steps of the canonical signaling cascade where Rumi may be exerting

regulation (8). Cell-surface staining of wing imaginal discs, a progenitor tissue in *Drosophila* development, that contained *Rumi* mutant clones showed Notch accumulating at the cell surface, indicating that *O*-glucosylation is not required for cell surface presentation of the receptor. Defects in ligand binding were ruled out because Notch generated in Rumi knockdown cells bind Delta as strongly as protein generated from control cells, and Notch activation is not observed in *rumi* mutant cells in spite of being cultured adjacent to cells over-expressing Delta or Serrate ligand. Interestingly, protein extracts from various tissues of *rumi* mutant animals showed significantly decreased amounts of Notch cleavage products, which is highly suggestive of a defect in proteolysis (S2 or S3 cleavage, Figure 1A) caused by lack of *O*-glucosylation. This temperature-sensitive Notch phenotype may be associated with protein instability. Importantly, in contrast to Ofut1, Rumi has not shown any chaperone-like activity. Defects in *rumi* mutants implicate *O*-glucosylation of Notch contributes to stabilization of the ECD, and that lack of *O*-glucose causes a defect in the ability of the receptor to couple ligand-binding to requisite conformational changes necessary for proteolysis.

Structural analysis of Notch signaling

The necessity of structural studies of the Notch family of receptors lies in gaining insights into the molecular mechanisms by which the structural features of this protein impart functionality. Recent innovative structural studies of interest describe both the ICD and ECD of mammalian and *Drosophila* Notch proteins. Novel insights into the mechanisms by which ICD mediates transcriptional activation of downstream Notch gene targets have been gained through crystallographic studies. Additional studies have allowed the elucidation of both the LNR and HD domain structures, which have allowed for a deeper understanding of how Notch activation

is triggered. Finally, NMR, X-ray crystallography, and STEM experiments revolving around defining EGF repeat architecture, the major structural motifs of the ECD have furthered our understanding of global Notch conformations, and will be explored here.

In the nucleus, the ICD can form a transcription complex with CSL family DNA-binding transcriptional regulators, along with additional co-factors including members of the Mastermind-like (MAML) family. CSL that is already bound to cognate DNA in head-to-head CSL recognition sites, can interact with the newly released ICD through a high-affinity site within the RAM domain, and then through a secondary interaction of the ankyrin repeats and CSL. X-ray crystallographic studies by Arnett *et al.* have recently revealed that two ICDs can dimerize via their Ankyrin domains, generating a flexible complex that could potentially rotate to accommodate varied spacing between CSL-binding sites at different promoters (67). MAMLs can then bind to the CSL/Ankyrin binary scaffold, whose C-terminal regions form platforms for additional transcriptional factors, including RNA Pol II, to join the growing complex and initiate transcription of downstream targets. It has been suggested that the cooperative dimerization of these transcription complexes, which occurs on paired-site promoter DNA, including the Hes gene family, coupled with the spacing of these CSL-binding sites, dually contribute to controlling the sensitivity of different promoters to Notch signaling.

As mentioned previously, LNRs are constituents of the NRR, an important regulatory region of the ECD, and mutations in this region results in a hyperactive form of Notch1 found at high frequency in cases of T-cell acute lymphoblastic leukemia lymphoma (T-ALL) (39). LNRs prevent premature ADAM10 cleavage at the S2 site by folding over and obscuring access to the heterodimerization (HD) domain, thereby serving as an autoinhibitory clamp for the HD domain (68-70). This has been further confirmed through brilliant studies utilizing novel Notch1

antibodies that recognize the NRR of normal and mutant Notch1 receptors. Aste-Amezaga *et al.* demonstrated that the anti-NRR antibody acts as a potent inhibitor for both wild type and single point-mutant Notch1 in which the autoinhibitory shielding that LNRs normally provide is decreased relative to wild-type (68). The antibody functions by mimicking the protective masking that LNRs provide to the cleavage/activation site, thereby acting as a surrogate negative regulator for the LNR. Further studies involving examination of Notch1 NRR in its autoinhibitory conformation by x-ray crystallography confirms a conserved hydrophobic clamp that sterically hinders access to the metalloprotease cleavage site S2. A novel mutation within this portion of the NRR was also examined. Although this mutation (H1545P) caused only minor destabilization of the NRR, it was found to invoke ligand-independent increased Notch signaling, indicative of its involvement in the autoinhibitory clamp on the HD (69).

To further explore the role of this region in Notch signaling, the S1 cleavage sites of Notch1 and Notch2, which lie within the NRR, were mutated to be resistant to proteolysis. While structurally similar to wild-type receptors, S1-resistant Notch1 and Notch2 receptors exhibited differential effects on cell surface expression and ligand-mediated receptor activation. Examination of T-ALL associated NRR mutations also showed variable dependence on S1 cleavage. Coupled together, these findings highlight the significant role of the NRR in Notch activation and regulation and show that significant conformational changes occur in the ECD subsequent to ligand binding to permit proteolytic activation.

The structural implications of *O*-glycosylation on Notch-ligand interactions have yet to be defined, but recent structural studies provide insights into Notch ECD global structure, and the possible impact of *O*-glycosylation on altering that structure. Several studies have focused on the structure of EGF 11-12, the ligand binding region of Notch (56,71). The Handford group

determined the structure of a bacterially expressed (therefore unglycosylated) form of EGF 11-13 from human Notch1 and showed that it can bind ligands *in vitro* (72). They performed docking studies with Jagged1 and made predictions about which residues in each molecule bind to one another. Interestingly, their docked structure has the *O*-fucose site on EGF12 outside of the binding interface suggesting that the O-fucose glycans may not directly participate in binding. Recently, the crystal structures of unmodified, *O*-fucosylated, and Fringe elongated mouse Notch1 EGF 12 was solved (72). Interestingly, Fringe elongation to the GlcNAc-β1,3-Fucose appears to alter the conformation of local residues significantly, with the fucose moiety believed to contribute to stability of the anti-parallel β-sheet, and interaction of GlcNAc with neighboring residues, causing local conformation shifts, and this may be a mechanism for how Fringe modification indirectly exerts its effects on Notch activity at this site. As *O*-fucose and *O*-glucose sites appear to be rotated away from each other on single EGF repeats, the possibility remains that *O*-glucose on EGF 12 may be interacting with Jagged1 at the Notch1 binding interface.

Although the Notch ECD is traditionally schematically represented as a linear and rigid structure (Figure 2), several recent studies indicate that this is unlikely an accurate depiction of the global structure of the receptor. Some EGF repeats are capable of coordinating calcium at the N-terminal linker region between two adjoining EGFs. These calcium-binding EGF repeats (cbEGF) are rich in acidic residues in the N-terminal linker region. Using EGF 11-13 of human Notch1, Hambleton and co-workers observed that in the presence of calcium, the calcium-binding motifs in the linker between two adjacent EGF repeats impart rigidity to this region (73). CbEGFs have a consensus sequence that includes the presence of conserved acidic/polar residues in the N-terminal linker region ($NNxNC^1$, where N can be D/E/Q/N, x any amino acid, and C^1 is

the first conserved Cysteine of the EGF repeat), and which allows for calcium ion coordination. The flexibility of non-cbEGF linkers, which do not coordinate calcium, is supported by the observation that consecutive non cbEGF repeats from *Plasmodium falciparum* merozoite surface protein-1 assume a U-shaped structure in which the non cbEGF domains share a hydrophobic interface. Based on these results, Hambleton and co-workers proposed that the Notch ECD consists of both rigid and flexible regions dependent on the presence or absence of cbEGFs. Of special interest is the fact that the organization of calcium-binding sites within the Notch ECD is highly conserved (56). This high degree of conservation of regions of structural rigidity and flexibility suggests their existence is essential for Notch ECD function (Chapter IV, Figure 29).

Support for this model of regions of flexibility and rigidity comes from electron microscopy (EM) studies by Kelly *et al.* Using the ECD of both *Drosophila* and human Notch1, affinity grids were used to prepare the proteins for EM analysis. The images depict structures that appear to have globular structures with occasional interspersed bends, consistent with the flexible regions predicted to exist in both proteins based on positioning of cbEGFs in these receptors. Density maps suggested that the receptors actually form dimers, so the authors confirmed this finding using STEM. While these EM studies do give insights into Notch ECD conformations, they do not address that the proteins being examined have not been demonstrated to be biologically active, either in their monomeric or dimeric forms. Further studies need to be performed to determine whether Fringe modification alters this structure.

Aims of the Dissertation

Although *O*-glycosylation of Notch has been established as essential for receptor function, the numerous predicted sites of *O*-fucosylation and *O*-glucosylation predicted to decorate this protein remain largely undefined. Confirmation of these sites of *O*-glycosylation and definition of the structures of glycans at these sites would allow a better understanding of where and how these sugars are able to impart functionality to the receptor. In addition, definition of these modification sites will also provide insights into how *O*-glycosylation impacts global Notch architecture, including potential effects on conformational changes. The primary aim of the present studies was to determine site occupancy of predicted sites of *O*-glucosylation and *O*-fucosylation on mouse Notch1 and *Drosophila* Notch, and to define glycan structures at any modified sites using mass spectrometry. To determine whether any *O*-glucose sites are important for ligand-mediated Notch1 activity, we used site-directed mutagenesis and cell-based signaling assays, parallel to studies carried out previously by our laboratory for conserved sites of *O*-fucosylation. We also explored Fringe and GXYLT modification on *Drosophila* Notch to determine the degree of elongation to disaccharide at these *O*-glycosylated sites, using a semi-quantitative mass spectral method known as Multiple Reaction Monitoring. Together, these studies will provide essential information necessary for molecular models of how these glycans affect Notch activity.

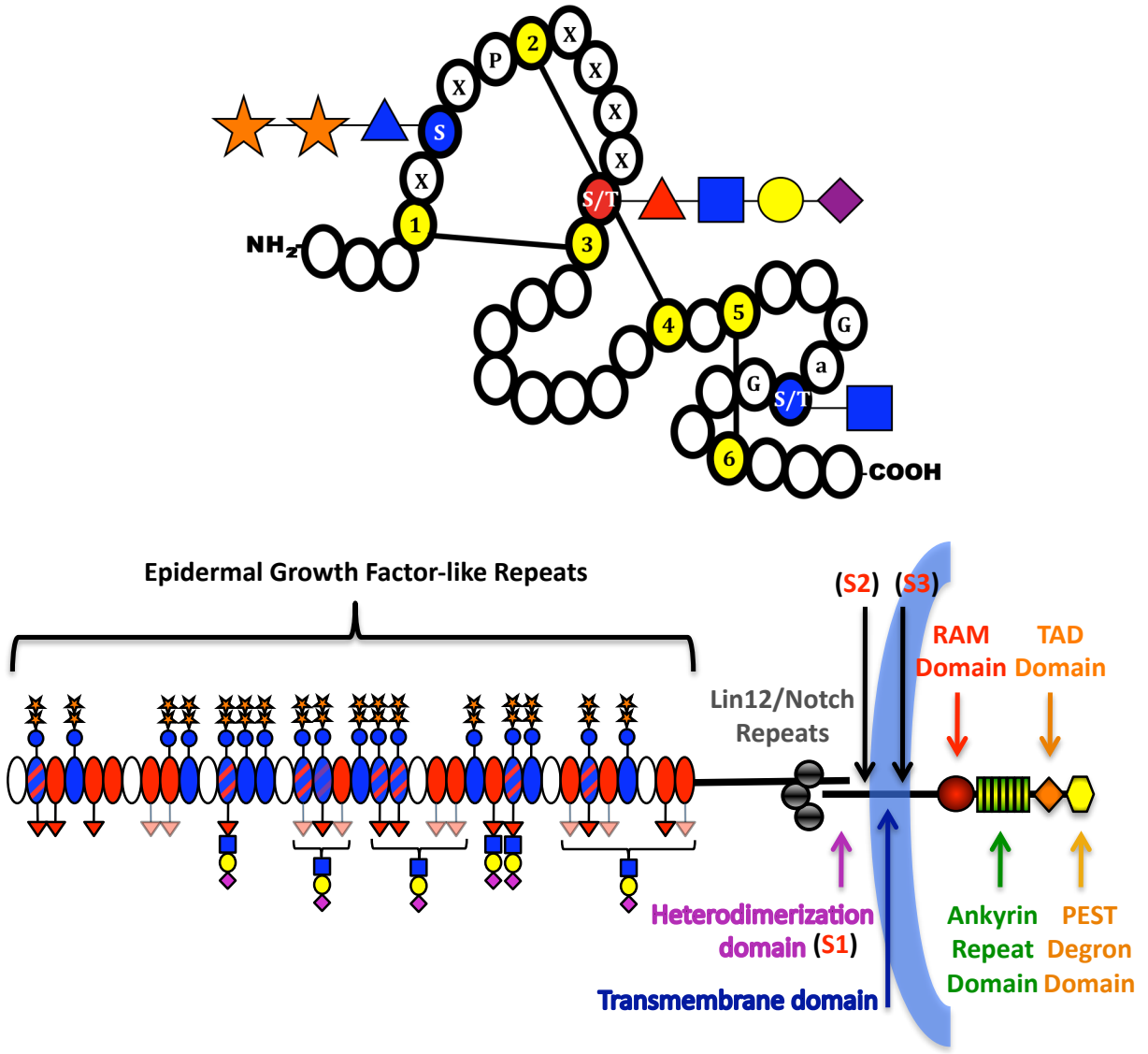


Figure 1. Schematic representation of an Epidermal Growth Factor-like (EGF) repeat, and domain structure of the Notch receptor.
A. Short 40 amino acid motifs known as EGF repeats are defined by containing six conserved Cysteines (yellow) that form three disulfide bridges. Potential sites of *O*-glucosylation (C^1XSXPC^2), *O*-fucosylation (C^2XXXXS/TC^3), and *O*-GlcNAcylation (no consensus sequence defined, but observed between Cys 5 and Cys6 on Serine or Threonine following Glycine and an aromatic residue (a), and preceding Glycine) are indicated in red or blue. Symbols for sugars: Fucose (red triangle), GlcNAc (blue square), Galactose (yellow circle), Sialic Acid (purple diamond), Glucose (blue circle), Xylose (orange star) **B.** The extracellular domains (ECD) of mouse Notch1 and *Drosophila* Notch consists of 36 tandem EGF repeats (blue rectangles) followed C-terminally by three Lin12/Notch repeats (LNRs) (grey circles) and a heterodimerization domain (HD) (black overlapping lines). The intracellular domain (ICD) consists of a RAM domain (red circle), 7 Ankryin repeats (green rectangles), a transcriptional activation domain (TAD) (orange diamond) and a PEST degron (yellow hexagon).

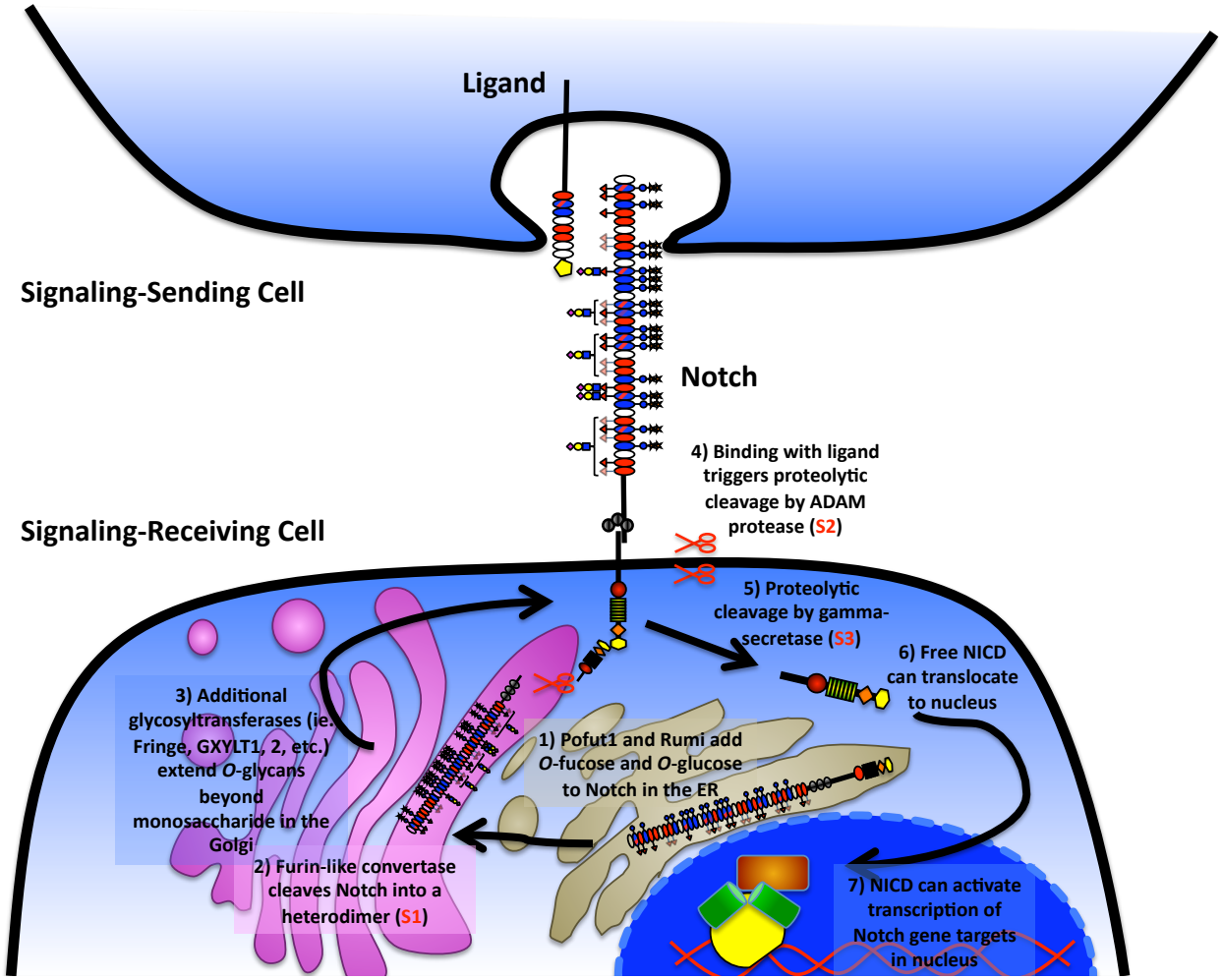


Figure 2: Notch activation pathway.
Notch is *O*-fucosylated and *O*-glucosylated in the ER, and elongated by additional glycosyltransferases in the Golgi. In this compartment, Notch is cleaved by a furin-like convertase (S1 cleavage) to form the active heterodimeric receptor which is subsequently expressed at the cell surface. Following interaction with a DSL ligand presented on an apposing cell, Notch ECD is cleaved by an ADAM protease (S2 cleavage), and ICD is cleaved by gamma-secretase within the cell membrane (S3 cleavage). Upon S3 cleavage, ICD is free to translocate to the nucleus and activate transcription by turning "off" active repression at transcription start sites of downstream Notch targets.

Chapter II: *O*-Glucose Trisaccharide is Present at High But Variable Stoichiometry at Multiple Sites on Mouse Notch1 [1]

Introduction

Notch activity is regulated by both *O*-fucosylation and *O*-glucosylation, and Notch receptors contain multiple predicted sites for both *O*-glucose and *O*-fucose modifications. Nonetheless, little is known about which predicted sites are occupied or what structures exist at each site. The Notch proteins consist of a large extracellular domain (ECD), transmembrane region and a large intracellular domain (28,40,74). The majority of the ECD consists of tandem epidermal growth factor-like (EGF) repeats (36 EGFs are present in mouse Notch1 and 2, 34 EGFs are found in mouse Notch3, and only 29 exist in mouse Notch4), which are conserved protein domains characterized by six Cysteine residues forming three disulfide bridges (75). EGF repeats are known to participate in protein-protein interactions, such as receptor-ligand binding. Many of the EGF repeats in Notch (and other EGF repeat-containing proteins) are modified with two unusual types of protein glycosylation: *O*-fucose and *O*-glucose (76). The *O*-fucose modification occurs at the consensus sequence $C^2xxxx(\underline{S/T})C^3$, located between the second and third Cysteines of the EGF repeat (77). *O*-Fucose is added by protein *O*-fucosyltransferase1 (Pofut1), which is a soluble protein in the ER (58,78-80). Knockout or knockdown of Pofut1 in either mice (81) or *D. melanogaster* (41,42) results in severe Notch-like phenotypes. Mouse Notch1 (mN1) is modified at multiple predicted sites with *O*-fucose glycans (19,82), and

[1] Portions of the work presented in this chapter have been previously published, and the remainder is in preparation for submission. Experiments performed by additional members of the Haltiwanger Lab are so noted in the appropriate figure legends.

elimination of individual *O*-fucose sites on mN1 (EGF repeats 12, 26, 27) alters activity in cell-based assays (82,83). Elimination of the *O*-fucose site on EGF12 *in vivo* results in defects in T cell development, confirming an important role for *O*-fucosylation of this site (57). These results are indicative of *O*-fucosylation being essential for Notch function, and form a paradigm for the role of glycosylation in Notch signaling.

The *O*-glucose modifications of EGF repeats have received much less attention to date. *O*-Glucose was first discovered on bovine blood clotting factors VII and IX (84) as a trisaccharide (Xyl-α1,3-Xyl-α1,3-Glc-β1-*O*-Ser) (85). The same trisaccharide was later found on human factors VII and IX, protein Z (86), thrombospondin (87) and murine fetal antigen-1/delta-like protein (FA-1-DLK) (88). Comparison of the sites of modification on these five proteins led to a proposed consensus sequence, C^1xSxPC^2, between the first and the second conserved Cysteine residues of an EGF repeat (89). To date, only Serine residues have been found to be modified with *O*-glucose, but it is unknown whether Threonine can be modified. This consensus site can be found in EGF repeats of numerous secreted and membrane-bound proteins (Table 2) (53) including the Notch receptors. Mouse Notch1 contains *O*-glucose consensus sequences in 16 of its 36 EGF repeats (Figure 3A). Endogenous Notch1 isolated from CHO cells is known to be modified with *O*-glucose trisaccharide (76), but the actual sites modified or structures of the modifications on Notch have not yet been established. A recent report confirmed that the *O*-glucose trisaccharide on Notch has the same structure as that reported above: Xyl-α1,3-Xyl-α1,3-Glc (21). An enzymatic activity capable of adding *O*-glucose to EGF repeats, protein *O*-glucosyltransferase (Poglut) was detected and initially characterized in extracts of cell lines and different rat tissues (53). The gene encoding Poglut, *Rumi*, was subsequently identified during a mutant screen in *Drosophila* for modifiers of Notch

activity (8). Loss of *rumi* results in a temperature sensitive Notch-phenotype in flies. Recent work shows that elimination of mouse Rumi also results in Notch-like phenotypes in mice (9). In addition, the genes encoding two mammalian homologs of the *O*-glucose-α1,3-xylosyltransferase, as well as the xyloside-α1,3-xylosyltransferase have recently been identified (55) (Sethi and Bakker, Personal Communication). Here we focus our examination on the occupancy of the predicted *O*-glucose sites on mouse Notch1 (mN1) using the consensus sequence C^1xSxPC^2, and address to what extent they are elongated with the xyloses. We also address which sites are functionally important by mutating individual sites and analyzing the effects in cell-based Notch signaling assays.

Experimental procedures

Materials

The Lec1-CHO cell line was developed by Dr. Pamela Stanley (90). COS7 and Lec1-CHO cells were from the American Type Culture Collection (Manassas, VA). COS7, NIH3T3, and HEK293T cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Lec1-CHO cells were grown in α-MEM (Invitrogen) supplemented with 10% bovine calf serum. [6-^{3}H]galactose (60 Ci/mmol) and UDP-[6-^{3}H]glucose were from American Radiolabeled Chemicals (St. Louis, MO). All other reagents were of the highest quality available.

Analysis of O-glucose sugars in Notch fragments

Preparation of plasmids encoding the EGF repeats 1-5, 6-10, 11-15, 16-18, 19-23, 24-28, 29-36 and 1-18 from mN1 with C-terminal Myc-His_6 tags (in pSecTag2, Invitrogen) was described previously (77). Lec1-CHO cells stably expressing each of these constructs were generated by treatment of transfected cells with 680 μg/ml hygromycin B (Invitrogen) for several weeks. Clones producing significant amounts of protein were isolated by dilution cloning and evaluated by immunoblot of media using anti-MYC antibody. Metabolic radiolabeling and purification of Notch fragments were performed essentially as described elsewhere (77), but the cells were labeled with 20 μCi/ml [6-^{3}H]galactose. Alkali-induced β-eliminations and size fractionation of *O*-linked sugars on the Superdex column (Amersham Pharmacia Biotech) were performed as described before (91). Analysis of peaks for the presence of glucitol following acid hydrolysis was performed by HPAEC-PAD analysis as described (18).

Analysis of O-glucose site occupancy in mouse Notch1 using CID mass spectrometry

Fragments of the mN1 ECD consisting of EGF repeats 1-5, 6-10, 11-15, 16-18, 19-23, 24-28, and 29-36 (as in the section above) and 1-18 were produced in stably transfected Lec1-CHO cells and purified using nickel-nitrilotriacetic acid chromatography as described previously (77). Some Notch1 fragments were also produced in HEK293T, COS7, and NIH3T3 cells by transient transfection for comparison. The purified fragments were reduced using Tris(2-carboxyethyl)phosphine hydrochloride (Pierce) and carboxyamidomethylated with iodoacetamide prior to in solution digest, or in-gel digest as described (8). For in solution chymotryptic digests, approximately 500 ng of protein was analyzed per run. Reduced and alkylated samples were digested with chymotrypsin (Princeton Separations) for 8 hours at 30ºC in 50 mM Tris-HCl and 2 mM $CaCl_2$. For in-gel digests, approximately 200-500 ng of protein was analyzed per run. The bands of interest were identified using zinc sulfate negative staining (Bio-Rad), excised, and digested with a protease (trypsin, chymotrypsin, Asp-N, V8) which, according to *in silico* digests (using PeptideMass proteomics tool from www.expasy.org), yielded fragments of appropriate size for analysis in the ion trap in the m/z range 400-2200. The digestions were performed for 8-16 hours at 37°C in 20 mM diammonium phosphate, pH 8.0. The digests were desalted on a C18 ZipTip microcolumn (Millipore), dried, and resuspended in 20% acetonitrile, 0.1% formic acid. Samples were then subjected to capillary LC-MS/MS or nano-LC-MS/MS using Collision Induced Dissociation (CID), leading to losses of labile modifications such as *O*-glycans first, followed by fragmentation of the more stabile peptide backbone into b and y-ions.

For capillary LC-MS/MS, digests were fractionated by reverse phase liquid chromatography coupled with electrospray ionization mass spectrometry using a Zorbax C8

capillary column (0.3 x 150 mm) (Agilent Technologies) with a 60 min linear gradient from 0 to 95% buffer B (buffer A: 0.1% formic acid; buffer B: 95% acetonitrile in 0.1% formic acid) at 5 µl/ml. The column effluent was sprayed into an XCT ion trap mass spectrometer (Agilent Technologies) equipped with a capillary electrospray source, operating in the positive ion mode, and set up to perform MS/MS on either the two or three most intense ions in each spectrum.

Samples were analyzed by nano-LC-MS/MS using an Agilent 6340 ion trap mass spectrometer equipped with an HPLC Chip-Cube interface. Sample volumes of 1.0-4.0 microliters were injected onto a Zorbax 300SB-C18 chip with a 40 nl enrichment column and a 43 mm x 75 mm separation column (Agilent). Samples were initially loaded onto the enrichment column at 4.0 ml/min in 5.0% buffer B, and then fractionated on the separation column at 0.45 ul/min with a 26 minute 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). The effluent from the HPLC-Chip was sprayed directly into the ion trap (Agilent), operating in the positive ion mode, and set up to perform MS/MS on the three most abundant ions in each scan, with 30 second active exclusion triggered after 2 consecutive spectra of a recurrent ion. All MS/MS experiments were carried out using the following settings: capillary voltage (-1700-1950V), End Plate Offset (-500V), Dry Gas (5.0 l/min) at 325ºC, trap drive (100), smart target (500,000), maximum accumulation time (150 ms), and scan range (300-2200 m/z).

Spectra were confirmed by searching databases with the MS/MS data using the Global Proteome Machine search engine (http://h777.thegpm.org/tandem/thegpm_tandem.html). All Cysteines were carboxyamidomethylated, and masses of all peptides were adjusted accordingly. The *O*-glucose modified peptides were identified by searching the MS/MS data for constant neutral losses of masses equivalent to the *O*-glucose mono-, di-, and trisaccharides (162, 294,

and 426 Da, respectively) relative to parent ions using the Data Analysis Tool (Agilent) as described (92). *O*-Fucose modified peptides were identified by constant neutral loss searches of 146 Da, and *O*-GlcNAc modified peptides by constant neutral loss searches of 203 Da. Subsequent searches of the MS/MS data were performed to identify peptides modified with mono- or disaccharide forms of *O*-glucose by generating extracted ion chromatograms using the masses of the unglycosylated peptide for the search.

Sequencing analysis of residue of O-glucosylation in mouse Notch1 using ETD mass spectrometry

Electron Transfer Dissociation (ETD) transfers electrons from an anionic reagent (fluoranthene) to protonated peptides causing fragmentation into c and z-ions. This lower energy fragmentation method leaves the labile *O*-glycan modification attached to its hydroxy amino acid, thereby retaining positional sequence information of the modification within a given peptide. Samples were characterized by nano-LC-MS/MS using an Agilent 6340 ion trap mass spectrometer equipped with ETD for peptide sequence analysis by MS/MS with glycans remaining intact. Parameters were as follows: reactant temperature was 60^0C, ionization energy 80eV, emission current 2uA, ionization chamber -3.7V, with an accumulation time of 25-35 ms and trap drive of 25.

Mutagenesis of O-glucosylation sites in mouse Notch1

The site-directed mutagenesis of all *O*-glucose sites in mN1 was performed using as a template a full-length construct of mN1 where the C-terminal PEST domain was replaced with six tandem MYC epitopes (Notch1-Myc_6) in the vector pCS2+ (a generous gift from Dr. Raphael

Kopan, Washington University School of Medicine). All glycosylation sites (Serine residues) were mutated to Alanine, and the mutated constructs were sequenced to confirm the presence of the mutation (Table 4). For EGF 28, the Ala residue was reverted back to Ser to demonstrate that the aberrant signaling was not due to any other inadvertently introduced mutation in Notch. All mutations were confirmed by sequencing.

Co-culture Assay

This assay was adapted from our previously described co-culture assay for NIH3T3 cells (Rampal et al. 2005). 1.0×10^5 NIH3T3 cells were seeded in a 24-well tissue culture plate and transiently transfected using Lipofectamine 2000 (Invitrogen), with 0.4 ug of wild type or mutant mouse Notch1-pCS2+ vector, 0.15 ug of TP-1 luciferase reporter construct (Ga981-6, a gift from Dr. Georg Bornkamm, Munich, Germany), and 0.15 ug of gWIZ B-galactosidase construct (Gene Therapy Systems) for transfection efficiency normalization. 4 h post-transfection, cells were allowed to recover in fresh DMEM. To begin co-culture, L-cells (control) or L-cells expressing Delta-like1 or Jagged1 (kind gift of Dr. Gerry Weinmaster, UCLA) were overlaid on the transfected NIH 3T3 cells at a density of 1.0×10^5 cells per well. 24 h post co-culture, cell lysates were prepared as previously described (82). All co-cultures were carried out in triplicate, with assays showing significant changes in Notch1 activation carried out at least twice.

Mutation of factor VII EGF Repeat

Mutation at the predicted *O*-glucose consensus site in human factor VII EGF was introduced by a conventional site-directed mutagenesis method, converting Serine 52 to Threonine (S52T). PCR was performed using the pET-20b(+) plasmid containing factor VII

EGF as a template and the primers, whose sequences were 5'-GACCAGTGTGCCACGAGTCCATGCCAG-3' and 5'-CTGGCATGGACTCGTGGCACACTGGTC-3', under the following conditions: PCR cycle number 18, denaturation at 95°C for 30 sec, annealing at 55°C for 45 sec, and elongation at 68°C for 4 min. Successful mutation was confirmed by DNA sequencing. Expression and purification of wild-type and mutated factor VII EGF proteins were performed as described previously (93). As a final step of purification, reverse phase HPLC was carried out. The properly folded factor VII EGF repeats were identified by their ability to be *O*-fucosylated by Pofut1 *in vitro* (53). The final concentration of factor VII EGF was determined by a BCA assay using BSA as a standard.

Protein O-glucosyltransferase assay

Protein *O*-glucosyltransferase assays were performed with slight modification as previously described (53). Briefly, 10 µl reaction mixture contained 50 mM HEPES pH 7.0, 10 mM $MnCl_2$, the indicated amounts of Factor VII EGF, 0.16 µM (0.01 mCi/ml) UDP-[^{3}H]glucose, 10 µM UDP-glucose, 0.5% NP-40, and partially purified protein *O*-glucosyltransferase from mouse brain extracts. The reaction was incubated at 37°C for 1 hour and stopped by adding 900 µl of 100 mM EDTA pH 8.0. The sample was loaded onto a C18 cartridge (100 mg). After the cartridge was washed with 5 ml of H_2O, the EGF repeat was eluted with 1 ml of 80% methanol. Incorporation of [^{3}H]glucose into the EGF repeat was determined by scintillation counting of the eluate. Reactions without substrates were used as background control.

Other methods

Western blots, fluorographs, and biotinylations were performed as described previously (53,82).

Results

Mouse Notch1 is extensively modified with O-glucose in Lec1-CHO cells

We have previously shown that endogenous Notch1 is modified with *O*-glucose trisaccharide in Lec8-CHO cells and that most of the *O*-glucose is in the trisaccharide form (76). However, it is not known which of the predicted *O*-glucose sites are modified, nor if the structure varies from one site to another. Mouse Notch1 contains 16 predicted *O*-glucose sites in its extracellular domain (Figure 3A). To determine whether *O*-glucose modifications are spread across the entire ECD, as suggested by the distribution of *O*-glucose sites (Figure 3A), we expressed fragments of mN1 ECD as described previously (77), containing EGF repeats 1-5, 6-10, 11-15, 16-18, 19-23, 24-28 and 29-36 with C-terminal Myc and His_6 tags, and N-terminal secretion signal sequence.

The fragments were stably expressed in Lec1-CHO cells and the cells were metabolically radiolabeled with [^{3}H]galactose. Since galactose is converted to UDP-galactose and then epimerized to UDP-glucose in cells, radiolabeling with [^{3}H]galactose permits specific labeling of the UDP-glucose pool without labeling all of the other products of glucose metabolism in cells (94). Since Lec1-CHO cells do not synthesize complex-type N-glycans (95), the radioactivity should be preferentially incorporated into O-glucose structures in Lec1-CHO cells. Radiolabeled Notch1 fragments were purified from conditioned medium using Ni^{2+}-nitrilotriacetic acid-agarose and analyzed by SDS-PAGE followed by Western blot and fluorography (Figure 3B). Under these conditions, each of the fragments expressed in Lec1-CHO cells was radiolabeled.

To confirm that the radiolabel incorporated into each Notch fragment was in the form of *O*-glucose, and to address the question of which glycan structures are present on each of the

Notch fragments, we released the *O*-linked sugars from the purified, radiolabeled fragments by alkali-induced β-elimination and analyzed the products by gel filtration chromatography (76,91). We found that both trisaccharide and monosaccharide species of *O*-glucose were present on all fragments, but trisaccharide was clearly the predominant species (Figure 4). The radiolabel in each of the monosaccharide and trisaccharide peaks was confirmed to be in the form of glucitol (the expected product from ß-elimination of *O*-glucose) using high-pH anion exchange chromatography following acid hydrolysis (data not shown) (76).

Additionally, some of the fragments contained larger, radiolabeled oligosaccharides, especially abundant in the fragments EGF16-18 and 29-36 (marked with *). Analysis of these oligosaccharide species by high pH anion exchange chromatography following acid hydrolysis revealed the presence of radioactive galactose (not glucitol), a building block of mucin-type oligosaccharides and glycosaminoglycan cores (data not shown). Thus, these high molecular weight species represent galactose-containing species other than *O*-glucose. Similar results were obtained in the original analysis of endogenous Notch1 in Lec8-CHO cells (76). These results indicate that *O*-glucose trisaccharide modifications are found at many sites scattered across the mN1 ECD.

Site-specific mapping of O-glucose modifications using mass spectrometry

To examine whether specific *O*-glucosylation sites were modified, we turned to mass spectral methodologies similar to those we have used to map *O*-fucose modification sites on Notch (8,22,92,96). We initially analyzed the same Notch fragments used in the radiolabeling experiments shown in Figures 3 and 4, although larger portions of mN1 (*e.g.* EGF1-18) were also analyzed to confirm that utilization of glycosylation sites was not affected by the size of the

fragment being analyzed. Each purified fragment was reduced, alkylated, and subjected to in-gel digestion with trypsin, chymotrypsin, or V8, depending on the suitability of peptides resulting from either digest for MS analysis, based on *in silico* digests of the mN1 sequence. The resulting peptide mix was subjected to LC-MS/MS as described in Experimental Procedures. Modified peptides were identified by neutral loss of the *O*-glucose saccharides upon low energy CID fragmentation.

An example of this analysis, performed on the EGF1-5 fragment from mN1 expressed in Lec1-CHO cells is shown in Figure 5. The top panel shows the base peak chromatogram (BPC), a measure of the most abundant ions eluting from the reverse-phase HPLC at each time point (Figure 5A). The data were searched for ions that lose an *O*-glucose trisaccharide (the most abundant form of *O*-glucose, Figure 4) upon fragmentation (Constant Neutral Loss). Figure 5B shows a Constant Neutral Loss search of 142 Da (loss of the *O*-glucose trisaccharide, 426 Da, from a triply charged peptide). A single molecular ion was found in this search eluting at 53.5 min, and both the MS spectrum and the CID fragmentation (MS/MS) spectrum for this ion are shown (Figure 5D). The MS spectrum (top panel of Figure 5D) reveals an ion (m/z 1254.5) corresponding to a triply charged tryptic peptide containing an *O*-glucosylation site from EGF repeat 4 modified with the *O*-glucose trisaccharide. Fragmentation (bottom panel, Figure 5D) results in major product ions corresponding to sequential loss of a pentose (xylose), another pentose (xylose), and a hexose (glucose) from the glycopeptide. This data indicates the presence of the *O*-glucose trisaccharide on this peptide. The most abundant product ion (m/z 1112.9) is the unglycosylated peptide. The mass of this ion corresponds to the triply charged form of a predicted tryptic peptide from EGF repeat 4 of mN1 that contains an *O*-glucose consensus

sequence (in bold): ^{137}SCQQADP**C**A**S**N**PC**ANGGQCLPFESSYICR^{165} (see Table 1 for predicted masses).

Some fragmentation of the peptide (after loss of the sugars) also occurs, resulting in b- and y-ions that confirm the identification of the peptide (several y-ions are indicated in Figure 5D, bottom). We do not reliably see b- or y-ions still modified with the glycans, so assignment of the modification to a specific residue is difficult to deduce from the data. Nonetheless, the presence of the *O*-glucose consensus site (C^1x**S**xPC^2) within the peptide is highly suggestive of the modified residue. Our analysis has revealed that the fragmentation pattern obtained here (major product ions resulting from the sequential losses of the two pentoses and a hexose) is diagnostic for identification of *O*-glucose trisaccharide-modified peptides. To search for the same peptide modified with *O*-glucose mono- or disaccharide, we searched the MS/MS data for the unglycosylated peptide (m/z 1112.8, Figure 3C, Extracted Ion Chromatogram or EIC). Two closely eluting peaks were detected (labeled "D" and "E" in Figure 5C). The larger peak, D, is the same species shown in Figure 5D (peptide modified with *O*-glucose trisaccharide). The smaller peak, E, corresponds to the peptide modified with *O*-glucose monosaccharide. The MS scan and CID fragmentation of E is shown in Figure 5E. This result is consistent with the presence of *O*-glucose monosaccharide on EGF1-5 in the radiolabeling experiments (Figure 4, EGF1-5). The peptide with *O*-glucose monosaccharide elutes slightly later (53.9 min) than the trisaccharide form (53.5 min), consistent with the absence of the xyloses.

By searching for neutral loss of the *O*-glucose saccharides with the characteristic losses of xylose, xylose, and glucose (as shown in Figure 5), we found ions corresponding to peptides containing all predicted *O*-glucosylation sites: EGF repeats 2, 4, 10, 12, 13, 14, 16, 17, 19, 20, 21, 25, 27, 28, 31, and 33 (Table 1, Figure 5, Figure 15A-O). The peptides were initially

identified based on the characteristic fragmentation pattern of the *O*-glucose trisaccharide (as in Figure 5D). The masses of the unglycosylated peptide were then matched to predicted masses of tryptic (or other proteases) peptides from mN1 that contain *O*-glucose consensus sites (Table 1). The identity of each peptide was confirmed by identifying unique b- and y-ions in the MS/MS spectra. One peptide, from EGF repeats 15 and 16, carried both the *O*-glucose trisaccharide and an additional HexNAc modification. (Table 1, Figure 15F). Sequential loss of masses corresponded to the masses of GlcNAc, Xyl, Xyl, and Glc, and totaled a loss of 629 Da. Either HexNAc or xylose can be lost first, indicating that both are terminal. Since this peptide contains a predicted site for *O*-GlcNAcylation from EGF repeat 15 (17), we believe the HexNAc is an *O*-GlcNAc modification. Several ions were identified corresponding to peptides containing both *O*-glucose and *O*-fucose modifications (EGF repeats 12, 17, 20, 21, 27, and 31, Table 1 and Figure 15). These peptides showed sequential losses of masses corresponding to fucose, xylose, xylose, and glucose, for a total loss of 572 Da. Either fucose or xylose can be lost independently, indicating that they are unlinked. A number of *O*-fucose sites were also mapped during this study using similar methods (EGF repeats 2, 3, 5, 12, 17, 20, 21, 27, 31, 35, Tables 1 and 2, and Figure 15).

All 16 peptides identified in this manner contain the consensus sequence $C^{1}x\underline{S}xPC^{2}$ (Table 1), indicating that the consensus sequence is a useful prediction tool for identification of potentially *O*-glycosylated proteins. Extracted Ion Chromatogram (EIC) searches for unglycosylated peptides (as in Figure 5C) revealed that the trisaccharide form of *O*-glucose predominates on the samples analyzed (data not shown), consistent with the radiolabeling results shown in Figure 4. One notable exception was EGF 27, which is modified by *O*-glucose as well as *O*-fucose. The most abundant form of this peptide contained the *O*-fucose monosaccharide

with no *O*-glucose modification (Figure 6A). Smaller amounts were observed bearing the *O*-glucose trisaccharide and *O*-fucose monosaccharide, and even less with the *O*-glucose monosaccharide alone or naked peptide. These results indicate that this peptide is *O*-fucosylated at high stoichiometry, but *O*-glucosylated at sub-stoichiometric levels. Alternatively, the presence of the O-glucose trisaccharide could suppress ionization, although we have not observed such suppression on other peptides. In contrast, the glycopeptide from EGF 12 also contains dual *O*-glycosylation sites, but in this case the predominant species contains both the *O*-glucose trisaccharide and the *O*-fucose monosaccharide (Figure 6B). EGF repeats 17, 20, 21, and 31 also have dual *O*-glycosylation sites, and like EGF 12, the major species is fully extended *O*-glucose trisaccharide with *O*-fucose monosaccharide (data not shown).

To determine whether the high stoichiometry of *O*-glucose site occupancy was cell-type specific, we analyzed the *O*-glucosylation of EGF 1-5 expressed in COS7 cells. In contrast to what we observed in Lec1-CHO cells (or HEK293T cells), a significant amount of unglycosylated peptide from EGF4 could be detected in the samples from COS7 cells suggesting under-glucosylation of this site (Figure 7). Interestingly, there is very little monosaccharide present in samples from either cell line, suggesting while the xylosyltransferases extend *O*-glucose with high stoichiometry, there is a degree of variability in stoichiometry of *O*-glucosylation that is cell or tissue type-dependent.

A non-traditional consensus site CASAAC is modified with O-glucose trisaccharide

While performing neutral loss searches for *O*-glucosylated peptides, a 17th glycopeptide was discovered that did not correspond to any of the masses of peptides predicted to be *O*-glucosylated based on the consensus sequence C^1xSxPC^2. When the search was expanded to

include masses of "non-glycopeptides" produced by chymotryptic digest of mouse Notch1, the glycopeptide was identified as including the C1-C2 region of EGF 9 (Figure 8A). To verify the fragmentation pattern was in fact that of *O*-glucosylated EGF 9, the sample was subjected to MS2 and Constant Neutral Loss-triggered MS3, both of which showed strong series of y-ions that correspond to the fragmentation pattern of EGF 9 (Figure 8B, 8C). Since the chymotryptic peptide contained several hydroxy amino acids, we wanted to confirm the position of *O*-glucosylation within the peptide by a second mass spectral method. Using ETD, we were able to map the modification to the only hydroxy amino acid (Serine) between C1 and C2 of EGF 9 (Figure 8D, 8D'). We next examined the extent of elongation of the O-glucose at this novel site by EIC searches of the data. Like the majority of the traditional consensus sites, the glycopeptide including EGF 9 is *O*-glucosylated and also elongated to trisaccharide with high stoichiometry (Figure 8E). These results suggested that the Proline in the consensus sequence may not be essential for *O*-glucosylation.

Mouse Notch1 contains one other similar site in EGF 35 without the Proline: C^1GSLRC^2. Although no additional ions with neutral losses of 426 Da (trisaccharide) were detected during our neutral loss searches, we searched the data for peptides from EGF35. Only the unmodified peptide was detected (Figure 9). EIC searches for the masses of the theoretical trisaccharide, disaccharide, and monosaccharide glycoforms of this peptide were unsuccessful. We have also confirmed that a similar site in EGF2 from *Drosophila* Notch (C^1NSMRC^2) is also unmodified (data not shown). These results reveal that Alanine, but not Arginine, can substitute for Proline in the consensus sequence for *O*-glucose: C^1x<u>S</u>x(P/A)C^2.

O-Glucosylation only occurs on Serine

Our analysis of *O*-glucose modification sites supports the proposal that only serine can be modified with *O*-glucose. To directly test this hypothesis, we mutated the serine in the *O*-glucose consensus sequence to a Threonine within a bacterially expressed form of an EGF repeat from human factor VII. We have used this EGF repeat in previous studies as an acceptor substrate for protein *O*-glucosyltransferase (53). Interestingly, the Threonine mutant did not serve as an acceptor substrate for protein *O*-glucosyltransferase in *in vitro* assays (Figure 10A). Proper folding of the mutated EGF repeat was confirmed by demonstrating that it functions as well as the wild type EGF repeat in assays with Pofut1 (Figure 10B), which only fucosylates properly folded EGF repeats. These results strongly support the concept that *O*-glucosylation only occurs on Serine residues in the context of the C^1x<u>S</u>x(P/A)C^2 consensus sequence. Consistent with this finding, we found that EGF repeat 21 from *Drosophila* Notch contains a Threonine instead of Serine in the conserved position (CVTNPC) and is not *O*-glucosylated (Chapter III, Figure 21).

Examination of other potential O-glucose sites

We have examined whether closely related sites can be *O*-glucosylated, including several where Serines are located in a different location between C^1 and C^2 of the EGF repeat than defined by the consensus sequence. Mouse Notch1 EGF 6 (CSPSPC) is found unmodified, as expected, due to misplacement of Serines between C1 and C2 (Figure 13). *Drosophila* Notch (dN) contains two EGF repeats (23: C^1SLSSPC^2, and 36: C^1SPNPC^2) with 1 or more Serine residues in non-consensus positions between C1 and C2, none of which are modified (Figure 14A-D). Therefore the revised consensus sequence accurately predicts the position of the

modified Ser residue and Ala/Pro residue between C1 and C2 of EGF repeats, and these amino acid positions are essential for modification. The consensus sequence for *O*-fucosylation, $C^2xxxx(\underline{S/T})C^3$, is also accurate. EGF 29 contains 3 amino acids between C2 and the hydroxy amino acid Thr, instead of the requisite 4 for *O*-fucosylation, and like EGF 15 from mouse Notch1 (19) it is also not modified (Figure 14E,F).

Mutation of most O-glucose sites has little or no effect on Delta-like1 and Jagged1-mediated Notch1 signaling except for EGF 28

In order to determine which *O*-glucose modifications influence Notch signaling, we mutated individual *O*-glucose modification sites and tested them in a cell-based Notch signaling assay (82). All 16 conserved *O*-glucosylation sites, as well as the newly discovered site at EGF 9, were mutated individually (S to A). Constructs for each site-mutant, as well as wild-type Notch1 and an empty vector control were used in previously established cell-based co-culture assays using COS7 cells. In efforts to obtain more robust signaling above endogenous signaling background noise, we optimized several parameters of these assays. This involved dose-dependent co-culture experiments where total amounts of DNA transfected into sending cells were incrementally increased to determine optimal amounts of vector necessary to obtain optimal signal-to-noise ratios. These were confirmed by reporter assays and by Western blot. Other parameters tested included altering Molar ratios of luciferase and β-galactosidase reporter vectors to each other and also by conducting assays using Molar equivalents of Notch1 and empty vectors, testing of alternative transfection reagents and alternative lysis conditions, comparison of alternative reporter plasmid systems, using cells of varying passage numbers, and finally adapting the assay for several different cell systems, including Lec1-CHO cells and

NIH3T3 cells. This assay development culminated in using NIH3T3 cells of low passage number as described in Experimental Procedures. The analyses revealed that elimination of the *O*-glucose site in EGF 28 reduced Delta-like 1 mediated Notch1 activation to near-background levels relative to WT Notch1 (Figure 11A, B). None of the *O*-glucose site mutants resulted in a statistically significant reduction of Jagged1-mediated Notch activation (Figure 11C, D). Mutations in other individual sites did not have any apparent effect on Notch signaling, including mutation at EGF 12. A revertant mutation (A to S) using the EGF 28 site mutant as the template, was generated to ensure that the effect on Notch1 activity was not due to inadvertently introduced random mutations. This revertant exhibited the same level of activation as the wild type Notch1, confirming that the effect was specifically due to the mutation of the *O*-glucosylation site (Figure 11E, F).

Loss of Delta-like1 or Jagged1-mediated Notch signaling could be caused either by alteration in the ability of ligands to activate Notch at the cell surface, or by reduction of the cell-surface expression of Notch. The effects of the mutations on cell-surface expression of mN1 were examined by a cell-surface biotinylation assay. All mutants, including EGF 28, were expressed on the cell surface at a similar level to the wild type protein (Figure 11G), suggesting that loss of *O*-glucosylation on EGF28 is affecting some other step in Notch activation.

Discussion

The consensus sequence for *O*-glucosylation originally proposed by Nishimura and co-workers in 1989 was formulated based on a comparison of the primary sequence of three glycoproteins confirmed to be *O*-glucosylated from humans and cows: blood coagulation factors VII and IX, and protein Z (86). Over 40 mammalian proteins are predicted to be *O*-glucosylated based on this consensus sequence, of which the Notch family of receptors contain the greatest number of predicted sites (9). Although we have previously shown Notch to be *O*-glucosylated, little was known to date about which of the 16 predicted sites on mN1 are actually modified, to what extent, or which ones, if any, play an essential role in Notch signaling. The results presented here confirm that mouse Notch1 extracellular domain is heavily decorated with *O*-glucose saccharides. Using the consensus sequence C^1xSxPC^2, we have identified peptides spanning this region of all 16 EGF repeats of mouse Notch1 predicted to be modified by Poglut/Rumi, and confirmed that all 16 bear the *O*-glucose-xylose-xylose trisaccharide, as detected using mass spectrometry (summarized in Figure 12A). Identifying ions that correspond to peptides containing the consensus sites C^1xSxPC^2, greater by a mass of 426 Da than the predicted mass of the naked peptide in an MS scan, with observed subsequent loss of two pentoses and a hexose in MS/MS, are sufficient to determine the sites of modification. Although this consensus sequence is a useful tool in predicting sites of *O*-glucose modification, using our constant neutral loss search method, we discovered an additional 17th glycopeptide whose naked peptide mass did not correspond to the masses of peptides that included any of the 16 predicted sites. *In silico* digest indicated that the glycopeptide corresponded in mass to a non-traditional site at EGF 9, C^1ASAAC^2. To verify the fragmentation pattern, MS2 was repeated and MS3 was carried out in CID mode. In addition, ETD confirmed the hydroxy amino acid within the peptide

bearing the *O*-glucose trisaccharide was between C1 and C2 of EGF 9. Because the original consensus sequence missed at least one modified site, the remaining 19 non-consensus EGF repeats were examined to determine if any other amino acids besides Pro and Ala are acceptable N-terminal to C2. Of these 18, only EGF 35 maintained a Serine in the appropriate position without a Proline residue before C2 (C^1GSLRC^2), but it was only detected as naked peptide. While mouse Notch2 and Notch3 do not contain any EGF repeats with C^1xSxxC^2 sites, mouse Notch 4 contains 1 site in EGF 20 (C^1VSASC^2), and *Drosophila* Notch contains similar sites at EGF 1 (C^1TSVGC^2), 2 (C^1NSMRC^2), and 22 (C^1ASNRC^2). Like mouse Notch1 EGF 35, *Drosophila* Notch EGF 2 has an Arg before C2, and similarly it has only been detected as naked peptide. In light of these analyses, we propose refinement of the consensus sequence to allow Alanine or Proline N-terminal to C2. As our neutral loss search approach allowed us to discover a novel non-consensus site that would otherwise have been overlooked, and no additional neutral losses were detected, we have confidence in our site-mapping approach and in the revised consensus sequence.

In further testing the consensus sequence, we decided to explore whether Serine is the only hydroxy amino acid modified with *O*-glucose. To accomplish this we used two approaches. The direct approach involved introducing a Ser to Thr mutation within human factor VII, a protein extensively studied and known to be *O*-glucosylated (53). Although properly folded, the Thr mutant did not serve as an acceptor substrate for *O*-glucosylation *in vitro*. The second approach involved examination of non consensus site EGF repeats that contained hydroxy amino acids in various positions between C1 and C2, and for EGF repeats that naturally have a Thr instead of Ser two amino acids after C1. EGF 21 from *Drosophila* Notch has a Thr instead of Serine between C1 and C2 (CVTNPC), and is not *O*-glycosylated. EGFs 6 and 23, also from

Drosophila Notch, contain multiple Serine residues between C1 and C2 flanking a non-hydroxy amino acid (CSPSPC, CSLSSPC), but are also not *O*-glycosylated. Based on all of these results, we propose the revised consensus sequence: $C^1xSx(P/A)C^2$.

Revision of the consensus sequence reveals 5 new potential *O*-glycosylation sites on Notch homologs (Table 3). Expanding the search to allow any amino acid N-terminal to C^2 reveals 36 additional potential sites, 20 of which are on new proteins not previously predicted to be *O*-glucosylated. Further studies need to be done to determine whether these additional proteins are in fact *O*-glucosylated.

Our data suggests that the predicted sites, including the novel site at EGF9, are modified at high stoichiometries. Very little unmodified peptide was typically observed, except at EGF 27 (Figure 6) or on EGF 4 when produced in COS7 cells (Figure 7). These results indicate that some EGF repeats may be more efficiently modified than others (*e.g.* comparison of EGF27 to others), and that efficiency may also be affected by cell type. Thus, the extent of modification at individual sites may be controlled by both the sequence of an individual EGF repeat as well as expression levels of Rumi in a given cell. Preliminary data suggests that Rumi levels do vary in different tissues (9).

Elongation of *O*-glucose past the monosaccharide is also quite efficient. Both the metabolically radiolabeling experiments and EIC comparisons of peptides with mono versus trisaccharide forms confirm that all sites are modified with trisaccharide. This is consistent with our previous work on full length endogenous Notch1 (76). These results suggest that the xylosyltransferases present in the cells examined here are also quite efficient. Of course, like Rumi, there may be tissue specific differences in their expression levels that alter the extent of elongation in a tissue specific or developmentally regulated fashion.

The tryptic peptide from EGF 15 and 16 contained multiple *O*-glycan modifications, including *O*-glucose trisaccharide (EGF16) and what appears to be the novel *O*-GlcNAc monosaccharide (on EGF15). *O*-GlcNAc has long been known to modify a variety of intracellular proteins, but it was recently reported on a non-intracellular protein for the first time when Okajima and co-workers observed this modification on *Drosophila* Notch (17). This is the first evidence of *O*-GlcNAc as an extracellular modification on a mammalian protein.

The mutagenesis studies suggest that elimination of most individual *O*-glucose sites (other than EGF 28) has very little effect on Notch1 activity. Thus, the major effects of *O*-glucose may be due to elimination of *O*-glucose at multiple sites, as would occur in *Rumi* mutants (8,9). It is somewhat surprising that elimination of the *O*-glucose site in EGF 12 has no effect on Notch activity in cell-based assays. EGF 12 is known to be part of the ligand-binding domain (56,71) and recent modeling studies suggest that the *O*-glucose modification site sits in the interface between human Notch1 and Jagged1 (72). The site is also present in most but not all known Notch receptors (Figure 3). The lack of an effect upon elimination of *O*-glucose at EGF 12 is in stark contrast to the elimination of *O*-fucose at EGF 12 (82,97).

The largest effect on Notch1 activity was observed when *O*-glucose at EGF 28 was eliminated. The site in EGF 28 is not highly conserved in Notch proteins from different species (Figure 3A), but is found in both human, mouse, and rat Notch1 (18), suggesting it may have mammalian-specific effects. In addition, the effect was seen only with Delta-like 1 as ligand. The EGF 28 mutant responded like wild type to activation with Jagged1. EGF 28 is in the *Abruptex* region of Notch, named for a series of mutations in *Drosophila* Notch that result in a hyperactive Notch refractory to Fringe (19,98). Recent structural studies have suggested that the presence of calcium-binding motifs reduces the flexibility of the linker between adjacent EGF

repeats (73). Most of the EGF repeats in Notch1 contain the calcium-binding motif, and as a result, the linkers are predicted to be rigid (Figure 12A). Interestingly, the *Abruptex* region has several non-calcium binding EGF repeats and may be a region of flexibility (Figure 12A). The fact that mutation of *O*-fucose sites (EGF 26 and 27) and an *O*-glucose site (EGF 28) within the *Abruptex* region affects Notch1 activity in cell-based assays suggests that these glycans may be affecting the flexibility of this region. We and others have previously proposed that the *O*-fucose glycans may affect the overall conformation of mouse Notch1 ECD (56,82). Similar proposals have been made regarding *Drosophila* Notch (56). The fact that the *O*-fucose and *O*-glucose glycans are fairly large with respect to the EGF repeat itself (Figure 12B) suggest that the presence or absence of these modifications could easily affect the overall conformation of the Notch ECD, especially in these flexible regions. Structural studies will be necessary to determine whether such conformational changes do indeed occur. Taken together, the individual *O*-glucose modifications seem to have distinct functions in mouse Notch ECD, which only adds to the complexity of glycan involvement in the Notch signaling pathway.

Table 1: Peptides from mouse Notch1 identified with *O*-glucose modifications. Peptides were identified by sequential neutral loss of masses corresponding to two xyloses and a glucose as shown in Figure 3. Spectra for each glycopeptide identified here are shown in Figures 5, 8, 9, and 15A-O. All masses were converted to the equivalent of singly charged ($M+H^+$) for the table. For each glycopeptide, the mass of the parent ion, the fully deglycosylated product (lacking all sugars, including fucose for those modified with both *O*-fucose and *O*-glucose, or HexNAc for EGF 15), and the difference between these (corresponding to the mass of the modification: *O*-glucose trisaccharide, 426 Da; *O*-glucose trisaccharide plus *O*-fucose monosaccharide, 572 Da; *O*-glucose trisaccharide plus *O*-HexNAc monosaccharide, 629 Da). The predicted mass of the unglycosylated peptide is also shown. All peptide masses are adjusted for carboxyamidomethylation of Cysteines. For peptides with a mass below 2000 Da, monoisotopic masses were used. For those above 2000 Da, average masses were used.

EGF	Sequence Consensus: CxSx(P/A)C	Parent Ion ($M+H^+$)	Deglyco Product ($M+H^+$)	Mass Δ	Pred. Mass ($M+H^+$)
2	^{57}CQDSNPCLSTPCK^{69}	1994.4	1566.8	427.6	1566.6
4	^{137}SCQQADPCASNPCANGGQCLPFESSYICR^{165}	3761.5	3336.7	426.9	3336.6
9	^{320}NCVCVNGWTGEDCSENIDDCASAACF^{345}	3441.4	3013.3	426.1	3013.2
10	^{366}TGLLCHLNDACISNPCNEGSNCDTNPVNGK^{395}	3759.8	3334.2	425.6	3333.7
12	^{456}CISNPCQNDATCLDQIGE^{473}	2669.8	2097.4	572.4	2096.2
13	^{483}EGVYCEINTDECASSPCL^{500}	2528.8	2102.5	426.3	2105.3
14	^{519}GFNGHLCQYDVDECASTPCK^{538}	2786.5	2359.9	426.6	2359.6
16	^{582}DGVATFTCLCQPGYTGHHCETNINECHSQPCR^{613}	4454.2	3823.4	630.8	3809.1
17	^{632}GTTGPNCEINLDDCASNPCDSGTCLDK^{658}	3401.8	2972.2	429.6	2973.2
19	^{714}LSEVNECNSNPCIHGACR^{731}	2544.4	2118.1	426.3	2118.3
20	^{739}CDCAPGWSGTNCDINNNECESNPCVNGGTCK^{769}	4121.8	3549.0	572.8	3549.8
21	^{781}EGFSGPNCQTNINECASNPCLNQGTCIDDVAGYK^{814}	4365.4	3791.8	573.6	3793.1
25	^{931}CDCLPGFQGAFCEEDINECASNPCQNGANCTDCVDSY^{967}	4781.4	4339.0	442.4*	4337.6
27	^{1021}DVNECDSRPCLHGGTCQDSYGTY^{1043}	3281.2	2709.1	572.1	2692.8
28	^{1063}CDSAPCKNGGRCW^{1075}	1996.9	1568.5	428.4	1567.6
31	^{1176}HGSNCSEEINECLSQPCQNGGTCIDLTNSY^{1205}	4021.4	3448.2	573.2	3446.6
33	^{1264}CEGDVNECLSNPCDPR^{1278}	2347.8	1922.0	425.8	1921.7
35	^{1333}ICRCPAGFEGATCENDARTCGSL^{1355}	2604.1	2604.1	0	2603.9

*Ion from EGF25 loses a water molecule in addition to *O*-glucose trisaccharide. See Figure 15K.

Table 2: Peptides from mouse Notch1 identified with *O*-fucose modifications. Peptides were identified by neutral loss of a mass corresponding to fucose as shown in Figure 5. Spectra for each glycopeptide identified here are shown in 15P-S. All masses were converted to the equivalent of singly charged ($M+H^+$) for the table. For each glycopeptide, the mass of the parent ion, the fully deglycosylated product and the difference between these (corresponding to the mass of the modification: *O*-fucose monosaccharide, 146 Da). The predicted mass of the unglycosylated peptide is also shown. All peptide masses are adjusted for carboxyamidomethylation of Cysteines. For peptides with a mass below 2000 Da, monoisotopic masses were used. For those above 2000 Da, average masses were used.

EGF	Sequence Cxxxx(S/T)C	Parent Ion ($M+H^+$)	Deglyco Product ($M+H^+$)	Mass Δ	Pred. Mass ($M+H^+$)
2	^{70}NAGTCHVVDHGGTVDYACSCPL GFSGPLCLTPLDNACLANPCR^{112}	4853.4	4707.0	146.4	4707.3
3	^{113}NGGTCDLLTLTEYK^{126}	1732.6	1587.2	146.4	1584.8
5	^{191}HGGTCHNEIGSYR^{203}	1635.1	1488.4	146.7	1487.6
35	^{1357}CLNGGTCISGPR^{1368}	1439.2	1292.6	146.6	1291.6

Table 3: Proteins predicted to be modified with *O*-glucose
The Swiss-Prot database was searched using the Motif search program at http://motif.genome.jp/MOTIF2.html using the following search strings in PROSITE format for EGF-like repeats containing the traditional, revised, and broader consensus sequences for *O*-glucosylation between the first and second Cysteines:

Traditional (CxSxPC):
C-{C}-{C}-P-C-{C} (1,70)-C-{C}(1,70)-C-{C} (1,6)-C-{C} (2)-G-[FYW]- {C} (0,24)-C-x.
and
C-{C}-S-{C}-P-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (2)-G-{C} (0,24)-C-x.
Alanine instead of Proline (CxSxAC):
C-{C}-S-{C}-A-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (2)-G-[FYW]- {C} (0,24)-C-x.
and
C-{C}-S-{C}-A-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (2)-G-{C} (0,24)-C-x.
Broader (CxSxxC):
C-{C}-S-{C}- (99)-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (2)-G-[FYW]- {C} (0,24)-C-x. And
C-{C}-S-{C}- (99)-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (2)-G-{C} (0,24)-C-x.
And
C-{C}-S-{C}- (99)-C-{C} (1,70)-C-{C} (1,70)-C-{C} (1,6)-C-{C} (3)-[FYW]- {C} (0,24)-C-x.
The search results were then culled for homologs between species with preference in the order human, mouse, other. The list was then sorted alphabetically by species and then protein name.

Protein Name	Species	Accession Number	Number of Predicted Sites		
			CXSXPC	CXSXAC	CXSXXC
Agrin	***Homo sapiens***	**O00468**	**1**		
C1QR1	***Homo sapiens***	**Q9NPY3**	**1**		
CELR1	***Homo sapiens***	**Q9NYQ6**	**1**		
CELR2	***Homo sapiens***	**Q9HCU4**	**2**		**1**
CELR3	***Homo sapiens***	**Q9NYQ7**	**2**		
CRUM1	***Homo sapiens***	**P82279**	**7**		**1**
CRUM2	***Homo sapiens***	**Q5IJ48**	**6**		
Cubilin	***Homo sapiens***	**O60494**	**1**		
Dickkopf-2	***Homo sapiens***	**Q9UBU2**			**1**
DLK-1	***Homo sapiens***	**P80370**	**1**		
DLK-2	***Homo sapiens***	**Q6UY11**	**1**		
Delta-like 1	***Homo sapiens***	**O00548**	**2**		
Delta-like 4	***Homo sapiens***	**Q9NR61**	**3**		
DNER	***Homo sapiens***	**Q8NFT8**	**3**		
Protein Eyes Shut Homolog	***Homo sapiens***	**Q5T1H1**	**3**		
Coagulation Factor IX	***Homo sapiens***	**P00740**	**1**		
Fibulin-1	***Homo sapiens***	**P23142**	**1**		
Fibulin-7	***Homo sapiens***	**Q53RD9**	**1**		
Fibrillin-2	***Homo sapiens***	**P35556**	**1**		
Fibrillin-3	***Homo sapiens***	**Q75N90**	**1**		
HEG1	***Homo sapiens***	**Q9ULI3**	**1**		

HGFA	***Homo sapiens***	**Q04756**	**1**		
Jagged-1	***Homo sapiens***	**P78504**	**3**		
Jagged-2	***Homo sapiens***	**Q9Y219**	**2**		
Notch1	***Homo sapiens***	**P46531**	**9**	**1**	**1**
Notch2	***Homo sapiens***	**Q04721**	**12**		
Notch3	***Homo sapiens***	**Q9UM47**	**8**	**1**	**1**
Notch4	***Homo sapiens***	**Q99466**	**9**		**1**
NT2NL9	***Homo sapiens***	**Q7Z3S9**	**1**		
Otogelin	***Homo sapiens***	**Q6ZRI0**			**1**
PAMR1	***Homo sapiens***	**Q6UXH**	**1**		
SLIT1	***Homo sapiens***	**O75093**	**2**		
SLIT2	***Homo sapiens***	**O94813**	**2**		
SNED1	***Homo sapiens***	**Q8TER0**	**6**		
SVEP1	***Homo sapiens***	**Q4LDE5**	**4**		
Thrombospondin-1	***Homo sapiens***	**P07996**	**1**		
Thrombospondin-2	***Homo sapiens***	**P35442**	**1**		
Thrombospondin-4	***Homo sapiens***	**P35443**	**1**		
Vitamin K-dependent protein Z	***Homo sapiens***	**P22891**	**1**		
Agrin	*Mus musculus*	**A2ASQ1**	**1**		
C1QR1	*Mus musculus*	**O89103**	**1**		
CELR1	*Mus musculus*	**O35161**	**1**		**1**
CELR2	*Mus musculus*	**Q9R0M0**			**1**
CELR3	*Mus musculus*	**Q91ZI0**	**2**		
Coagulation Factor X	*Mus musculus*	**O88947**	**1**		
Coagulation Factor IX	*Mus musculus*	**P16294**	**1**		
CRUM1	*Mus musculus*	**Q8VHS2**	**7**		**1**
CRUM2	*Mus musculus*	**Q80YA8**	**6**		
Delta-like 1	*Mus musculus*	**Q61483**	**2**		
Delta-like 3	*Mus musculus*	**O88516**	**1**		
Delta-like 4	*Mus musculus*	**Q9JI71**	**3**		
DLK-1	*Mus musculus*	**Q09163**	**1**		
DLK-2	*Mus musculus*	**Q8K1E3**	**1**		
DNER	*Mus musculus*	**Q8JZM4**	**3**		
Fibrillin-2	*Mus musculus*	**Q61555**	**1**		
Fibulin-7	*Mus musculus*	**Q501P1**	**1**		
HGFA	*Mus musculus*	**Q9R098**	**1**		
Jagged-1	*Mus musculus*	**Q9QXX0**	**3**		
Jagged-2	*Mus musculus*	**Q9QYE5**	**1**		
LRP2	*Mus musculus*	**A2ARV4**			**1**
Milk fat globule-EGF factor 8	*Mus musculus*	**P21956**			**1**
Notch1	*Mus musculus*	**Q01705**	**10**	**1**	**2**
Notch2	*Mus musculus*	**O35516**	**11**		
Notch3	*Mus musculus*	**Q61982**	**11**		
Notch4	*Mus musculus*	**P31695**	**7**		**1**
Otogelin	*Mus musculis*	**O55225**			**1**

PAMR1	*Mus musculus*	**Q8BU25**	**1**		
Vitamin K-dependent protein Z	*Mus musculus*	**Q9CQW3**	**1**		
SLIT1	*Mus musculus*	**Q80TR4**	**2**		
SLIT2	*Mus musculus*	**Q9R1B9**	**2**		
SNED1	*Mus musculus*	**Q70E20**	**7**		
SSPO	*Mus musculus*	**Q8CG65**			**1**
Stabilin-1	*Mus musculus*	**Q8R4Y4**			**1**
SVEP1	*Mus musculus*	**A2AVA0**	**5**		
Thrombospondin-1	*Mus musculus*	**P35441**	**1**		
Thrombospondin-2	*Mus musculus*	**Q03350**	**1**		
Thrombospondin-4	*Mus musculus*	**Q9Z1T2**	**1**		
VWCE	*Mus musculus*	**Q3U515**			**1**
Crumbs	*Drosophila*	**P10040**	**3**		
Delta	*Drosophila*	**P10041**	**1**		
Protein Eyes Shut	*Drosophila*	**A0A1F4**	**5**		
FAT	*Drosophila*	**P33450**	**1**		
Notch	*Drosophila*	**P07207**	**15**		**3**
STAN	*Drosophila*	**Q9V5N8**	**1**		
Agrin	*Gallus gallus*	**P31696**	**1**		**1**
Agrin	*Discopyge ommata*	**Q90404**	**1**		
Agrin	*Rattus norvegicus*	**P25304**	**1**		
C1QR1	*Rattus norvegicus*	**Q9ET61**	**1**		
Cadherin-4	*C. elegans*	**Q19319**	**1**		
CELR2	*Rattus norvegicus*	**Q9QYP2**	**2**		**1**
CELR3	*Rattus norvegicus*	**O88278**	**2**		
Coagulation Factor X	*Gallus gallus*	**P25155**	**1**		
Coagulation Factor X	*O. caniculus*	**O19045**	**1**		
Coagulation Factor X	*Rattus norvegicus*	**Q63207**	**1**		
Coagulation Factor X	*T. carinatus*	**Q4QXT9**	**1**		
Coagulation Factor VII	*Bos taurus*	**P22457**	**1**		
Coagulation Factor IX	*Bos Taurus*	**P00741**	**1**		
Coagulation Factor X	*C. familiaris*	**P19540**	**1**		
Coagulation Factor X	*Gallus gallus*	**Q804X6**	**1**		
Coagulation Factor X	*Felis catus*	**Q6SA95**	**1**		
Coagulation Factor X	*Pan troglodytes*	**Q95ND7**	**1**		
COMC	*D. discoideum*	**Q55AP8**			**1**
CSPG2	*Gallus gallus*	**Q90953**	**1**		
Cubilin	*C. elegans*	**Q20911**	**1**		
Cubilin	*Canis familiaris*	**Q9TU53**	**1**		
Delta-like 3	*Rattus norvegicus*	**O88671**	**1**		
Delta-like A	*Danio rerio*	**Q6DI48**	**1**		
Delta-like C	*Danio rerio*	**Q9IAT6**	**3**		
Delta-like D	*Danio rerio*	**Q8UWJ4**	**1**		
Fibulin-1	*C. elegans*	**Q8MJJ9**	**1**		**2**
Fibulin-1	*Gallus gallus*	**O73775**	**1**		

Fibropellin-1	*S. purpuratus*	**P10079**	**17**		
Fibropellin-3	*S. purpuratus*	**P49013**	**6**		
FP2	*M. galloprovincialis*	**Q25464**	**1**		
GLP-1	*C. elegans*	**Q9ULI3**	**1**		
HGFA	*Canis familiaris*	**Q6QNF4**	**1**		
Jagged-1a	*Danio rerio*	**Q90Y57**	**3**		
Jagged-1b	*Danio rerio*	**Q90Y54**	**3**		
Jagged-1	*Rattus norvegicus*	**Q63722**	**2**		
Jagged-2	*Rattus norvegicus*	**P97607**	**1**		
Lin-12	*C. elegans*	**P14585**	**2**		
LRP	*C. elegans*	**Q04833**			**1**
LRP2	*Rattus norvegicus*	**P98158**			**1**
MFGM	*Bos taurus*	**Q95114**			**1**
MFGM	*Sus scrofa*	**P79385**			**1**
MFGM	*Rattus norvegicus*	**P70490**			**1**
Micronemal protein 6	*T. gondii*	**Q9XYH7**	**1**		
MUC13	*Rattus norvegicus*	**P978811**	**1**		
Notch1	*Danio rerio*	**P46530**	**10**	**1**	**2**
Notch1	*Rattus norvegicus*	**Q07008**	**11**	**1**	**1**
Notch2	*Rattus norvegicus*	**Q9QW30**	**10**		
Notch3	*Rattus norvegicus*	**Q9R172**	**10**		
Notch	*Xenopus laevis*	**P21783**	**11**		**2**
PAMR1	*Bos Taurus*	**Q5E9P5**	**1**		
PAMR1	*Pongo abelii*	**Q5RDI1**	**1**		
PAMR1	*Xenopus tropicalis*	**Q6DIV5**	**1**		
Protein Cueball	*Aedis aegypti*	**Q16YE7**			**1**
Protein Cueball	*C. pungens*	**B0WH58**			**1**
SLIT1	*Rattus norvegicus*	**O88279**	**2**		
SNED1	*Rattus norvegicus*	**Q5ZQU0**	**7**		
SP63	*S. purpuratus*	**Q07929**	**1**		
SSPO	*Rattus norvegicus*	**Q700K0**			**1**
SUREJ	*S. purpuratus*	**Q26627**	**1**		
SVEP1	*Rattus norvegicus*	**P0C6B8**	**5**		
Thrombospondin-1	*Bos taurus*	**Q28178**	**1**		
Thrombospondin-1	*Xenopus laevis*	**P35448**	**1**		
Thrombospondin-2	*Bos taurus*	**Q95116**	**1**		
Thrombospondin-2	*Gallus gallus*	**P35440**	**1**		
Thrombospondin-4	*Bos taurus*	**Q3WW8**	**1**		
Thrombospondin-4	*Rattus norvegicus*	**P49744**	**1**		
Vitamin K-dependent protein Z	*Bos taurus*	**P00744**	**1**		
Vitellogenin receptor	*Solenopsis invicta*	**Q6X0I2**			**1**
VWA2	*Xenopus laevis*	**Q6DCQ6**	**1**		
WAKLO	*Arabidopsis thaliana*	**Q8RY67**			**1**

Table 4: Primers for site-directed mutagenesis

EGF repeat	Sequence
2 (fwd)	CCTTGCCTCGCCACACCGTGTAAGAATGCTGGAACGTG
2 (rev)	ACACGGTGTGGCGAGGCAAGGATTGGGGTCCTGGC
4 (fwd)	GACCCCTGTGCCGCCAACCCCTGTGCC
4 (rev)	GGCACAGGGGTTGGCGGCACAGGGGTC
9 (fwd)	ACTGTGCCGCTGCCGCCTGTTTCCAGGGTG
9 (rev)	AGGCGGCAGCGGCACAGTCATCAATGTTCTC
10 (fwd)	CATGCGTGCATCGCCAACCCCTGCAAC
10 (rev)	GTTGCAGGGGTTGGCGATGCACGCATG
12 (fwd)	AGTGCATCGGCCAACCCATGTCAGAATGATGCCACTTGC
12 (rev)	TGACATGGGTTGGCCGATGCACTCATTAACATCAATCTC
13 (fwd)	GAGTGCGCCGCCAGCCCCTGTCTGCACAATGGCC
13 (rev)	GACAGGGGCTGGCGGCGCACTCATCCGTGTTG
14 (fwd)	GAGTGTGCCGCCACACCATGCAAGAACGGTGC
14 (rev)	ATGGTGTGGCGGCACACTCATCCACATCATACTGGCA
16 (fwd)	AGTGCCACGCCCAACCGTGCCGCCATG
16 (rev)	ACGGTTGGGCGTGGCACTCATTGATGTTGGTCTC
17 (fwd)	GACTGCGCCGCCAACCCCTGTGACTCTGGCACCTG
17 (rev)	CACAGGGGTTGGCGGCGCAGTCATCCAGGTTGATC
19 (fwd)	GAGTGCAACGCTAACCCCTGCATCCACGGAGCTTGCC
19 (rev)	GCAGGGGTTAGCGTTGCACTCGTTGACCTCGG
20 (fwd)	AACGAGTGTGAGGCCAACCCTTGTGTC
20 (rev)	GACACAAGGGTTGGCCTCACACTCGTT
21 (fwd)	GAATGTGCCGCCAACCCCTGCCTGAACCAGGGGACCTG
21 (rev)	GCAGGGGTTGGCGGCACATTCGTTGATGTTGGTC
25 (fwd)	GAATGTGCCGCCAATCCCTGCCAAAATGGTGCCAATTG
25 (rev)	GCAGGGATTGGCGGCACATTCATTGATGTCCTC
27 (fwd)	AATGAGTGTGATGCCCGGCCCTGTCTG
27 (rev)	CAGACAGGGCCGGCCATCACACTCATT
28 (fwd)	TGGTGCGACGCGGCTCCCTGCAAGAATGGTGGCAGG
28 (rev)	GCAGGGAGCCGCGTCGCACCACCAGCGCACAAGG
31 (fwd)	GAGTGCCTGGCCCAGCCCTGCCAGAATGGGGGTACC
31 (rev)	GCAGGGCTGGGCCAGGCACTCGTTGATCTCCTCTG
33 (fwd)	AATGAATGTCTTGCCAACCCCTGTGAC
33 (rev)	GTCACAGGGGTTGGCAAGACATTCATT

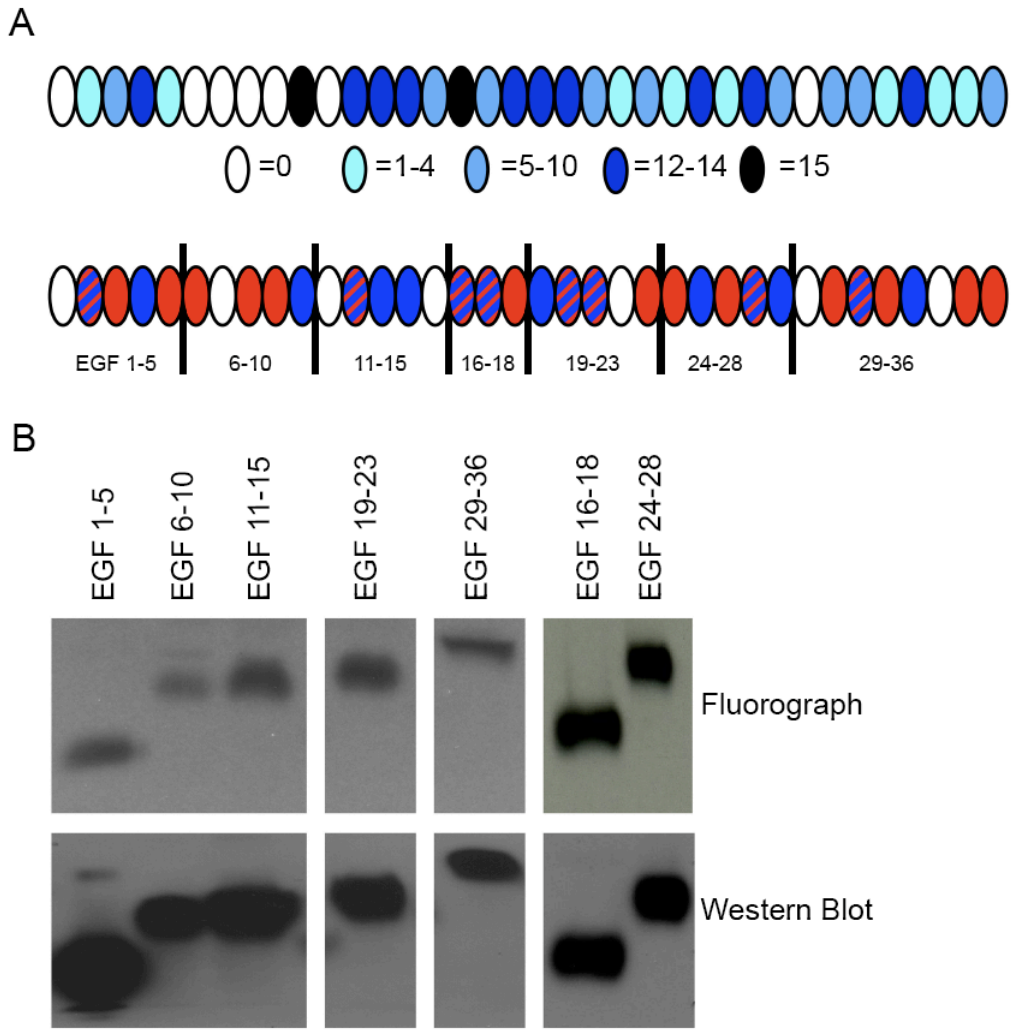


Figure 3. Expression and *O*-glucosylation of EGF fragments from mouse Notch1. **A.** Domain map showing predicted *O*-glucose sites on Notch. The top panel shows degree of conservation of predicted *O*-glucose sites (C^1x<u>S</u>xPC^2) on 15 Notch proteins containing 36 EGF repeats. The intensity of the shading correlates with the degree of conservation (modified from Haines and Irvine (100)). The lower panel shows the predicted *O*-glucose sites in mouse Notch1. Blue ovals indicate EGF repeats bearing the consensus sequence for *O*-glucosylation, red ovals indicate those with predicted sites of *O*-fucosylation (C^2xxxx(<u>S/T</u>)C^3), and blue and red hatched ovals bear both consensus sequences. The vertical lines demarcate the fragments expressed in Lec1-CHO cells. **B.** Plasmids encoding 7 Notch fragments of mouse Notch1 were transfected into Lec1-CHO cells, the cells were metabolically radiolabeled with [^{3}H]-galactose, and the secreted fragments were purified from the conditioned media as described in Experimental Procedures. Based on radioactivity, approximately equivalent amounts of each fragment were loaded on two SDS-PAGE gels and analyzed by fluorography (top panel) and Western blot using anti-myc antibodies (bottom panel). The left three panels were experiments performed by Dr. Aleksandra Nita-Lazar.

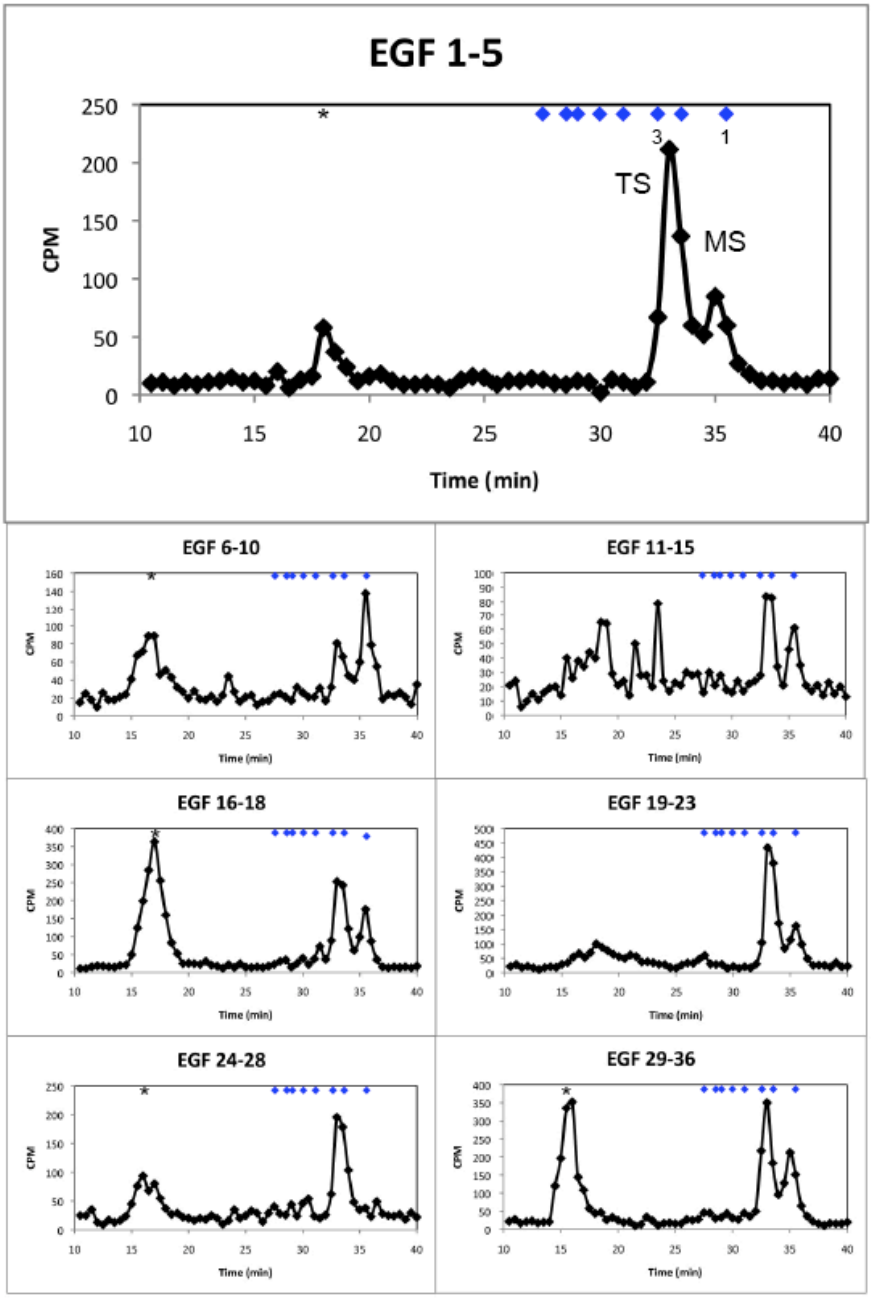


Figure 4. All Notch1 fragments are modified with *O*-linked glucose.
O-linked saccharides were released from purified mouse Notch1 fragments (as indicated at the top of each panel) expressed in Lec1-CHO cells radiolabeled with [^{3}H]-galactose by alkali-induced β-elimination. The β-eliminated glycans were separated on a Pharmacia Superdex sizing column calibrated with partially hydrolyzed dextran standards. Positions of glucose units from dextran standard are indicated on top of each graph with blue diamonds. MS, monosaccharide peak, TS, trisaccharide peak. * indicates high molecular weight species containing [^{3}H]galactose but not [^{3}H]glucitol following acid hydrolysis (thus not derived from *O*-glucose glycans). These experiments were performed by Dr. Aleksandra Nita-Lazar.

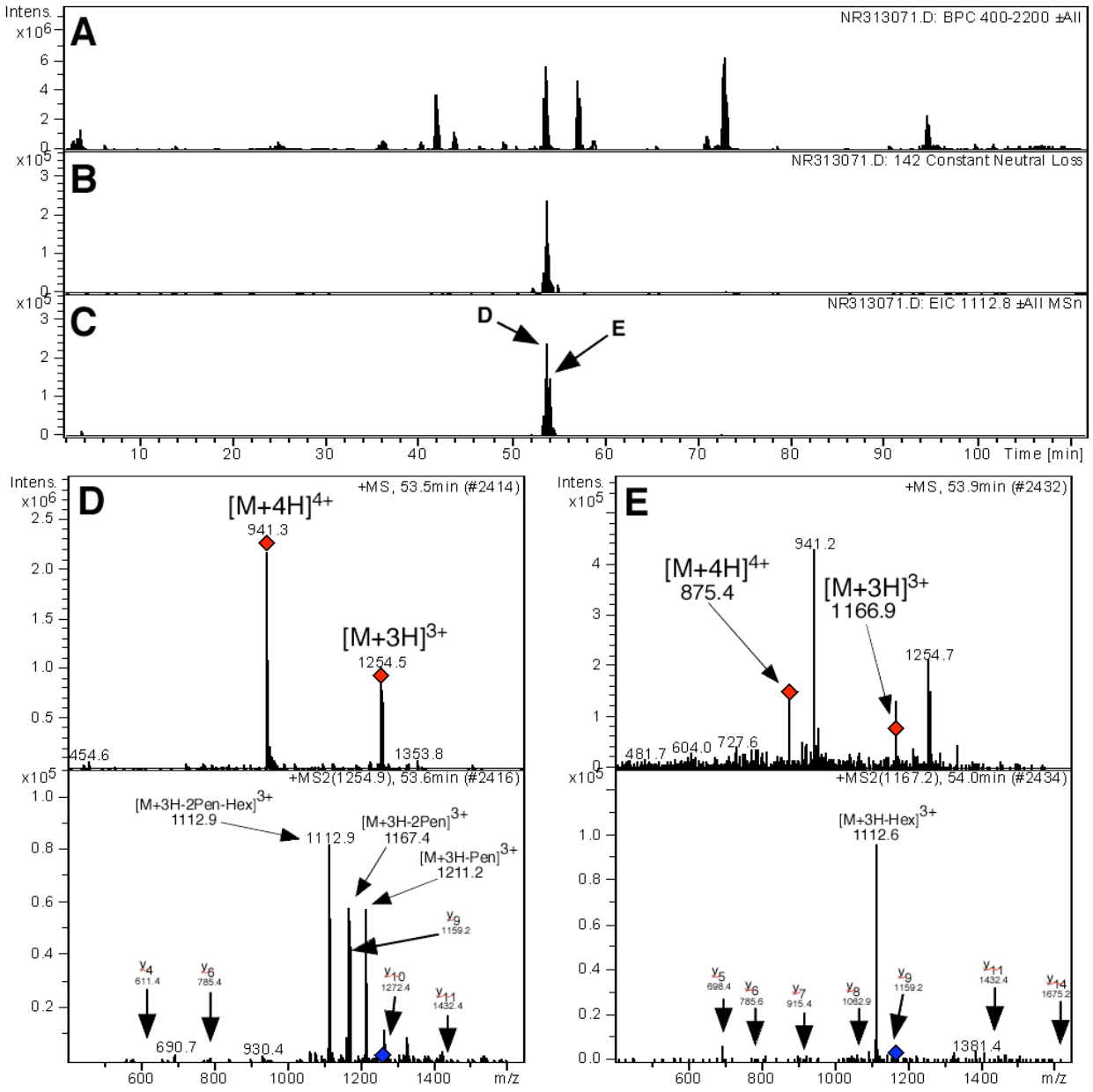

A
NR313071.D: BPC 400-2200 ±All
B
NR313071.D: 142 Constant Neutral Loss
C
NR313071.D: EIC 1112.8 ±All MSn
D
E
Time [min]
D
+MS, 53.5min (#2414)
[M+4H]4+
941.3
[M+3H]3+
1254.5
454.6
1353.8
+MS2(1254.9), 53.6min (#2416)
[M+3H-2Pen-Hex]3+
1112.9
1112.9
[M+3H-2Pen]3+
1167.4
[M+3H-Pen]3+
1211.2
y9
1159.2
y4
611.4
y6
785.4
y10
1272.4
y11
1432.4
690.7
930.4
m/z
E
+MS, 53.9min (#2432)
941.2
[M+4H]4+
875.4
[M+3H]3+
1166.9
1254.7
481.7
604.0
727.6
+MS2(1167.2), 54.0min (#2434)
[M+3H-Hex]3+
1112.6
y5
698.4
y6
785.6
y7
915.4
y8
1062.9
y9
1159.2
y11
1432.4
y14
1675.2
1381.4
m/z

Figure 5. Identification of an *O*-glucosylated peptide from EGF1-5 of mouse Notch1 by LC-MS/MS. EGF1-5 from mouse Notch1 was produced in Lec1-CHO cells and purified from the medium as described in Experimental Procedures. The protein was reduced, alkylated, purified by SDS-PAGE, and subjected to in-gel tryptic digestion. The resulting peptides were analyzed by LC-MS/MS as described in Experimental Procedures. **A.** Base Peak Chromatogram (BPC) showing the most abundant ions in each MS scan. **B.** Constant neutral loss chromatogram showing all ions that lost 142 Da upon CID fragmentation. This search reveals any triply charged peptide bearing an *O*-glucose trisaccharide. The *O*-glucose trisaccharide has a mass of 426 Da, thus loss from a triply charged peptide is one third, or 142 Da. **C.** Extracted ion chromatogram (EIC) showing the presence of m/z 1112.8 ion in any MS/MS spectrum. This search reveals any spectrum where fragmentation results in the formation of the m/z 1112.8 ion. Two species were identified, designated "D" and "E". The spectra for these species are shown in panels D and E, respectively. **D.** MS (top) and MS/MS (bottom) spectra of the ion identified in the constant neutral loss search (panel B) and in the EIC search (D in panel C) eluting at 53.5 min of the chromatogram. The mass spectrometer selects the two most abundant ions for CID fragmentation in each MS scan, and these ions are identified by red diamonds. M corresponds to the mass for the peptide plus the *O*-glucose trisaccharide. One of the selected ions corresponds to the triply charged form of this glycopeptide ($[M+3H]^{3+}$, m/z 1254.5), and the other corresponds to the quadruply charged form of the same glycopeptide ($[M+4H]^{4+}$, m/z 941.3). The bottom panel shows the MS/MS spectrum resulting from CID fragmentation of the m/z 1254.5 ion from the top panel. The location of the parent ion (m/z 1254.5) prior to fragmentation is indicated in the MS/MS spectrum with a blue diamond. Ions corresponding to the sequential losses of a pentose ($[M+3H-Pen]^{3+}$), a second pentose ($[M+3H-2Pen]^{3+}$), and a hexose ($[M+3H-2Pen-Hex]^{3+}$), from the parent ion are indicated. The major product ion ($[M+3H-2Pen-Hex]^{3+}$, m/z 1112.9) matches the predicted mass for the triply charged form of ^{137}SCQQADP**C**A**S**N**PC**ANGGQCLPFESSYICR^{165}, a tryptic peptide from EGF repeat 4 that contains an *O*-glucose consensus sequence (see Table 1 for predicted masses). No other predicted tryptic peptides from EGF1-5 of mouse Notch1 correspond to this mass. Further confirmation for the assignment of the peptide comes from peptide fragment ions. Several y-ions from fragmentation of the peptide are shown here. **E.** MS (top) and MS/MS (bottom) spectra of the ion identified in the EIC search (E in panel C) eluting at 53.9 min of the chromatogram. M corresponds to the mass for the peptide plus the *O*-glucose monosaccharide. One of the selected ions corresponds to the triply charged form of this glycopeptide ($[M+3H]^{3+}$, m/z 1166.9), and the other corresponds to the quadruply charged form of the same glycopeptide ($[M+4H]^{4+}$, m/z 875.4). The bottom panel shows the MS/MS spectrum resulting from CID fragmentation of the m/z 1166.9 ion from the top panel. The location of the parent ion prior to fragmentation is indicated in the MS/MS spectrum with a blue diamond. An ion corresponding to the loss of a hexose ($[M+3H-Hex]^{3+}$) from the parent ion is indicated. This major product ion ($[M+3H-Hex]^{3+}$, m/z 1112.8) matches the predicted mass for the triply charged form of the same tryptic peptide from EGF repeat 4 identified in panel D (96).

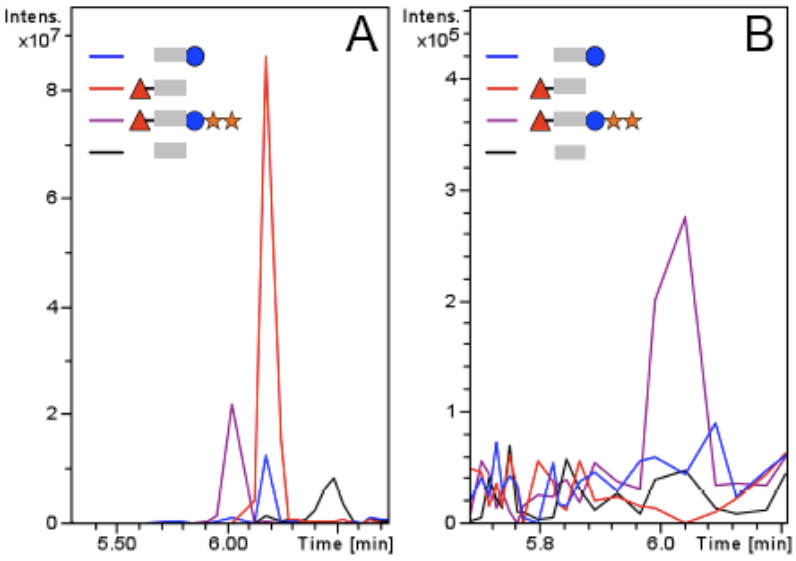


Figure 6. Extent of *O*-glucosylation is site specific.
A. Extracted ion chromatogram (EIC) traces of the ions corresponding to the naked (black), *O*-fucose alone (red), *O*-glucose alone (blue), and *O*-fucose + *O*-glucose trisaccharide (purple) forms of a peptide from EGF27 reveal that the major species is *O*-fucosylated with no *O*-glucose modification. **B.** EGF 12 also bears dual *O*-glycosylation sites, but unlike EGF 27, EIC traces show that the major species here is fully modified with *O*-fucose monosaccharide and O-glucose trisaccharide. Both *O*-glycan monosaccharide and naked peptide species are not detected, indicating both sites are extensively modified. Symbols: peptides, gray line; glucose, blue circle; xylose, orange star; fucose, red triangle.

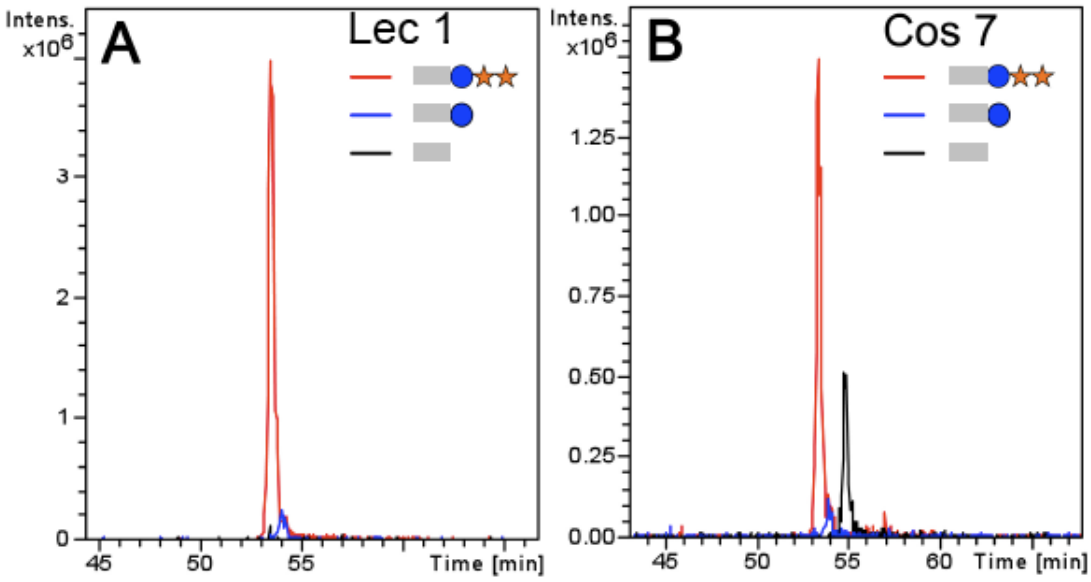


Figure 7. Extent of *O*-glucosylation is cell-line dependent.
Mouse Notch1 EGF 1-5 was over-expressed and purified from Lec1-CHO and COS7 cells and analyzed by LC-MS/MS as described in Figure 3. **A.** EIC searches for the ions corresponding to a peptide from EGF 4 modified with *O*-glucose trisaccharide (red), monosaccharide (blue), and naked peptide (black) show EGF4 is highly *O*-glucosylated and extended to trisaccharide in Lec1-CHO cells, with naked peptide and monosaccharide species at almost undetectable levels, indicating glucosylation and xylosylation is occurring at high efficiency. **B.** Parallel EIC searches for ions corresponding to the same EGF 4 peptide (trisaccharide (red), monosaccharide (blue), and naked peptide (black)) reveal a significant amount of naked peptide in COS7 cells, indicating under-glucosylation of EGF4 in these cells. Monosaccharide species is almost undetectable here, suggesting xylosylation is efficient. Symbols: peptides, gray line; glucose, blue circle; xylose, orange star.

320 NCVCVNGWTGEDCSENIDDCASAACF 345

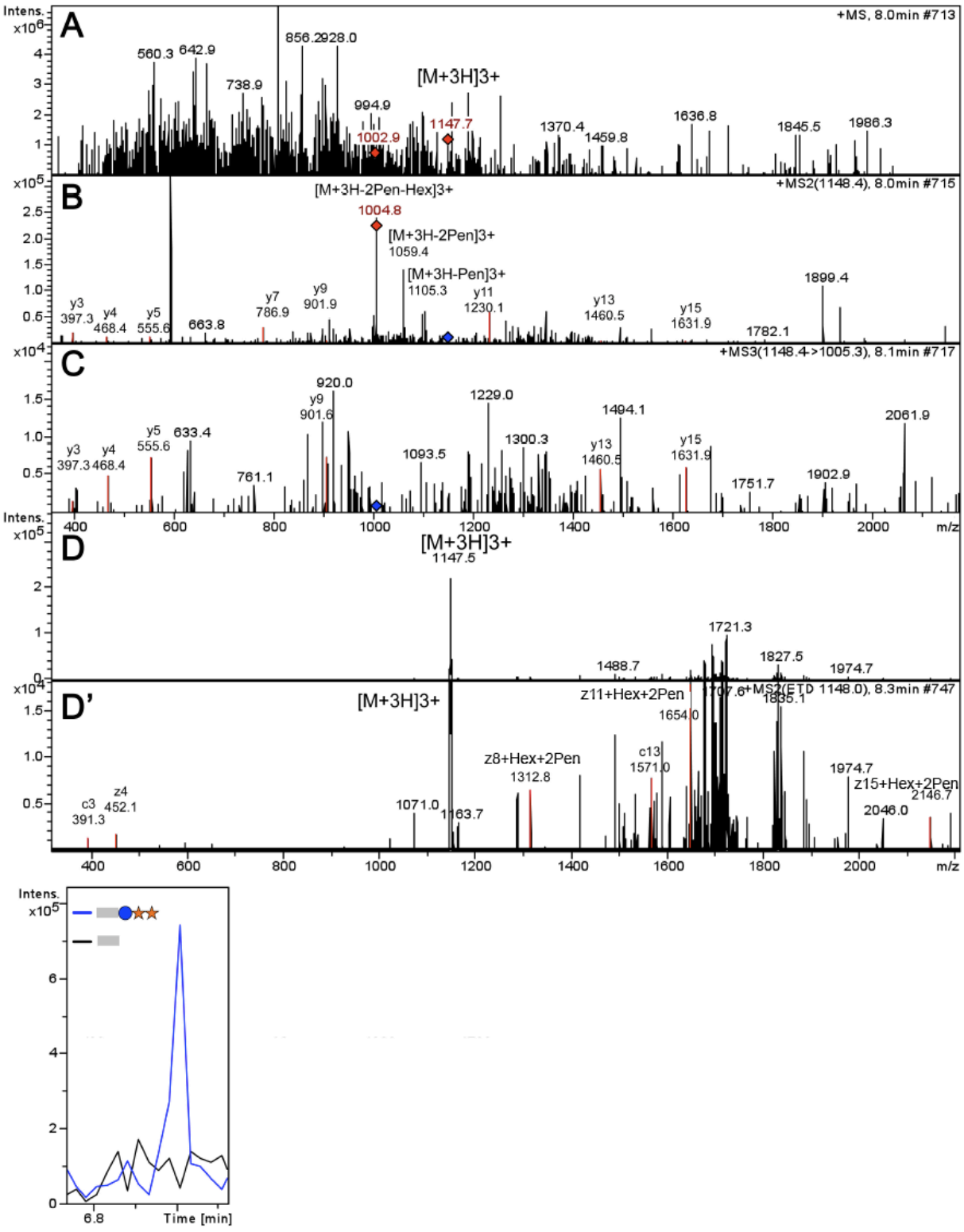


Figure 8. A non-traditional consensus site is modified with *O*-glucose trisaccharide at EGF 9.

Mouse Notch1 EGF 1-18 was subjected to reduction, alkylation, and in-solution digestion as described in Experimental Procedures. The resulting chymotryptic peptides were analyzed by LC-MS/MS. **A.** MS scan showing parent ion 1147.7 (red diamond) being selected for fragmentation. **B.** MS/MS reveals constant neutral loss of xyl-xyl-glc and identifies a peptide from EGF9 as the modified site. The blue diamond shows the position of the parent ion prior to fragmentation. The major fragments show the sequential loss of a Pentose ($[M+3H-Pen]^{3+}$, m/z 1105.3), a second Pentose ($[M+3H-2Pen]^{3+}$, m/z 1059.4), and a Hexose ($[M+3H-2Pen-Hex]^{3+}$, m/z 1004.8) from the parent ion. Further confirmation of this assignment comes from peptide fragment ions (y-ions, highlighted in red) in the MS/MS spectrum, several of which are indicated. **C.** CNL-triggered MS3 of parent ion confirms glycopeptide is from EGF 9 of mouse Notch1, with increased signal from y-ions (red). **D.** ETD fragmentation of the parent ion leaves glycopeptide intact, allowing confirmation of Ser modified with O-Hex-Pen-Pen with glycopeptide fragment ions (**D'** shows increased scale with c and z-ions highlighted in red). **E.** Extracted Ion Chromatograms of ions corresponding to the O-glucose trisaccharide (blue) and naked peptide (black) indicate this site is highly modified. Symbols: peptides, gray line; glucose, blue circle; xylose, orange star.

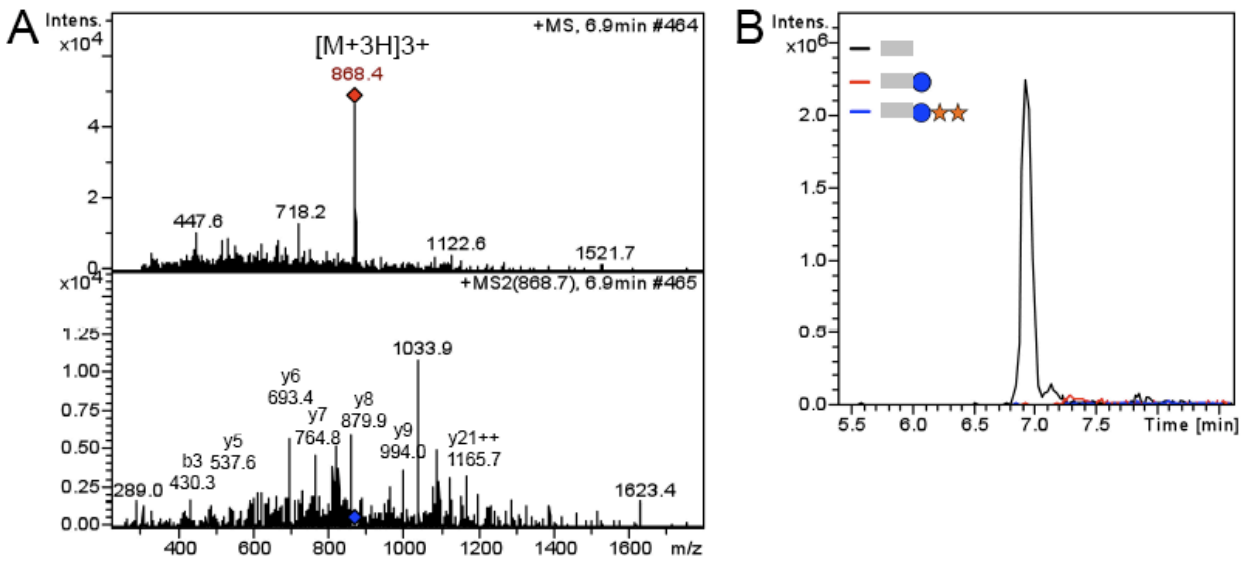


Figure 9. EGF 35 is not *O*-glucosylated
Mouse Notch1 EGF 29-36 was reduced, alkylated, and subjected to in-solution chymotryptic digest. The resulting peptides were analyzed by nano-LC-MS/MS as described in Experimental Procedures. **A.** Top panel: MS scan showing parent ion 868.4 (red diamond) being selected for fragmentation. Bottom panel: MS/MS of the parent ion, where the blue diamond shows the position of the parent ion prior to fragmentation. Peptide fragment ions (b- and y-ions) in the MS/MS spectrum confirm peptide is from EGF 35. **B.** Extracted Ion Chromatogram traces of ions corresponding to the trisaccharide (blue), monosaccharide (red), and naked peptide (black) show the only species present is unmodified naked peptide. Symbols: peptides, gray line; glucose, blue circle; xylose, orange star.

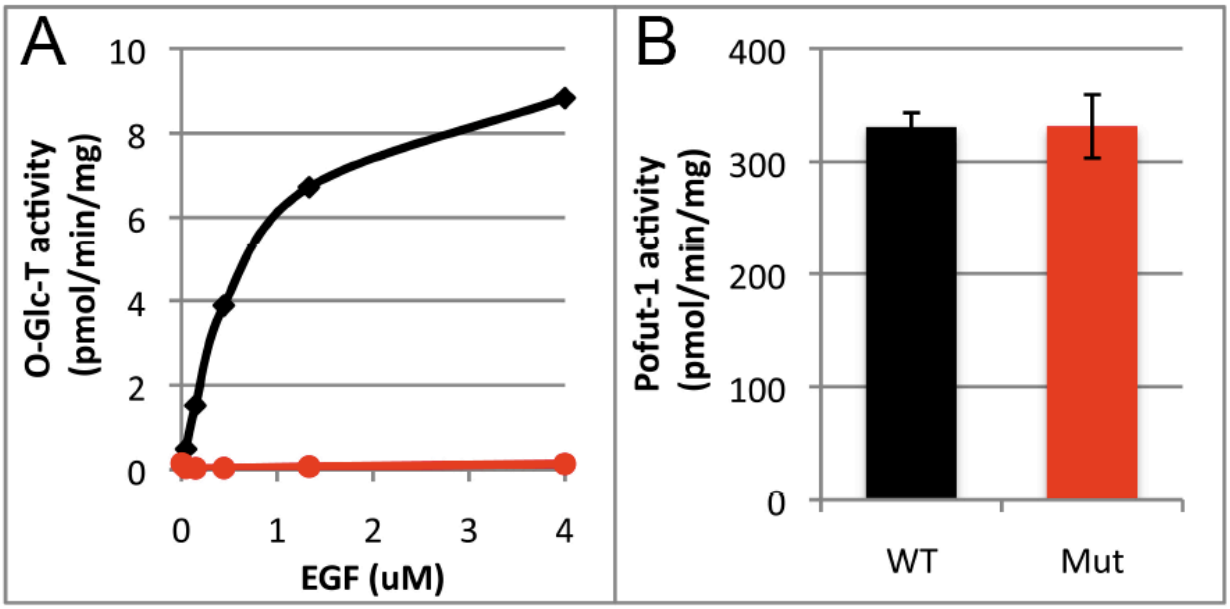
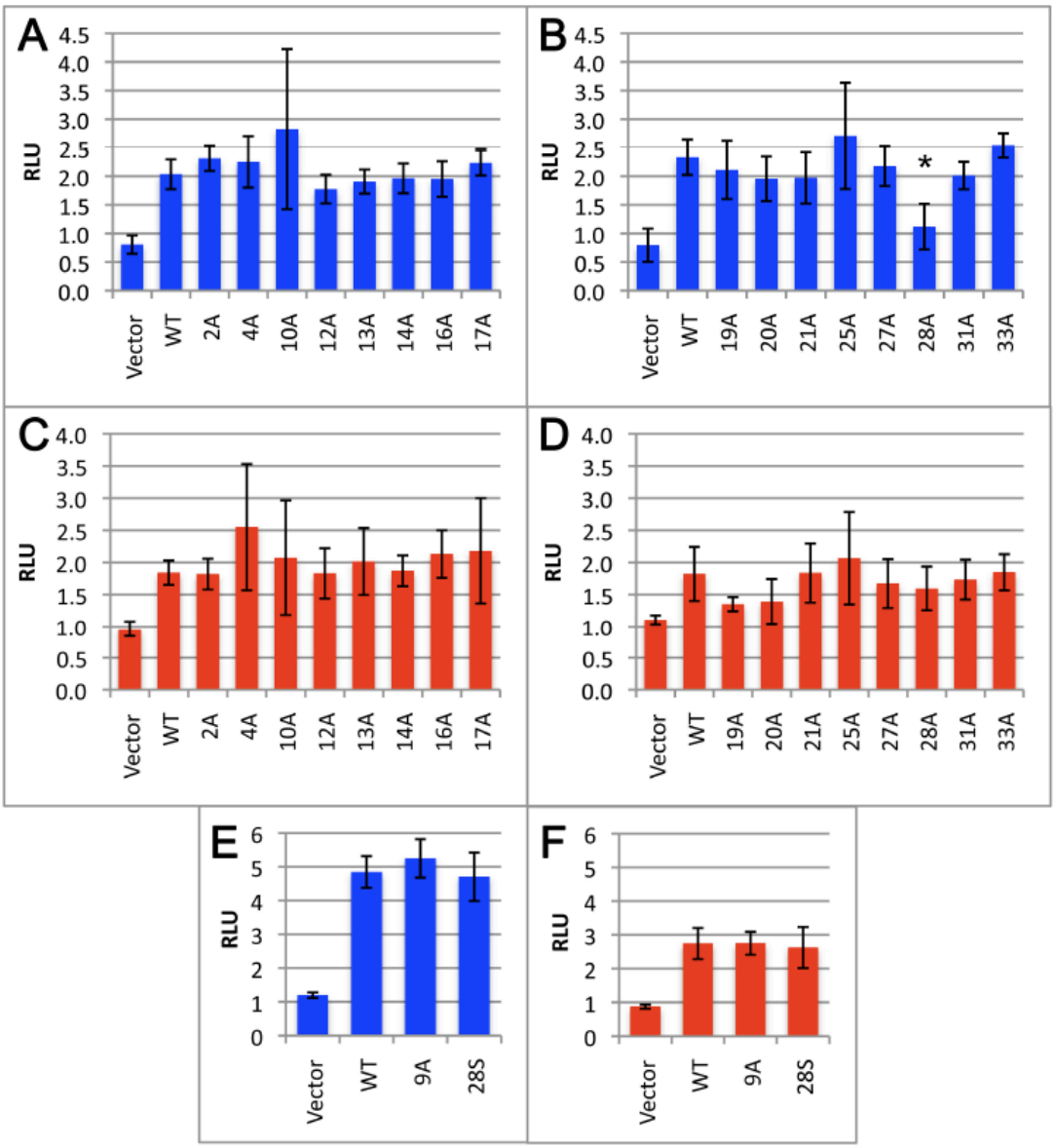


Figure 10. *O*-Glucose only modifies Serine, not Threonine, on EGF repeats. A. Protein *O*-glucosyltransferase assays were performed as described (53) using crude lysates of mouse brain as enzyme source and increasing amounts of bacterially expressed factor VII EGF repeat, or a mutated form of the EGF repeat, S52T, where the S in the consensus *O*-glucose consensus sequence was replaced with a T, converting $C^1A\underline{S}SPC^2$ to C^1ATSPC^2. **B.** Pofut1 assays were performed as described (79) using 4 μM wild-type or mutated EGF repeat. All assays were performed in duplicate. Error bars represent the range of duplicates. This experiment was performed by Dr. Hideyuki Takeuchi.


A
RLU
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Vector
WT
2A
4A
10A
12A
13A
14A
16A
17A
B
RLU
*
Vector
WT
19A
20A
21A
25A
27A
28A
31A
33A
C
RLU
Vector
WT
2A
4A
10A
12A
13A
14A
16A
17A
D
RLU
Vector
WT
19A
20A
21A
25A
27A
28A
31A
33A
E
RLU
6
5
4
3
2
1
0
Vector
WT
9A
28S
F
RLU
Vector
WT
9A
28S

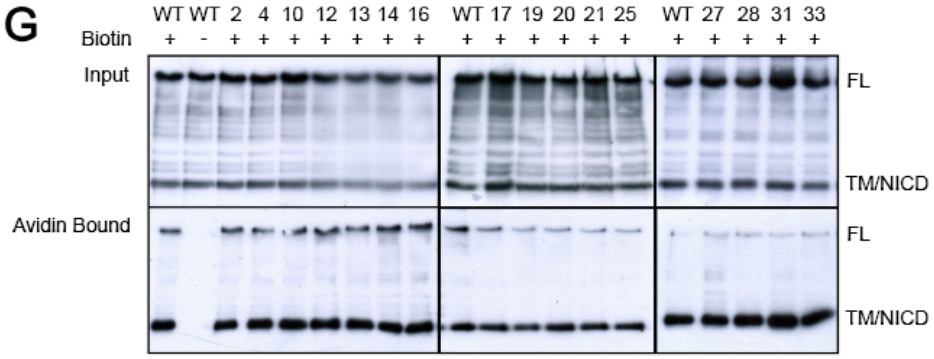
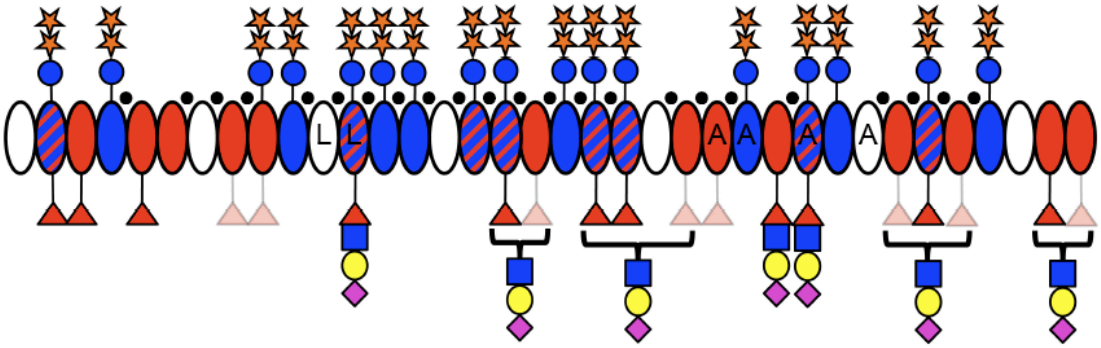

G
Biotin
WT WT 2 4 10 12 13 14 16 WT 17 19 20 21 25 WT 27 28 31 33
+ - + + + + + + + + + + + + + + + + +
Input
FL
TM/NICD
Avidin Bound
FL
TM/NICD

Figure 11. *O*-glucose site mutation in EGF 28 of mouse Notch1 alters Notch activation. Single *O*-glucose site mutants were evaluated in comparison to wild type mouse Notch1 in a co-culture signaling assay, where NIH3T3 cells were transiently transfected with mN1 plasmids, and co-transfected with luciferase to measure Notch1 activation by Delta-like1 (**A, B, E**) or Jagged1 (**C,D, F**). The results are expressed as Relative Luciferase Units (RLU), which reflects fold-activation induced by ligand expressing cells over that obtained with L cells. The co-cultures in which Notch bearing an *O*-glucose site mutant resulted in a statistically significant difference in RLU when compared to wildtype, as determined by one-way ANOVA analysis, are indicated with the symbol (*). To ensure no inadvertent mutations contributed to the decrease in Notch1 activation that resulted from the EGF 28 mutant, this mutant was reverted to wild type and used in the co-culture assays (**E, F**). To evaluate cell-surface expression, NIH3T3 cells were transiently transfected with wild type Notch1-Myc_6 or Notch1-Myc_6 *O*-glucose site mutants. 24h post-transfection, cell surface biotinylation was carried out as described (82). Cell lysates and streptavidin-bound fractions were analyzed by immunoblot using an anti-myc antibody to detect transfected Notch1(**G**). Immunoblotting with anti-PAN-cadherin and anti-ß-actin were used as positive and negative controls, respectively (not shown). Mutants were generated jointly with Dr. Aleksandra Nita-Lazar. These assays were carried out by Dr. Shinako Kakuda.

A



B

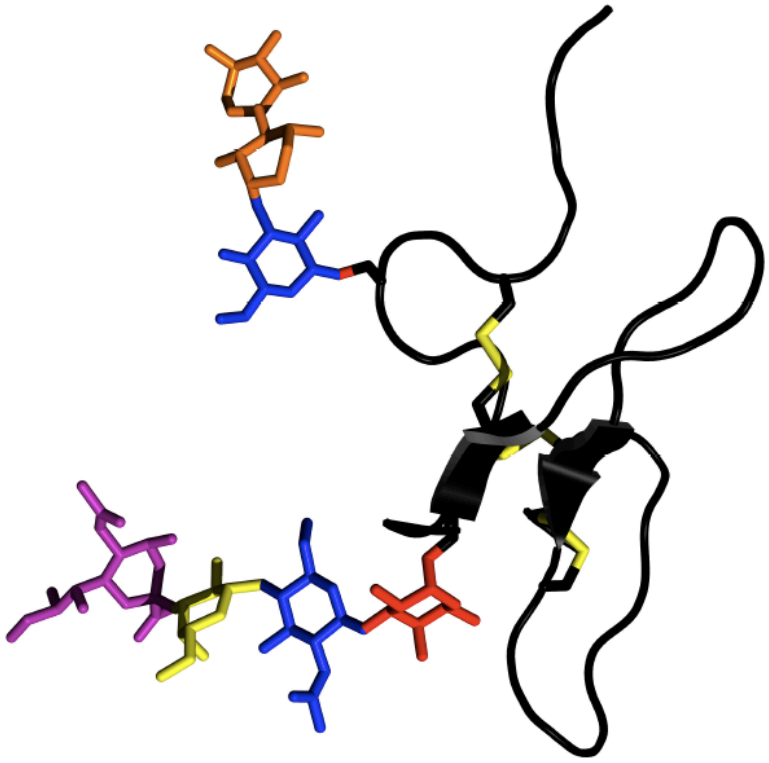


Figure 12. Summary of *O*-fucose and *O*-glucose glycans on mouse Notch1 (A), and model of an EGF repeat with both *O*-fucose tetrasaccharide and *O*-glucose trisaccharide (B). A. Domain map of the 36 EGF repeats in mouse Notch1. EGF repeats containing the revised *O*-glucose consensus sequence, $C^1xSx(P/A)C^2$, are shaded blue. All predicted *O*-glucose sites are modified with *O*-glucose trisaccharide, although some monosaccharide form also exists at lower levels. Calcium binding motifs between EGF repeats, which reduce flexibility between the two EGF repeats (73), are indicated by small black circles. L indicates EGF repeats implicated in ligand binding (101), and A indicates EGF repeats involved in *Abruptex* mutations (98). Predicted *O*-fucose sites are shaded red. The presence of *O*-fucose glycans at specific sites is based on data from Table 1 and other sources (19,102). Glycans at sites that have not been definitively shown to be modified are shaded. Elongation of *O*-fucose past the monosaccharide is based on work from (19,82). EGF repeats predicted to be modified by *O*-glucose and *O*-fucose are filled with blue and red hatching. Symbols: glucose, blue circle; xylose, orange star; fucose, red triangle; GlcNAc, blue square; Galactose, yellow circle; Sialic Acid, purple diamond. Faded red triangle represents EGF repeats *O*-fucose is predicted to modify but remain to be confirmed. **B.** Structure of the EGF 12 from mouse Notch1 with *O*-fucose disaccharide as recently determined by (103) (http://www.rcsb.org ID:2RQZ). The galactose and sialic acid, and the *O*-glucose trisaccharide structures, were modeled on using the Sweet II program (http://www.dkfz-heidelberg.de/spec/sweet2/doc/index.php#), and the picture was rendered with PyMOL. The side chains of the six cysteines are shown in stick form and all bonds coloring. For the *O*-fucose tetrasaccharide, the Thr-*O*-fucose, and the tetrasaccharide are shown in stick form with fucose in red, N-acetylglucosamine in blue, galactose in yellow, sialic acid in purple. Similarly, for the *O*-glucose trisaccharide, the Ser-*O*-glucose and the xyloses are shown in stick form, with the glucose in blue and the xyloses in orange. Adapted from (104).

^{208}ATHTGPHCELPYVP**CSPSPC**QN^{228}

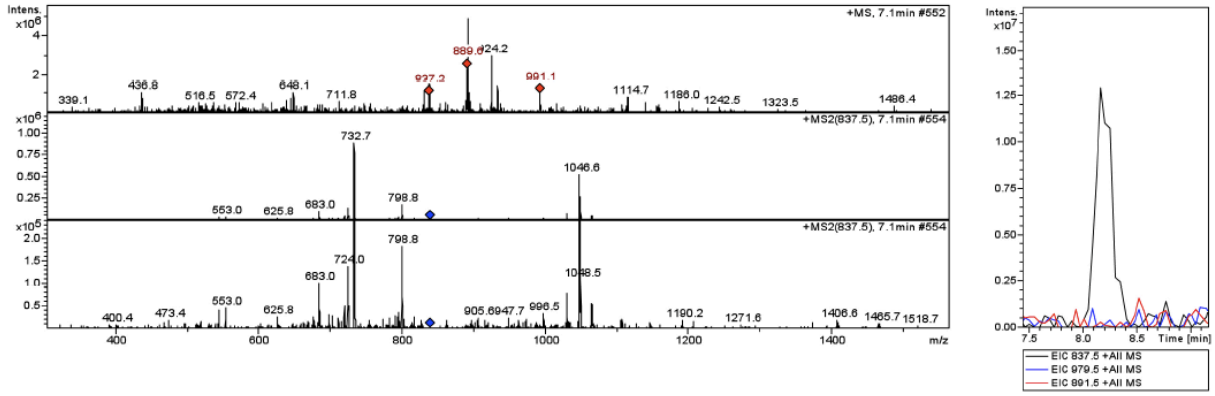


Figure 13. Mouse Notch1 EGF 6 is not *O*-glucosylated.
Mouse Notch1 EGF 1-18 was reduced, alkylated, and subjected to in-solution Asp-N digest. The resulting peptides were analyzed by nano-LC-MS/MS as described in Experimental Procedures. **A.** Top panel: MS scan showing parent ion 837.5 (red diamond) being selected for fragmentation. Bottom panel: MS/MS of the parent ion, where the blue diamond shows the position of the parent ion prior to fragmentation. Peptide fragment ions (b- and y-ions) in the MS/MS spectrum confirm peptide is from EGF 6. **B.** Extracted Ion Chromatogram traces of ions corresponding to the trisaccharide (blue), monosaccharide (red), and naked peptide (black) show the only species present is unmodified naked peptide.

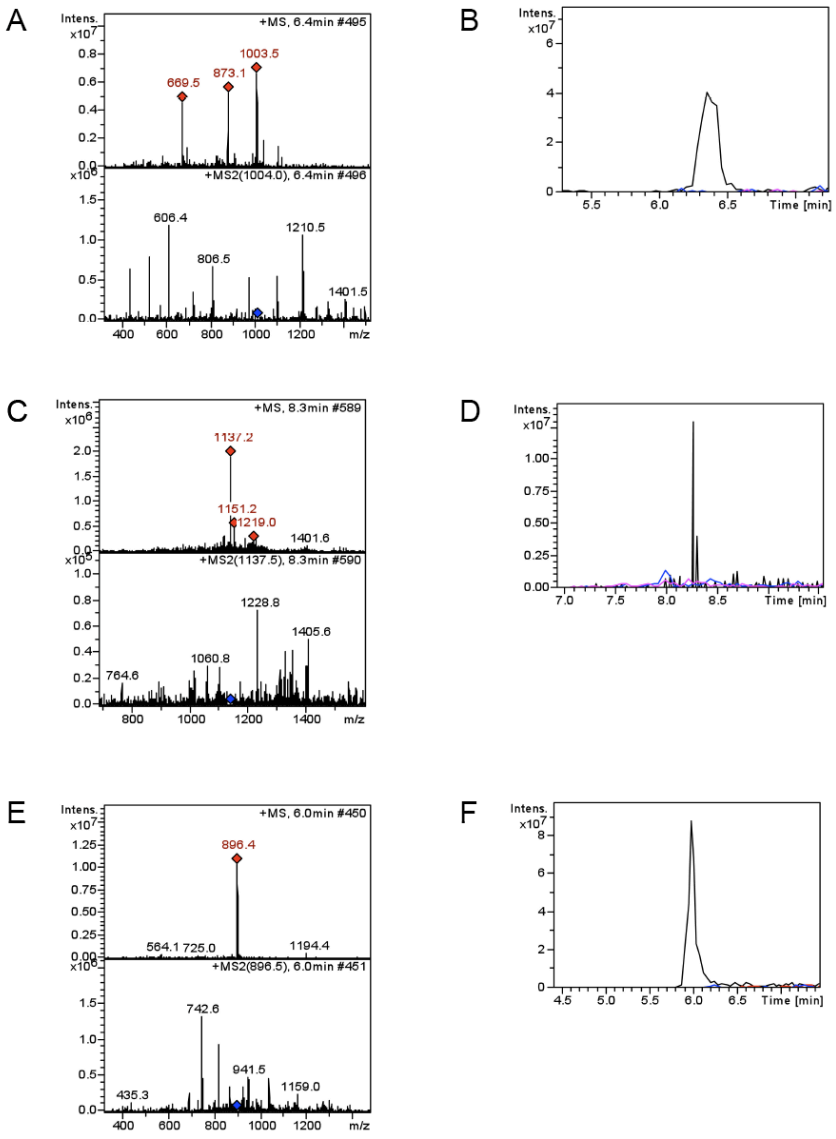


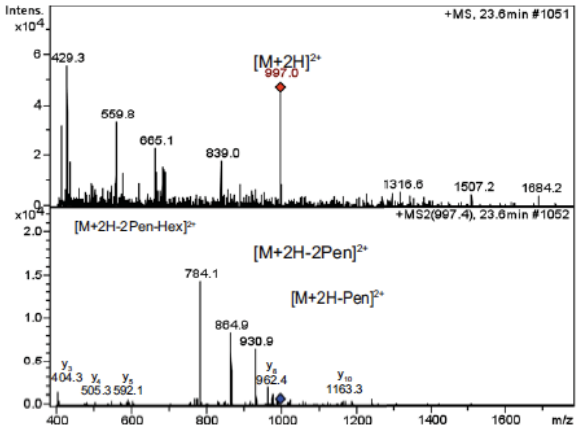
Figure 14. Non-traditionally positioned hydroxy amino acids are not modified with *O*-glucose or *O*-fucose on mouse Notch1.
Purified EGF repeats 24-28 and 29-36 of mouse Notch1 ECD were reduced, alkylated, and subjected to in-solution tryptic and chymotryptic digests. The resulting peptides were analyzed by nano-LC-MS/MS as described in Experimental Procedures. **A, C, E.** Top panels: MS scans showing parent ions 1004.0, 1137.5, and 896.5 (red diamonds) being selected for fragmentation. Bottom panel: MS/MS of the parent ion, where the blue diamond shows the position of the parent ion prior to fragmentation. Peptide fragment ions (b- and y-ions) in the MS/MS spectrum confirm peptides are from EGF repeats 23, 36, and 29 respectively. **B, D, F.** Extracted Ion Chromatogram traces of ions corresponding to the trisaccharide (blue), monosaccharide (red), and naked peptide (black) for EGF repeat 23, 36, and 29 respectively, show the only species present are unmodified naked peptide.

Figure 15. Identification of *O*-glucosylated peptides from EGF repeats 2, 10, 12, 13, 14, 16, 17, 19, 20, 21, 25, 27, 28, 31, and 33, and *O*-fucosylated peptides from EGF repeats 2, 3, 5, and 35 of mouse Notch1 by LC-MS/MS. *O*-Glycosylated peptides were identified using the methods described in Figure 5.For each peptide, an MS spectrum showing the selection of the parent ion for fragmentation (top) and an MS/MS spectrum showing the resulting CID fragmentation (bottom) is shown. M is the mass of the peptide with modifications. Predicted masses of singly charged ions are shown in Table 1. Ions in the MS/MS spectrum showing losses of the modifications are indicated. The sequence of the peptide and the modifications is provided at the top of each spectrum. **A.** Peptide from EGF 2 was identified in digests of EGF 1-5 and EGF 1-18. **B.** Peptide from EGF 10 was identified in digests of EGF6-10 and EGF 1-18. **C.** Peptide from EGF 12 was identified in digest of EGF 11-15, and is modified with *O*-glucose trisaccharide and *O*-fucose. **D.** Peptide from EGF 13 was identified in digest of 11-15. **E.** Peptide from EGF14 was identified in digest of EGF 1-18. **F.** Peptide from EGF 16 was identified in digest of EGF 1-18, and is modified with *O*-glucose trisaccharide and *O*-GlcNAc. Ions matching loss of *O*-glucose trisaccharide (bold text) followed by *O*-GlcNAc, as well as *O*-GlcNAc followed by *O*-glucose trisaccharide (normal text) can be seen. **G.** Peptide from EGF 17 was identified in digest of EGF1-18. **H.** Peptide from EGF 19 was identified in digest of EGF19-23. **I.** Peptide from EGF 20 was identified in digest of EGF 19-23. Because this peptide is modified with *O*-fucose as well as *O*-glucose, ions matching loss of the fucose followed by the *O*-glucose trisaccharide (plain text) or the *O*-glucose trisaccharide followed by *O*-fucose (bold text) can be seen. **J.** Peptide from EGF 21 was identified in digest of EGF 19-23, also modified with *O*-fucose. As with the peptide from EGF 20 shown in F, ions matching loss of the fucose followed by the *O*-glucose trisaccharide (plain text) or the *O*-glucose trisaccharide followed by *O*-fucose (bold text) can be seen. **K.** Peptide from EGF 25 was identified in digest of EGF 24-28. Here we see loss of Xylose+water, followed by sequential loss of remaining xylose and *O*-glucose. **L.** Peptide from EGF 27 was identified in digest of EGF 24-28. Ions matching loss of *O*-glucose trisaccharide and then *O*-fucose can be seen (bold text), as well as loss of *O*-fucose first followed by *O*-glucose trisaccharide (plain text).**M.** Peptide from EGF 28 was identified in digest of EGF 24-28. **N.** Peptide from EGF 31 was identified in digest of 29-36. **O.** Peptide from EGF33 was identified in digest of EGF29-36. **P-R.** Peptides from EGF 2, 3, and 5 were identified in a digest of EGF 1-5. **S.** Peptide from EGF 35 was identified in digest of EGF 29-36. B, E, G, H, J, M, and P-S were derived from experiments performed by Dr. Aleksandra Nita-Lazar.

A

EGF 2

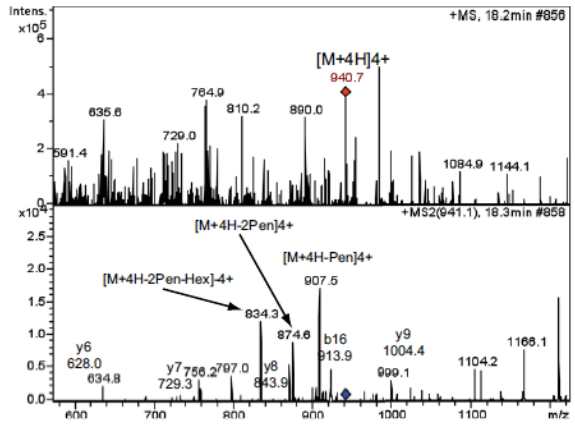
CQDSNPCLSTPCK



B

EGF 10

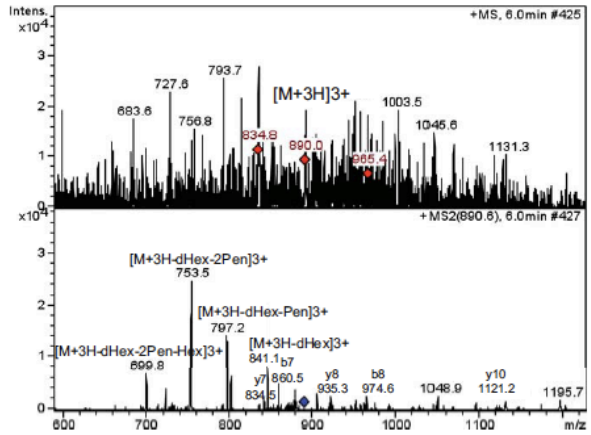
TGLLCHLNDACISNPCNEGSNCDTNPVNGK



C

EGF 12

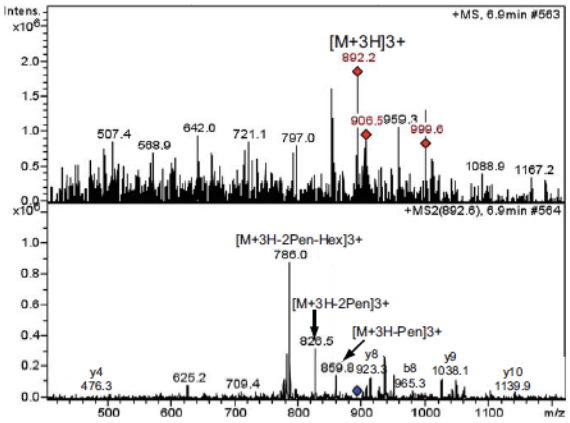
CISNPCQNDATCLDQIGE



D

EGF 13

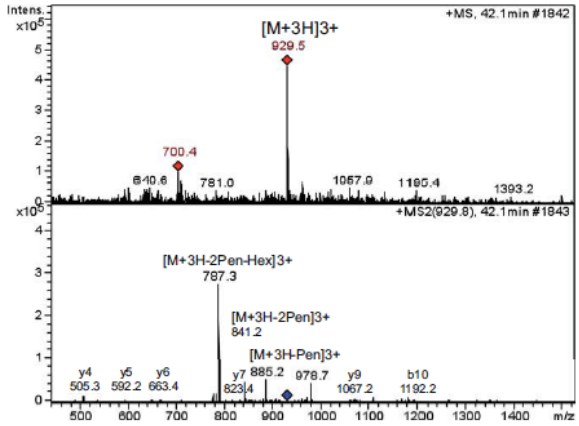
EGVYCEINTDE**CA**S**S**P**C**L



E

EGF 14

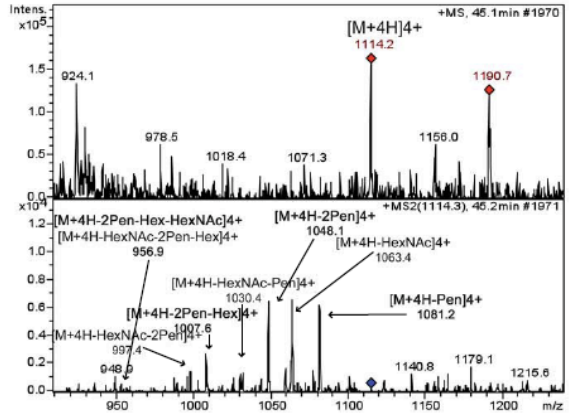
GFNGHLCQYDVDE**CA**S**T**P**C**K



F

EGF 16

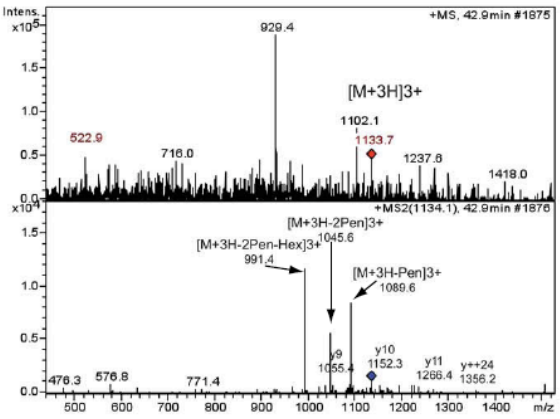
DGVATFTCL**CQ**P**G**Y**T**G**HHC**ETNINE**CH**S**QPC**R



G

EGF 17

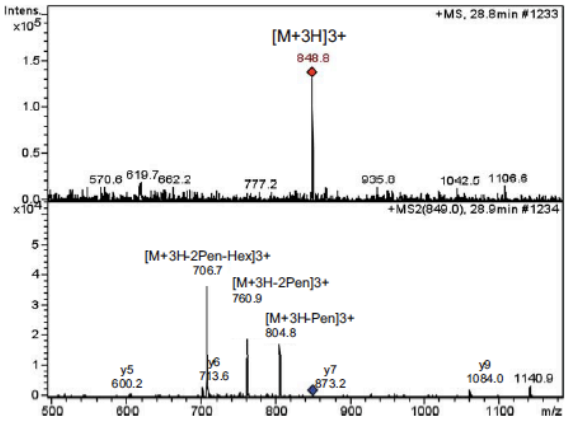
GTTGPNCEINLDD**CA**S**NPC**DSGTCLDK



H

EGF 19

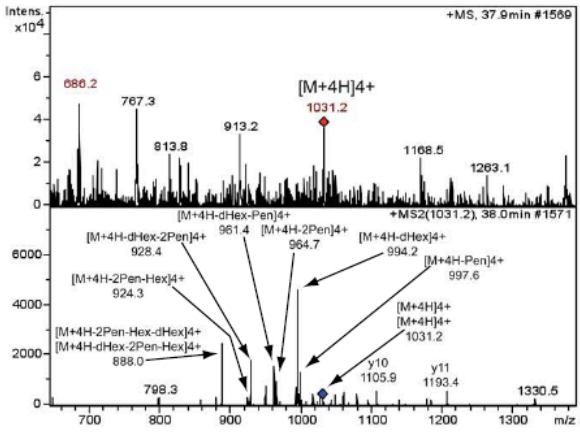
LSEVNE**CN**S**NPC**IHGACR



I

EGF 20

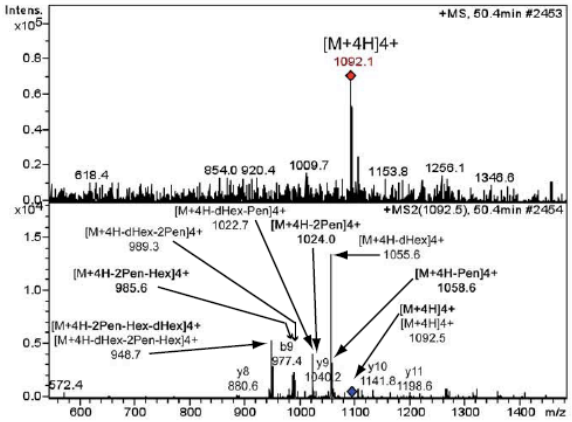
CDCAPGWSGTNCDINNNE**CE**S**NPC**VNGG**T**CK



J

EGF 21

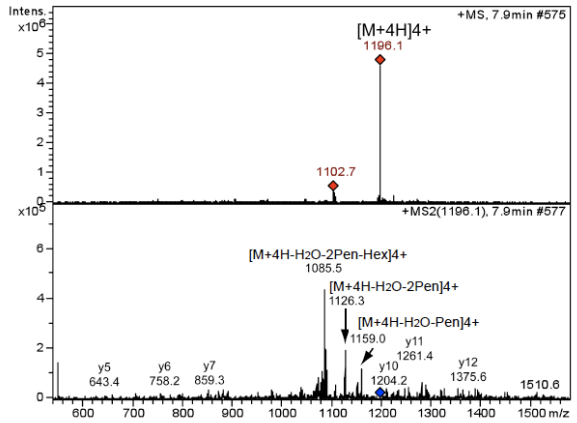
EGFSGPNCQTNINE**CA**S**NPCLNQGTC**IDDVAGYK



K

EGF 25

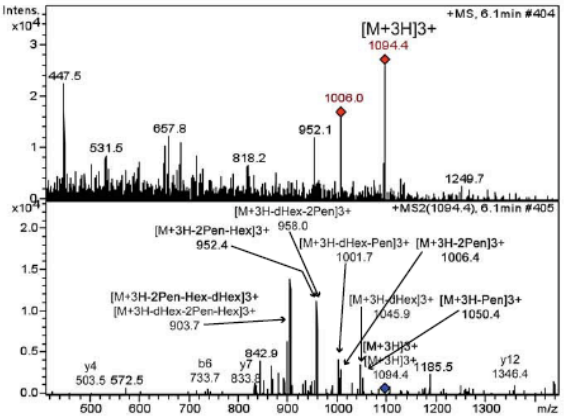
CDCLPGFQGAFCEEDINE**CA**S**NPC**QNGANCTDCVDSY



L

EGF 27

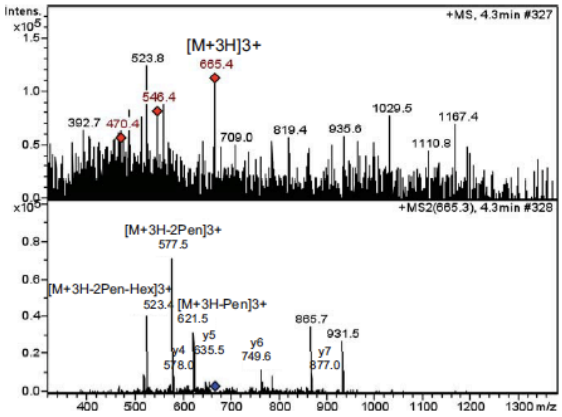
DVNE**CD**S**RPCLHGGTC**QDSYGTY



M

EGF 28

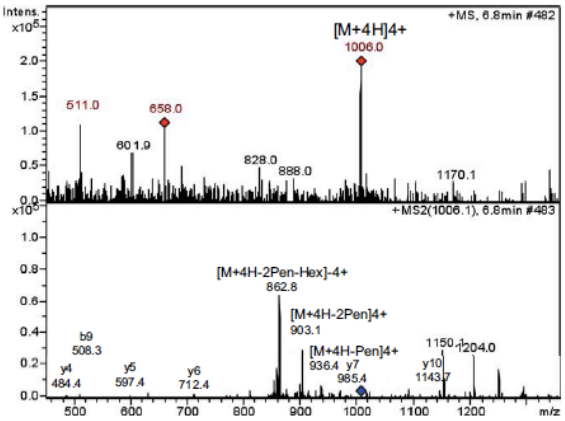
CDSAPCKNGGRCW



N

EGF 31

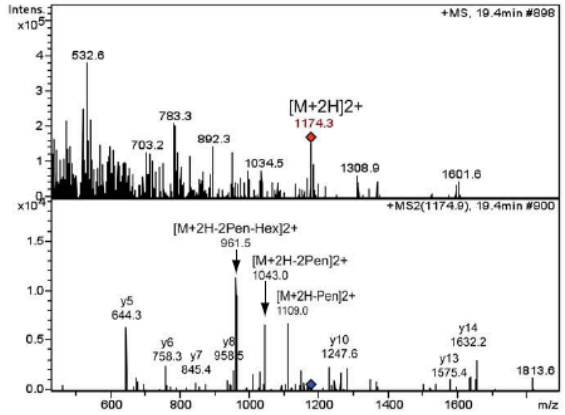
HGSNCSEEINE**CLSQPCQNGGTC**IDLTNSY



O

EGF 33

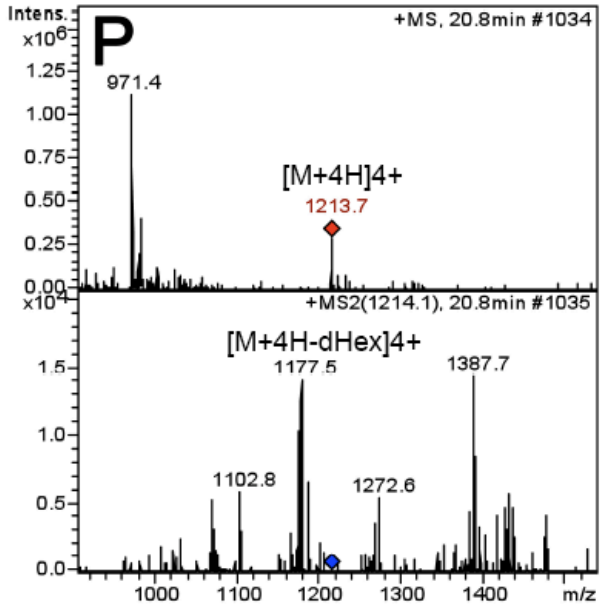
CEGDVNE**CLSNPC**DPR



EGF 2

M= ^{70}NAGTCHVVDHGGTVDYACSCPLGFS-GPLCLTPLDNACLANPCR^{112} +dHex

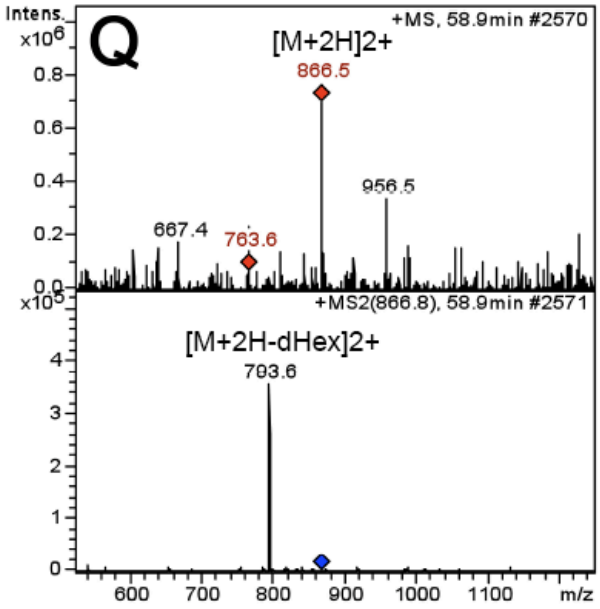
[M+4H]4+ = 1214.1



EGF 3

M=^{113}NGGTCDLLTLTEYK^{126} +dHex

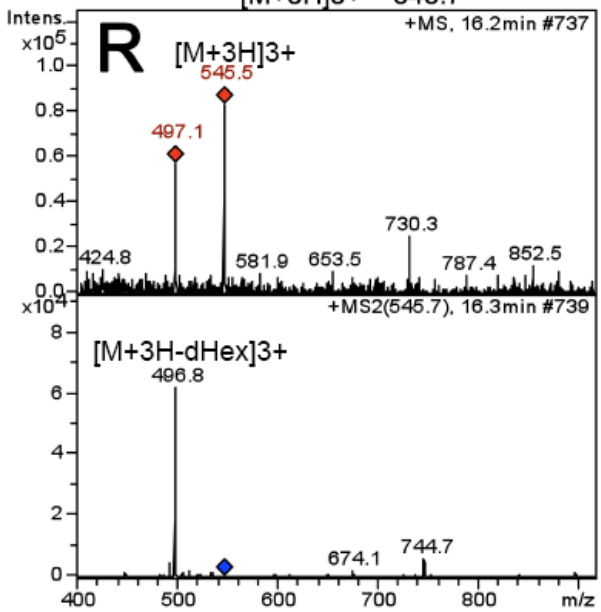
[M+2H]2+ = 866.8



EGF 5

M= ^{191} HGGTCHNEIGSYR^{203} +dHex

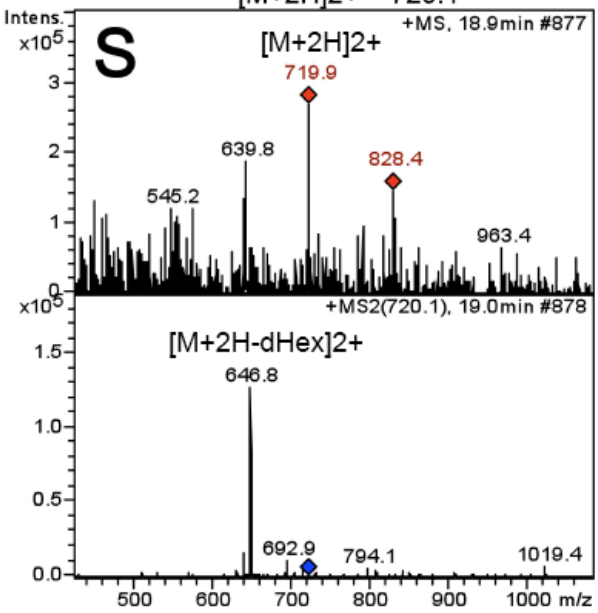
[M+3H]3+ = 545.7



EGF 35

M= ^{1357}CLNGGTCISGPR^{1368} +dHex

[M+2H]2+ = 720.1



Chapter III: On the "Fringe" of Notch: Mapping *O*-Fucose and *O*-Glucose Glycans at Specific Sites on the Extracellular Domains of *Drosophila* Notch Using Nano-LC-ESI-MS/MS

Introduction

Although *Drosophila* Notch has been shown to be modified with *O*-fucose and *O*-glucose glycans, little is know about which sites are occupied, or the extent to which Fringe or glucoside xylosyltransferase (GXYLT) elongate the monosaccharides at any or all of these sites. Here we have conducted a comprehensive mapping of the extent of *O*-fucosylation and *O*-glucosylation on the EGF repeats of *Drosophila* Notch, and the elongation of these glycans to disaccharide structures. *O*-fucosylation occurs at high stoichiometry at predicted sites across the extracellular domain (ECD), while Fringe elongation to the GlcNAc-β1,3-Fucose disaccharide occurs at sub-stoichiometric levels in a site-specific manner. This indicates Fringe shows a preference for elongation depending on context, and that there are in fact hotspots for modification with GlcNAc along the Notch receptor. Like *O*-fucose modifications, *O*-glucosylation also occurs stoichiometrically, but elongation with xylose modifications appear to cluster to the middle region of the ECD. GXYLT elongation to the Xyl-α1,3-Glc disaccharide is also observed to be substoichiometric. To examine this further, we implemented a semi-quantitative mass spectral method, Multiple Reaction Monitoring, to confirm if there is a hierarchy among fucosylated sites to serve as substrates for Fringe elongation, and likewise for glucosylated sites as substrates for GXYLT elongation. This has allowed us to determine that some sites are in fact preferred for elongation by Fringe over others.

The extracellular domain (ECD) of *Drosophila* Notch consists largely of 36 tandem epidermal growth factor-like (EGF) repeats, each of which is approximately 40 amino acids in length, motifs which are found in a wide variety of both cell-surface and secreted proteins, including all Notch homologs, and the Delta, Serrate, and Lag2 (DSL) family of Notch ligands. These EGF repeats are characterized by six conserved Cysteines that form three disulfide bonds, giving rise to a distinct three-dimensional folding pattern. EGF repeats can also bear a variety of post- and co-translational modifications, including three forms of *O*-linked glycosylation: *O*-fucose, *O*-glucose, and *O*-GlcNAc (13,16,17). Putative consensus sequences were established for these modifications: *O*-fucosylation occurs at a site located between the second and third conserved Cysteine residues on Serines or Threonines $C^2xxxx(S/T)C^3$, (19,105), while the newly revised *O*-glucose consensus sequence can be found between the first and second conserved Cysteine residues ($C^1xSx(A/P)C^2$) (13) (See Chapter 2). The *O*-GlcNAc modification on EGF repeats has recently been observed to occur between Cysteines five and six (17), but a consensus sequence has yet to be defined. Over 100 proteins are predicted to be modified with these forms of *O*-glycosylation, with the Notch proteins containing the highest number of predicted sites of *O*-fucosylation and *O*-glucosylation of any protein by database searching using the consensus sequences defined (2), (Chapter II, Table 3).

Protein *O*-fucosyltransferase1 (Pofut1, OFUT1 in flies) and Protein *O*-glucosyltransferase (Poglut, Rumi in flies) are the glycosyltransferases that add *O*-fucose and *O*-glucose monosaccharides to Notch, respectively. Loss of *Pofut1/Ofut1* or *Poglut/rumi* in mice (9,24) or flies (8,41,42) phenocopies Notch-null phenotypes, suggesting both forms of *O*-glycosylation are essential for Notch function. The β1,3-GlcNAc transferase Fringe (Fng) can extend *O*-fucose to a disaccharide in flies and mammals. In flies, the Fng mutant phenotype is

characterized primarily by loss of wing development, due to a lack of differential Notch activation at the dorsal-ventral boundary during development of the wing imaginal disc. In addition, it has also been shown that loss of Fringe affects development of other fly tissues, including eyes, legs, and ovaries, and is therefore essential in the regulation of the development of multiple tissues in the *Drosophila* system (106-111). These studies convey that Fng is an important modulator of Notch function. In mammals, the GlcNAc-β1,3-Fucose disaccharide can be elongated by a β1,4-Galactosyltransferase, and subsequently by a sialyltransferase, to the tetrasaccharide Sia-α2,6-Gal-β1,4-GlcNAc-β1,3-Fuc-*O*-Ser/Thr. In flies, *O*-fucose has not been observed on Notch in a more elongated form than the GlcNAc-β1,3-Fucose-*O*-Ser/Thr disaccharide. Though there are two members of the β4GalT family in flies, they actually encode GalNAc transferases, and double mutant animals exhibit no *fng*-like phenotype (112).

Similar to Fringe elongation of *O*-fucose, parallel extension of *O*-glucose occurs by two recently characterized glucoside α1,3-xylosyltransferases, GXYLT1 and GXYLT2, which extend the monosaccharide structure to Xyl-α1,3-Glc-*O*-Ser in mammals (26). Although the function of GXYLT1,2 are not yet known, by analogy we suspect they will also affect Notch activity. A second, newly characterized mammalian xyloside α1,3-xylosyltransferase (Sethi and Bakker, Personal Communication) can further elongate this structure to a trisaccharide by addition of a terminal xylose, generating a Xyl-α1,3-Xyl-α1,3-Glc-*O*-Ser trisaccharide in mammals, which is the predominant *O*-glucose glycoform observed in mammalian systems/tissues (see Chapter II). In an *O*-glycomics analysis of total protein extracts from fly embryos, Aoki *et al.* (20) showed *O*-glucose mono- and disaccharide structures exist within the *Drosophila* glycome, though no trisaccharide was detected, rendering whether elongation of *O*-glucose occurs on *Drosophila* Notch a major unanswered question. In these same studies, Aoki

et al. did not detect any *O*-fucose mono- or disaccharide, but instead report of a novel *O*-fucose-GlcNAc-GlcA trisaccharide. Though this trisaccharide has never been reported on Notch, it is of note that the Glucuronic Acid occurs on what appears to be a Fringe-elongated glycan structure, and is observed in Fringe-rich tissues during fly development (20). Whether this novel trisaccharide occurs on Notch remains to be resolved.

To gain insights into how these essential modifications impart functionality to the receptor, it is vital to identify sites of *O*-glycosylation, and define structures of the glycans at each of these sites. We have previously shown Notch to be *O*-fucosylated (96) and *O*-glucosylated (8) at multiple sites (Figure 12). In continuation of this work, here we address the question of site-occupancy at predicted sites of *O*-fucosylation and *O*-glucosylation by extending our site-mapping efforts to include 20 predicted *O*-fucose and 16 predicted *O*-glucose sites using mass spectrometry. We show that Fringe-mediated elongation of *O*-fucose is substoichiometric and site specific. We also show that O-glucose is elongated with a single xylose at a subset of sites, but we find no evidence for elongation to the trisaccharide observed in mammals. We further explored relative levels of site-specific elongation to disaccharide for both *O*-glycan modification using a semi-quantitative mass spectral method known as Multiple Reaction Monitoring.

Experimental Procedures:

Expression Constructs:

pMTHy plasmids encoding all 36 EGF repeats in the *Drosophila* Notch ECD with a C-terminal triple FLAG tag (N-EGF:FLAG) and *Drosophila* Fringe with a C-terminal His_6 tag (Fng) were generous gifts from Dr. Kenneth Irvine (Rutgers University). Construction of these plasmids, transfection of *Drosophila* S2 cells, and purification of the resulting proteins from medium has been described (25,61,96). pMTHy empty vector, used as a control for transfections, was a generous gift from Dr. J. Peter Gergen (Stony Brook University).

Cell culture and protein purification:

Stably transfected S2 cells were generated and cultured as described previously (96). Briefly, S2 cells were stably co-transfected with either 4.0 µg each of N-EGF:FLAG and pMTHy empty vector or 4.0ug each of N-EGF:FLAG and Fng, using Cellfectin II (Invitrogen) to generate unmodified (N-EGF:FLAG) or *in vivo* Fng-elongated N-EGF:FLAG (N-EGF:FLAG+Fng). Cells were cultured in Schneider's *Drosophila* media (Invitrogen) supplemented with 10% FBS, 1% penicillin-streptomycin.

For purification of N-EGF:FLAG (-/+Fng), stably transfected S2 cells were cultured to 250ml in Schneider's *Drosophila* media, and expression was then induced by addition of 0.7 mM $CuSO_4$ for 3 days. Media was cleared by centrifugation at 4000 rpm for 20 min at 4ºC, and conditioned media was incubated with 250 µl of anti-FLAG beads (Sigma) overnight at 4ºC with rotation. Purified protein was eluted from beads using 3XFLAG peptide (Sigma) as per manufacturer's instructions. To remove excess 3XFLAG peptide, samples were dialyzed overnight against TBS supplemented with 2 mM CaCl2 and 10% glycerol. 10 µl of eluate was

run on a 12% SDS-PAGE with BSA standards of varying concentrations and stained with Coomassie blue. Protein yields were estimated by densitometry using ImageJ software (NIH).

Analysis of Fringe elongation using Galactosyltransferase labeling:

Galactosyltransferase assays were used to measure the amount of GlcNAc being added to N-EGF:FLAG by Fringe. *In vitro* radiolabeling of GlcNAc with galactose was carried out using saturating levels of bovine β4GalT-1 (Sigma), and UDP-[^{3}H]galactose (60 Ci/mmol; American Radiolabeled Chemicals, Inc.). To quantify and chemically analyze the *O*-fucose disaccharide on Notch (generated *in vivo*) we used mammalian ß4GalT-1. FLAG-tagged Notch (N-EGF:FLAG) was expressed and purified from WT (N-EGF:FLAG) or Fringe-expressing S2 cells (N-EGF:FLAG+Fng) and then glycosylated *in vitro* using saturating levels of ß4GalT-1 and UDP-[^{3}H]-galactose, in a standard galactosyltransferase assay. Briefly, each reaction mixture (final volume=500ul) contained 1.0ug of N-EGF:FLAG unmodified or elongated *in vivo* by Fringe, 50mM HEPES, pH 7.4, 5mM $MnCl_2$, 10mM Galactose, 2% NP-40, and 2uCi UDP-[^{3}H]-Galactose (American Radiolabeled Chemicals, Inc.). 400 milliunits of bovine ß1,4-galactosyltransferase was added to each tube, and allowed to incubate at 37^0C for 1h. Reactions were terminated by addition of 50ul stop solution (0.1M EDTA, 10% SDS). Samples were then separated on a Sephadex G-50 column in 1.0ml fractions, and aliquots analyzed for radioactivity by scintillation counting. After peptide *N*-glycosidase F (PNGaseF) was used to remove any contaminating *N*-glycans, *O*-glycans were released from N-EGF:FLAG using alkali-induced β-elimination, and were subsequently analyzed by gel filtration chromatography on a Superdex peptide column as described (96).

Analysis of glycopeptides by nano-LC-ESI-MS:

Nano-LC-ESI/MS and MS/MS, Constant Neutral Loss (CNL) manual searches, and Extracted Ion Chromatogram (EIC) manual searches were all performed as described (Chapter 2, Experimental Procedures) (8,22,96). Briefly, 1.0 µg N-EGF:FLAG -/+ Fng was acetone precipitated, subjected to reduction/alkylation and in-gel or in-solution protease digest with trypsin, chymotrypsin, Asp-N, or V8. In-gel digests were concentrated using ZipTip C18 cartridges (Millipore), and in solution digests were cleaned up by spin-filtering with a 0.22um filter (Agilent).

Constant Neutral Loss (CNL) and Extracted Ion Chromatogram (EIC) Searches:

The *O*-glucose modified peptides were identified by searching the MS/MS data for constant neutral losses of masses equivalent to the *O*-glucose mono-, di-, and trisaccharides (162, 294, and 426 Da, respectively) relative to parent ions using the Data Analysis Tool (Agilent) as described (92). *O*-Fucose mono- and disaccharide modified peptides were identified by constant neutral loss searches of 146 and 349Da, respectively. *O*-GlcNAc modified peptides identified by constant neutral loss searches of 203 Da. Subsequent searches of the MS/MS data were performed to identify peptides modified with mono- or disaccharide forms of *O*-glucose by generating extracted ion chromatograms using the masses of the unglycosylated peptide for the search.

For Multiple Reaction Monitoring (MRM) (113), 1.0 µg quantities of N-EGF:FLAG -/+ Fng were prepared by in-solution digest in parallel, as described above. Based on MS and MS/MS data, glycopeptide ions were followed in pairs with a transition ion to diminish background of non-specific ions matching the m/z value of our glycopeptide of interest. For example, a parent ion bearing *O*-Fucose-GlcNAc disaccharide was monitored with the

unmodified naked peptide as the transition ion. As a nonspecific ion sharing the same m/z value as our ion of interest is unlikely to also generate the same naked peptide m/z as a daughter ion, this in effect reduced background signal more significantly than an EIC search is capable of. Constant neutral loss intermediates were also followed, again with naked peptide as the transition ion, and finally naked peptide was monitored with a prominent b- or y-ion as a transition ion. This allowed us to monitor the amount of disaccharide, monosaccharide, and unmodified naked peptide in a minus or plus Fng sample in a semi-quantitative manner. In addition to quantifying glycopeptides, non-glycosylated peptides were also quantified in minus and plus Fng samples as an injection control. Since a given peptide lacking glycosylation sites should ionize to an equivalent efficiency independent of glycosylation state of the sample (ie. minus or plus Fng), using an internal non-glycosylated peptide as our control allowed us to normalize for run-to-run variations and also served a dual function as an injection control.

Results

Notch is modified extensively with O-fucose and O-glucose in S2 cells

Previously, we have confirmed modification of 7 predicted sites of *O*-fucosylation and 5 predicted sites of *O*-glucosylation on the ECD of Notch expressed in S2 cells (Figure 28) (8,96). Here we sought to determine whether the other predicted *O*-glycosylation sites are modified and whether the structures vary from one site to another. To determine how extensively the receptor is modified with both forms of *O*-glycosylation, we used the previously refined consensus sequence for *O*-fucosylation ($C^2xxxx(\underline{S/T})C^3$), and the newly revised consensus sequence for *O*-glucosylation ($C^1x\underline{S}x(P/A)C^2$), to predict additional potential sites of modification. *Drosophila* Notch has a total of 23 predicted sites of *O*-fucosylation and 18 predicted sites of *O*-glucosylation (Figure 16A).

To confirm whether only a subset or all of the predicted sites of *O*-glycosylation were actually modified, and define the *O*-glycan structures at modified sites, we implemented a mass spectral approach that we have used recently to map all predicted, plus one novel site of *O*-glucosylation, on mouse Notch1 (Chapter 2). The comprehensive nature of this approach enabled us to revise the consensus sequence for *O*-glucosylation, and we have used similar approaches to map the twelve previously reported *O*-glycosylation sites on Notch (8,96). To map novel sites on *Drosophila* Notch, N-EGF:FLAG was co-expressed with Fng in *Drosophila* S2 cells. Fng-modified N-EGF:FLAG was purified using anti-FLAG beads. N-EGF:FLAG+Fng was reduced, alkylated, and subjected to in-gel or in-solution digest with a protease to generate glycopeptides of appropriate size and charge states amenable to analysis by mass spectrometry. Trypsin, chymotrypsin, Asp-N, or V8 were used to generate peptides from every region of the ECD based on *in silico* digests, allowing greater than 70% sequence coverage of the EGF repeats

(Figure 16B). Glycopeptides were identified by constant neutral loss (CNL) searches for the *O*-glucose mono- and disaccharides and or *O*-fucose mono- and disaccharides resultant from CID-induced low energy fragmentation. A summary of the site mapping data can be found in Table 5.

Drosophila Notch is modified with O-glucose mono- and di-saccharides in S2 cells

To confirm the location of *O*-glucose modifications on Notch, and to determine whether any or all of these sites are elongated by the addition of any xyloses, we searched MS/MS data for neutral losses of *O*-glucose mono- and disaccharides. For detection of *O*-glucose monosaccharide, CNL searches included m/z losses of 81, 54, and 40.5 (loss of *O*-glucose, 162Da from a doubly, triply, or quadruply charged ion), and for Xyl-Glc disaccharide, searches were inclusive of m/z losses of 147, 98, 73.5 (loss of Xyl-Glc, 294Da from a doubly, triply, or quadruply charged ion). An example of these CID analyses for a chymotryptic digest can be seen in Figure 17A. The top panel in 17A shows the MS spectrum with an ion corresponding to a triply charged chymotryptic peptide containing an *O*-glucosylation site from EGF repeat 16, modified with *O*-glucose disaccharide. The mass of this ion corresponds to the triply charged form of a predicted chymotryptic peptide from EGF 16 of Notch that contains an *O*-glucose consensus sequence (in bold): ^{635}ICQKQINE**CE<u>S</u>NPC**QF^{650} (see Table 5 for predicted masses). The bottom panel shows CID fragmentation (MS/MS) of the parent ion, m/z 784.3, resulting in major product ions that correspond to the sequential losses of a pentose (xylose) (m/z= 740.3) and hex (glucose) (m/z= 686.1) from the glycopeptide that was found during CNL searches for disaccharide. This data is indicative of *O*-glucose disaccharide on this peptide, as b- and y-ions confirm that this is Notch EGF 16, modified with *O*-glucose-xylose. Our analysis confirms that

the fragmentation pattern seen here (major product ions resulting from the sequential losses of a pentose followed by a hexose) is diagnostic for identifying *O*-glucose disaccharide-modified peptides. To examine whether this same peptide exists with other modifications, we searched the MS/MS data for the unglycosylated peptide (m/z= 686.1), monosaccharide (m/z= 740.3), disaccharide (m/z= 784.3), and trisaccharide (m/z= 828.3) forms (black, blue, pink, and green traces respectively, Figure 17C). A pair of co-eluting peaks was detected (labeled blue, pink in Figure 17C). The larger blue peak is the same species as the parent ion shown in Figure 17A (peptide modified with *O*-glucose disaccharide). The minor pink peak corresponds to the peptide modified with *O*-glucose monosaccharide (corresponding MS and MS/MS spectra shown in Figure 17B). Importantly, the peptide modified with monosaccharide elutes slightly later (6.25 min) than the disaccharide form (6.35 min), consistent with the absence of the xylose, which contributes to the hydrophilicity of the peptide and therefore earlier elution time. This is consistent with minor elution time differences observed for *O*-glucose mono- and trisaccharides modifying mouse Notch1 (22)(Chapter 2, Figure 3C). No significant amount of ions corresponding to naked peptide nor peptide with *O*-glucose trisaccharide were observed (Figure 17C). Similar EIC searches comparing relative levels of monosaccharide in comparison to naked peptide, as carried out in Figure 17C, reveal *O*-glucose modifies the receptor at all 18 observed sites, indicating Rumi modifies Notch at high stoichiometries (Table 5, Figure 23, Figure 28). These EIC traces also confirm that *O*-glucose is not elongated beyond the Xyl-Glc disaccharide on Notch expressed in S2 cells, consistent with the glycomic analysis performed by Aoki and coworkers (20).

Drosophila Notch is modified with O-fucose mono- and di-saccharides in S2 cells

It has been established that Fng addition of GlcNAc is sufficient to modulate Notch-ligand binding in S2 cells, and that *in vitro* elongation using mammalian β4GalT-1 does not significantly alter Notch-ligand binding (96). To determine saccharide sizes of *O*-glycans present on *in vivo* Fng-modified Notch, and evaluate the extent of Fringe elongation relative to unmodified Notch, N-EGF:FLAG produced -/+Fng in S2 cells were glycosylated *in vitro* using saturating levels of β4GalT-1 and UDP [^{3}H]Gal. In the absence of Fng, approximately 3.6 mol of Galactose were added to each mol of N-EGF:FLAG. In the presence of Fng, approximately 15.9 mol of Galactose were added to each mol of protein (Figure 18). This means Fng is responsible for a 12.3 mol/mol increase in Gal incorporation, or roughly 12 GlcNAc residues were transferred to each molecule of Notch. Product characterization was carried out on these samples by first treating with peptide *N*-glycosidase F (PNGaseF) to remove contaminating *N*-linked glycans, followed by alkali-induced β-elimination to release *O*-glycans (Figure 19). β-eliminated *O*-glycans in their sugar-alcohol forms were then subjected to size separation by Superdex column chromatography (Figure 20). Radiolabeled *O*-glycans from the Fng-modified samples co-migrated with Gal-β1,4-GlcNAc-β1,3-fucitol standard, the predicted product of Fng-modified disaccharide post-β4GalT-1 radiolabeling. As expected, this trisaccharide peak was not detected in –Fng samples, consistent with a lack of GlcNAc-β1,3-fucose disaccharide in these samples. A minor disaccharide peak was observed, attributable to incorporation of galactose into a monosaccharide species such as *O*-GlcNAc, known to modify Notch (17). This disaccharide structure remains to be verified. Of importance is the fact that no larger *O*-fucose saccharides, including the novel GlcA-GlcNAc-Fuc trisaccharide, were detected in these studies,

demonstrating that no further elongation of the GlcNAc-Fuc disaccharide occurs in Fng-expressing S2 cells.

To identify glycopeptides from Notch modified with *O*-fucose glycans, constant neutral loss searches were performed as described above. An example of these CNL analyses of tryptic digest can be seen in Figure 21. The top panel in Figure 21A shows an MS spectrum with an ion corresponding to a triply charged tryptic peptide bearing a GlcNAc-β1,3-Fucose disaccharide from EGF 21. The bottom panel shows CID fragmentation (MS/MS) of this ion (m/z= 1138.6) resulting in major product ions that correspond to the sequential losses of a HexNAc (GlcNAc) (m/z= 1070.9) and dHex (fucose) (m/z= 1021.9) from the glycopeptide that was found during CNL searches for disaccharide. The mass of this ion corresponds to the triply charged form of a predicted tryptic peptide from EGF 21 of Notch that contains an *O*-fucose consensus sequence (in bold): ^{826}CETNIDDCVTNP**CGNGGTC**IDK^{847} (see Table 5 for predicted masses). This data is indicative of the presence of *O*-fucose disaccharide on this peptide, and b- and y-ions confirm that this is Notch EGF 21. Consistent with previous work (96), our analysis confirms that the fragmentation pattern seen here (major product ions resulting from the sequential losses of a hexNAc followed by dhex) is diagnostic for identifying *O*-fucose disaccharide-modified peptides. To verify whether this same peptide exists with other modifications, we searched the MS data for the unglycosylated peptide (m/z= 1021.9), monosaccharide (m/z= 1070.9), disaccharide (m/z= 1138.6), and potential novel trisaccharide (m/z= 1192.6) forms (black, red, blue, and yellow traces respectively, Figure 21C). A pair of closely eluting adjacent peaks were detected (labeled blue, red in Figure 21C). The smaller blue peak is the same species shown in Figure 21A (peptide modified with *O*-fucose disaccharide). The larger red peak corresponds to the peptide modified with *O*-fucose monosaccharide. The black trace shows levels of naked

peptide are negligible. As expected, no signal for trisaccharide (yellow trace) is detected, consistent with the galacosyltransferase radiolabeling shown in Figure 20 that *O*-fucose is not elongated beyond the disaccharide on Notch expressed in S2 cells. The MS scan and CID fragmentation of the monosaccharide is shown in Figure 21B. The peptide modified with monosaccharide elutes slightly later (5.5 min) than the disaccharide form (5.3 min), consistent with the absence of the GlcNAc, which contributes to the hydrophilicity of the peptide and therefore earlier elution time. A closer look at the levels of *O*-fucosylation across the receptor using EIC searches to compare relative levels of *O*-fucose monosaccharide and naked peptide, as seen in Figure 21C, reveals *O*-fucose modifies the receptor at high stoichiometries. Of the 22 predicted sites of *O*-fucosylation, 20 sites have been identified to date and have been shown to bear the glycan modification, indicating OFUT1 modifies Notch at high stoichiometries (Table 5, Figure 23, Figure 28). Additionally, no novel trisaccharide was detected in these MS studies, consistent with independent biochemical data showing only *O*-fucose mono- and disaccharides modify N-EGF:FLAG+Fng in S2 cells, and highly suggestive of this modification not existing in this cell system.

Summary of O-fucose and O-glucose site mapping on Drosophila *Notch from S2 cells*

Constant neutral loss searches of the *O*-fucose disaccharides for the hallmark losses of GlcNAc and fucose, revealed ions corresponding to peptides containing the vast majority of the previously unreported predicted *O*-fucosylation sites: EGF repeats 1, 2, 4, 8, 9, 21, 24, 25, 26, 27, 28, 30, and 31 (Table 5, Figures 21, 23, 28). Initial identification of glycopeptides was carried out by manual searches for hallmark fragmentation of *O*-fucose disaccharide or monosaccharide (as in Figure 21). Masses of naked peptide were then matched to predicted

monoisotopic or average masses of *in silico* digested N-EGF:FLAG peptides that contained *O*-fucose consensus sites (Table 5). Confirmation of identified glycopeptides was carried out by the identification of unique b- and y-ions in every MS/MS spectra. Elongation by Fringe appears to occur in a site-specific manner. *O*-Fucose sites within EGF repeats 1, 3, 4, 5, 7, 8, 9, 12, 17, 20, 21, 23, 27, 28, 30, and 31 are all detected with extension to the disaccharide, while exclusively monosaccharide was detected at EGF repeats 2, 24, 25, and 26 (Figure 28). The extent of elongation at individual sites will be examined in more detail below.

Equivalent searches for *O*-glucose disaccharides and their hallmark losses of xylose and glucose revealed ions corresponding to peptides containing all of the unmapped predicted *O*-glucosylation sites: EGF repeats 4, 5, 10, 12, 15, 18, 20, 24, 25, 27, and 30 (Table 5, Figures 17, 23, 28). Initial identification of glycopeptides was carried out in a similar fashion as was done for *O*-fucose sites. In some cases, glycopeptides included sequences spanning C1 through C3, allowing for identification of both *O*-glucose and *O*-fucose from the same ion (Table 5, Figure 28). These peptides demonstrated sequential loss of masses corresponding to GlcNAc, fucose, xylose, and glucose, for a total loss of 643 Da. Either GlcNAc or xylose can be lost first, indicating each modification is terminal. One peptide from EGF 11-12 carried an *O*-GlcNAc modification from C5-C6 of EGF 11, as well as *O*-glucose monosaccharide and GlcNAc-Fucose disaccharide from C1-C3 of EGF 12 (Figure 22).

Using the previously refined consensus sequence C^2xxxxS/TC^3, we have identified 12 additional *O*-fucose sites to bring the total to 20 of 22 predicted sites. Similarly, using the newly revised consensus sequence C^1XSX(P/A)C^2, we have also identified 11 additional *O*-glucose sites, to bring that total to 16 of 18 predicted sites. Elongation at sites of *O*-glucosylation appears to be site-specific, mainly at EGF repeats 14-20. In contrast to Fringe elongation of *O*-

fucose sites, extension by the XXYLT results in a clustering of *O*-glucose disaccharide structures towards the middle of the ECD.

Fringe exhibits site-to-site hierarchy of elongation with sub-stoichiometric elongation at multiple sites

We were interested in knowing whether each site of Fringe-elongation served as a substrate for extension to disaccharide equivalently, or if Fringe showed preferential elongation towards a subset of these sites. To examine which sites, if any, are preferentially elongated by Fringe, we used Multiple Reaction Monitoring, a targeted label-free approach to semi-quantitative mass spectrometry that allows monitoring of multiple glycoforms of a peptide within a single chromatographic run (114). The rationale for using MRM instead of a simple EIC is that quantification during a standard MS experiment from which EIC traces are derived relies on the parent ion mass alone. Although MS/MS spectra can confirm that signal being observed is from the parent ion of interest and not an unrelated peptide of same m/z, background noise can still contribute to an EIC trace (see Figures 17C and 21C). MRM follows specified ions in a coupled format, for example a glycoform would be paired with a major product ion (naked peptide), known as the transition ion, and the signal reported is a function of the product ion intensity, therefore a double filter is being applied to each ion being monitored. Unless fragmentation of the parent ion results in generation of the transition ion, it will not contribute to the trace, thereby reducing background noise to inconsequential levels (115). In addition to using equivalent starting amounts of N-EGF:FLAG -/+ Fng for our comparative studies, a non-glycosylated N-EGF:FLAG peptide common to all samples was monitored during each MRM experiment, and

served as an internal standard to ensure similar amounts of sample were being analyzed in separate runs (Figure 24K).

First, to verify Fng elongation was equivalent to previously studied levels, unmodified and Fng modified N-EGF:FLAG were subjected to *in vitro* glycosylation via a galactosyltransferase assay to elongate any GlcNAc present along either receptor, as previously described (Experimental Procedures). Addition of [^{3}H]galactose allows specific quantitation of GlcNAc present along either receptor. Levels were comparable to those described previously, indicating addition of approximately 12 GlcNAc moieties per molecule of N-EGF:FLAG. Samples were then quantitated by SDS-PAGE and densitometry, and amounts equivalent to 1.0 ug of each were acetone precipitated to concentrate. N-EGF:FLAG -/+Fng were then reduced, alkylated, and subjected to in solution tryptic or chymotryptic digests to generate glycopeptides as described above.

To demonstrate that the method is working, we performed MRM analysis on the same *O*-fucosylated glycopeptide bearing disaccharide generated shown in Figure 21A. The disaccharide parent ion (m/z 1138.6), and monosaccharide parent ion (m/z 1070.9) were monitored with naked peptide (m/z 1021.9) serving as the transition ions for both. Naked peptide was also monitored, with an abundant fragmentation ion (m/z 708.8) selected as the transition ion. Figure 24F shows an MRM analysis of EGF 21 from N-EGF:FLAG not modified with Fringe, while 24F' shows an MRM experiment on the same peptide derived from an equivalent amount of Fringe-modified N-EGF:FLAG. As one would expect, no disaccharide is detectable in the –Fng sample. The peptide is found predominantly in the monosaccharide *O*-fucose form, with negligible amounts of naked peptide detected, indicating EGF 21 is *O*-fucosylated at a high stoichiometry, consistent with EIC searches (Figure 21C). In the +Fng sample, we observed the

amount of monosaccharide-modified EGF 21 has decreased by greater than 50%, and we detected disaccharide glycoform amounts increased by an equivalent amount, indicating this site is modified by Fng to an intermediate degree. The amount of disaccharide detected is not equivalent to the amount of monosaccharide that disappears, suggesting that the addition of GlcNAc suppresses ionization of the glycopeptide. Nonetheless, even with variable amount of suppression of ionization caused by the GlcNAc, we can still directly compare levels of glycopeptides with *O*-fucose monosaccharide between samples. Since monosaccharide-modified peptide should ionize to equivalent degrees, and monosaccharide levels should be decreased in a +Fng sample due to either deglycosylation (which would result in increased levels of naked peptide and is unlikely to occur) or because of extension with GlcNAc (which would result in increased levels of disaccharide), a direct comparison of the monosaccharide glycoform can be made to determine the extent of elongation at a specific site.

Figure 24C shows EGF 7 from N-EGF:FLAG not modified with Fringe, while 24C' shows the same peptide derived from an equivalent amount of Fringe-modified N-EGF:FLAG. As expected, we again see no disaccharide detectable in the –Fng sample, and the major glycoform present is the *O*-fucose monosaccharide. Interestingly, in the +Fng sample, we see the amount of monosaccharide-modified EGF 7 has decreased by greater than 50% again, but in contrast to EGF 21, at this site, disaccharide levels increased by an equivalent amount in the +Fng sample, indicating that the GlcNAc is not causing significant suppression of ionization of this glycopeptide. These results suggest addition of GlcNAc causes variable levels of suppression depending on the peptide. Several *O*-fucosylated sites examined resulted in significant decreases in monosaccharide-modified ions in +Fng samples, coupled with increases in disaccharide-modified ion, including EGF repeats 7, 9, 21, 28, and 31, indicating a relatively

high extent of Fringe elongation at these sites (greater than 50% decrease in monosaccharide signal). In contrast, EGF repeats 2, 3, 5, 17, 20, 23, 26, and 30 showed less dramatic decreases in monosaccharide-modified ions in +Fng samples, and had even lower levels of GlcNAc-fucose-modified peptides, indicating a lower preference of Fringe for modification at these sites (less than 50% decrease in monosaccharide signal). A third and final set including EGF repeats 2, 24, 25, and 26 showed negligible amounts of Fringe elongation, with minimal or no change in monosaccharide signal, and minimal or no disaccharide-modified peptides detectable (less than 10% decrease in monosaccharide signal).

Similar analyses were conducted on an *O*-glucosylated glycopeptide bearing disaccharide generated from a chymotryptic digest. Mass to charge ratios to be monitored were determined based on MS/MS analysis (Figure 17 or 28). The disaccharide glycoform parent ion (m/z 784.3) and monosaccharide glycoform parent ion (m/z 740.3) were monitored with naked peptide (m/z 686.1) acting as the transition ion for both glycoforms. Naked peptide was also monitored, with an abundant fragmentation ion (m/z 551.3) serving as the transition ion.

Figure 25C shows an MRM analysis of EGF 16 from N-EGF:FLAG not modified with Fringe, while 25C' shows an MRM experiment on the same peptide derived from an equivalent amount of Fringe-modified N-EGF:FLAG. Both *O*-glucose and *O*-glucose-xylose modified peptide are detected in the –Fng sample. The peptide is found in almost equivalent amounts of monosaccharide and disaccharide-modified glycoforms, with negligible amounts of naked peptide detected, indicating EGF 16 is *O*-glucosylated at a high stoichiometry. Interestingly, in the +Fng sample, we can see the amount of monosaccharide-modified EGF 16, as well as disaccharide-modified EGF 16, have both more than doubled. Nonetheless, the relative ratios of *O*-glucose mono- and disaccharides do not vary between – and + Fng samples. Figure 25B

shows EGF 5 from N-EGF:FLAG not modified with Fringe, while 25B' shows the same peptide derived from an equivalent amount of Fringe-modified N-EGF:FLAG. Consistent with CNL data for this site (Figures 23 and 28), no disaccharide is detected in the –Fng or +Fng sample, and the major glycoform present is the *O*-glucose monosaccharide. In the presence of Fng, we do see an increase in the total amount of monosaccharide-modified EGF 5.

The virtually undetectable levels of naked peptide coupled with the detection of *O*-glucosylation at EGF repeats 5, 16, and additional sites (Figure 28) indicates that Rumi is modifying Notch at high stoichiometries across the receptor. The lack of xylose at EGF 5 indicates *O*-glucose is elongated in a site-specific manner. When all sites of xylosylation are examined, a pattern of substoichiometric xylosylation emerges, with sites in the middle region of the ECD, from EGF 15-20 being extended to disaccharide, but not at any of the other *O*-glucose sites (Figure 23).

Discussion

Of the over 50 proteins predicted to be *O*-fucosylated and or *O*-glucosylated, Notch contains the greatest number of consensus sites for both forms of *O*-glycosylation. Although *Drosophila* Notch has been shown to be modified with *O*-fucose and *O*-glucose glycans (8,96), the majority of predicted sites remained unmapped, and the extent to which Fringe and GXYLT elongate all or only a subset of these sites has not been explored. The results presented here confirm that the extracellular domain of N-EGF:FLAG is extensively *O*-fucosylated and *O*-glucosylated. Using the consensus sequence $C^2xxxx(\underline{S/T})C^3$, we have identified peptides spanning this region for 13 of the 15 unmapped EGF repeats of N-EGF:FLAG predicted to be modified by OFUT1, and confirmed that each of these 13 sites is *O*-fucosylated using nano-LC-ESI-MS/MS (Figure 23). EIC searches for *O*-fucosylated and non-glycosylated (naked) peptide confirmed that the major species for each site was the glycosylated ion (Figure 21C), indicating N-EGF:FLAG is *O*-fucosylated at high stoichiometries across the receptor. EIC searches were also used to confirm that no *O*-fucose site reported here was elongated beyond the disaccharide (data not shown). These findings are not surprising, as previous studies we have carried out on Notch and its mammalian homologs have demonstrated that peptides detected from regions predicted to be *O*-fucosylated are effectively occupied with this *O*-glycan moiety, and Galactosyltransferase assays confirm that further elongated species are not present on *Drosophila* Notch expressed in S2 cells (Figure 20) (96).

To determine whether there is a hierarchy among fucosylated sites to serve as substrates for Fringe elongation, we implemented Multiple Reaction Monitoring mass spectrometry. Since suppression of ionization is a concern for ions bearing GlcNAc modifications (compare Figures 24C and F), we used decrease in the level of *O*-fucose monosaccharide glycoforms as a measure

of Fringe-mediated elongation at individual sites. Using this method, we categorized *O*-fucosylation sites into highly modified by Fringe (>50% decrease in monosaccharide signal between – and +Fng N-EGF:FLAG samples), less highly modified (<50% decrease in monosaccharide signal), and largely unmodified by Fringe (<10% disaccharide signal). Using Multiple Alignment using Fast Fourier Transform (MAFFT) with CLUSTALW output to compare these sites, we find that there is a strong conservation of an acidic residue in position 18 and non-polar in position 20 of EGF repeats that are highly Fng elongated (>50% disaccharide) (Figure 27). This conservation is observed to a lesser degree in EGF repeats not as greatly modified (<50% disaccharide), suggesting that these residues may play a role in recognition of the EGF repeats by Fringe.

Studies on mouse Notch1 to elucidate preferential sites of elongation by Fringe have included examination of sites of Fringe elongation by metabolic radiolabeling, alkali-induced β-elimination, and gel filtration chromatography. In these studies, short fragments of Notch1 ECD were expressed in attempts to isolate which regions were Fringe-elongated (19). No elongation by Manic Fringe was detected on EGF repeats 1-5, but elongation was detected at other regions of the ECD, particularly between EGF 24-28. Further examination of EGF repeats in this region revealed mammalian Fringes are capable of recognizing individual EGF repeats for modification, as evidenced by EGF 26 serving as a substrate for elongation but not EGF 24. Interestingly, unlike the case with mouse Notch1, we see *in vivo* Fringe elongation across the *Drosophila* Notch ECD, including at the N-terminal region. This is consistent with a previous report that demonstrated Fng can modify *O*-fucose on *Drosophila* Notch EGF 1-3 fragment *in vitro* (60). Rampal *et al.* examined specificity determinants for Fringe recognition in a mammalian system. In these studies, our lab concluded that residues 24 and 28, flanking the outside of C4 and C5,

were contributing to Fringe substrate recognition, though the specific residues preferred for elongation varied between the three mammalian Fringe homologs (93). In contrast, residues 18 and 20 in Notch appear just after C3, in close proximity to the site of *O*-fucosylation. Taken together, these data are reflective of differences in Fringe specificities between mammals and *Drosophila*. Further studies will need to be performed to examine the molecular basis of Fringe specificity in more detail.

While we have reported on elongation of *O*-glucose to a Xyl-α1,3-Xyl-α1,3-Glc-*O*-Ser structure previously (22), and more extensively for a recent analysis of *O*-glucosylation on mouse Notch1 (Chapter II), xylosylation of *O*-glucose to a di- or trisaccharide structure has never been reported on *Drosophila* Notch. Findings from our previous site-mapping work on Notch involving sites of *O*-glucosylation were restricted to reports of monosaccharide modifying five sites along the receptor (EGF repeats 14, 16, 17, 19, and 36) (8). To verify occupancy and glycan structures at the 18 predicted sites of *O*-glucosylation on *Drosophila* Notch, the newly revised consensus sequence for *O*-glucosylation, $C^1xSx(P/A)C^2$ was used to identify peptides spanning this region for 16 of the 18 EGF repeats predicted to be modified by Rumi. All 16 predicted sites detected were found to be *O*-glucosylated using mass spectrometry. A comprehensive summary of all site-mapping can be seen in Figure 23. Interestingly, although *O*-glucosylation is seen at sites across the length of the ECD, *O*-glucose elongation by GXYLT is observed to be substoichiometric, and restricted to one region of the receptor from EGF 15-20. This xylosylated region coincides with a stretch of the ECD that is calcium-binding EGF repeat-rich (56). EGF repeats 14-20 are all predicted to coordinate calcium between the flexible linker regions that connect neighboring EGF repeats, which has been previously shown to impart structural rigidity to these motifs in human Notch1 (73). Analysis by MRM confirmed that there

appears to be a hierarchy amongst *O*-glucose sites for xylosylation (Figure 25). Independent of the presence of Fng, each predicted site was *O*-glucosylated, and *O*-glucose found on EGF repeats 14-20 was elongated to disaccharide, as seen with EIC and MS/MS data. Surprisingly, the amount of mono- and or disaccharide detected in the Fng-modified samples was increased by greater than 200% for all *O*-glucosylation sites monitored, in spite of normalization with an internal standard peptide. One possible explanation includes Fringe elongation at a neighboring site may be increasing protease efficiency at sites generating these glycopeptides in +Fng samples. As Fng is elongating *O*-fucosylation sites to disaccharides, it may be causing conformational changes in the overall structure of the receptor, exposing a different set of sites to modification by glycosyltransferases than would occur in the absence of Fng. This scenario does not explain why glucosylation levels are also increasing though, as *O*-glucose is transferred in the ER, while Fng modifies later in the Golgi. Although the reason for this change is unclear, the data strongly suggests that Fringe does not influence the relative amount of elongation on *O*-glucose consistent with the idea that the elongation of *O*-glucose and *O*-fucose is not linked.

Alignments were also carried out for *O*-glucosylated EGF repeats. Sites were again distributed into two categories: group 1 included all sites that where no elongation to disaccharide was detected, whereas group 2 consisted of all *O*-Glc-Xyl modified sites. Using MAFFT, we find that there are multiple residues within each set that may contribute to GXYLT recognition and elongation (Figure 27). While mono- and disaccharide-modified EGF repeats have an acidic residue 2 positions N-terminal to C1, only xylosylated EGF repeats contain acidic residues in all 4 positions N-terminal to C1. In addition, xylosylated EGF repeats also have Asn or Ser between the Serine and Proline of the *O*-glucose consensus sequence. Other residues of interest include a highly conserved Asp two positions C-terminal to C3, followed by a neutral

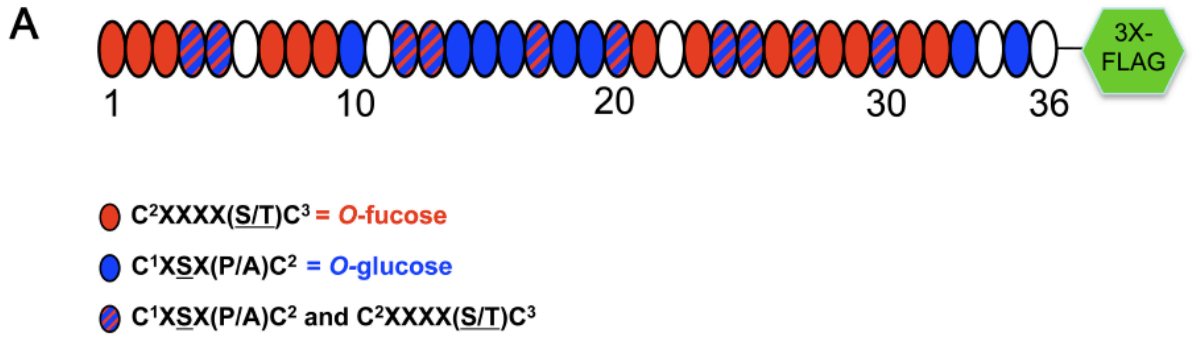
non-polar residue to positions C-terminal to the Asp. Non-xylosylated EGF repeats contain random amino acids at these positions. Further work needs to be done to test whether these sites play an important role in recognition of O-glucosylated EGF repeats by GXYLT enzymes.

In *Drosophila* there appears to be only one homolog to the mouse GXYLT enzymes (CG9996). It shares 51.9% homology with GXYLT1 and 50% homology with GXYLT2. In contrast to Notch expressed in S2 cells, in the murine system we observe all *O*-glucose sites are elongated on mouse Notch1. It is conceivable that each homolog is responsible for elongation of a subset of these *O*-glucose sites. If that is the case, it is plausible that the single fly GXYLT is capable of modifying only a subset of *O*-glucose sites, in common with one of the mammalian homologs, and would explain why we do not see global xylosylation at all *O*-glucose sites across the Notch ECD. The results of the alignments may be reflective of key amino acids important for this potential differential recognition between GXYLT homologs. Since S2 cells are an embryonic fly cell line that do not express endogenous Notch or Fng, it is conceivable that an as yet unidentified, and possibly more distantly related second homolog exists in flies, which is not expressed specifically in S2 cells. Although homology searches within the fly genome should rule out this possibility, unexpected glycosyltransferases that do not share a high degree of homology with their glycosyltransferase class/family have been previously identified. Differential expression of one GXYLT in S2 cells could also account for the substoichiometric elongation we observe in this system. If this is the case, tissue specific expression of GXYLTs in mammals has the potential for causing the same effect.

Table 5: Peptides from *Drosophila* Notch identified with *O*-fucose, *O*-glucose, and or *O*-GlcNAc modifications.

Peptides were identified by sequential neutral loss of masses corresponding to GlcNAc, fucose, xylose, and glucose, as shown in Figures 17 and 21. Representative spectra for each glycopeptide identified here are shown in Figures 17, 21, 22, and 28 (A-X). All masses were converted to the equivalent of singly charged ($M+H^+$) for the table. For each glycopeptide, the mass of the parent ion, the fully deglycosylated product (lacking all sugars), and the difference between these (corresponding to the mass of the modifications: *O*-glucose disaccharide, 294 Da; *O*-glucose monosaccharide, 162Da; *O*-fucose disaccharide, 349 Da, *O*-fucose monosaccharide, 146 Da, *O*-GlcNAc monosaccharide, 203 Da). The predicted mass of the unglycosylated peptide is also shown. All peptide masses are adjusted for carboxyamidomethylation of Cysteines. For peptides with a mass below 2000 Da, monoisotopic masses were used. For those above 2000 Da, average masses were used.

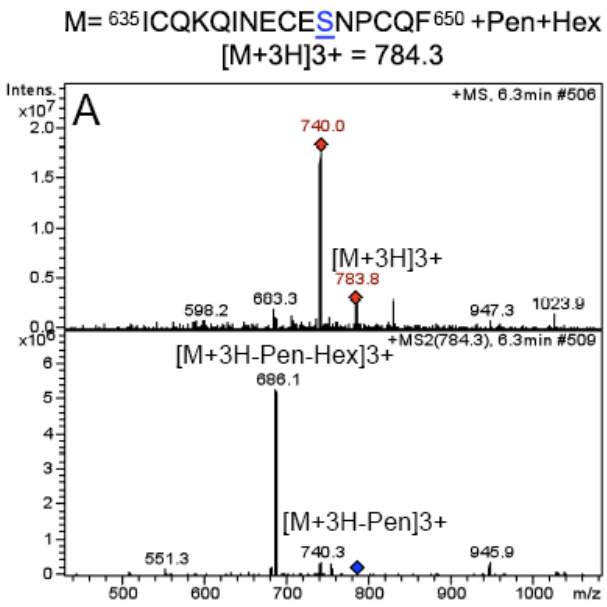
EGF	Sequence	Parent Ion ($M+H^+$)	Deglyco Product ($M+H^+$)	Mass Δ	Pred. Mass ($M+H^+$)
1	^{83}CACDSHYVGDYCEHRNPCNSMRCQNGGTCQVTFR^{116}	4501.0	4193.0	308.0	4199.5
2	^{105}CQNGGTCQVTFR^{116}	1574.6	1428.0	146.6	1428.6
3	^{126}CPLGFDESLCEIAVPNACDHVTCLNGGTCQLK^{157}	3988.3	3636.7	351.6	3638.1
4	^{116}CASSPCRNGATCALAGSSSF^{134}	2472.1	2163.1	309.0	2163.4
5	^{217}DIEECQSNPCKY^{228} ^{228}YGGTCVNTHGSYQCMCPTGYTGK^{250}	1705.6 2950.3	1543.2 2600.5	162.4 349.8	1542.6 2600.9
7	^{289}NCEQNYDDCLGHLCQNGGTCIDGISDYTCR^{318}	3947.8	3597.7	350.1	3597.8
8	^{328}CQDDVDECAQRDHPVCQNGATCTNTHGSY^{356}	3746.8	3397.0	349.8	3397.5
9	^{378}QAACFYGATCIDGVGSFYCQCTK^{400}	2812.4	2665.6	146.8	2665.9
10	^{403}TGLLCHLDDCTSNPCHADAICDTSPINGSYACSCATGYK^{442}	4931.4	4785.4	146.0	4782.8
11-12	^{481}TGPRCETNINECESHPCQNEGSCLDDPGTF^{510}	4197.7	3487.0	710.7	3483.7
14	^{561}CQINIDDCQSQPCR^{574}	2098.4	1794.0	295.4	1793.7
15	^{577}GICHDSIAGYSCECPPGYTGTSCEININDCDSNPCHR^{613}	4571.8	4277.0	294.8	4276.6
16	^{639}QINECESNPCQFDGHCQDR^{657}	2689.9	2395.3	294.6	2395.5
17	^{673}NCEVNVNECHSNPCNNGATCIDGINSYK^{700}	3754.3	3242.5	511.8	3242.5
18	^{715}NVDECISSPCANNGVCIDQVNGYK^{738}	3142.3	2714.8	427.8	2714.9
19	^{744}GFYDAHCLSDVDECASNPCVNEGR^{767}	3068.2	2772.7	295.5	2773.9
20	^{788}CELDIDECSSNPCGHGGTCYDK^{809}	3158.5	2647.0	511.5	2646.8
21	^{826}CETNIDDCVTNPCGNGGTCIDKVNGYK^{852}	3413.8	3063.7	350.1	3063.3
23	^{919}NGASCLNVPGSYR^{931}	1745.6	1396.0	349.6	1394.6
24	^{942}DCAINTDDCASFPCQNGGTCLDGIGDYSCLCVDGFDGK^{979}	4545.8	4236.2	309.6	4234.6
25	^{972}CVDGFDGKHCETDINECLSQPCQNGATCSQY^{1002}	3961.4	3652.6	308.8	3652.9
26	^{1033}NGGSCIDGINSY^{1044}	1373.2	1226.8	146.4	1226.5
27	^{1064}CDSNPCLNGATCHEQNNEYTCHCPSGFTGK^{1093}	4029.8	3517.4	512.4	3517.8
28	^{1094}QCSEYVDWCGQSPCENGATCSQMK^{1117}	3030.4	2884.0	146.4	2884.1
30	^{1180}CQKEIDECQSQPCQNGGTCRDLIGAY^{1205}	3398.8	3089.5	309.3	3089.4
31	^{1211}QGQNCELNIDDCAPNPCQNGGTCHDRVM^{1239}	3511.0	3364.0	147.0	3364.6
35	^{1371}NCELSGQDCDSNPCR^{1385}	1975.2	1812.2	163.0	1811.6

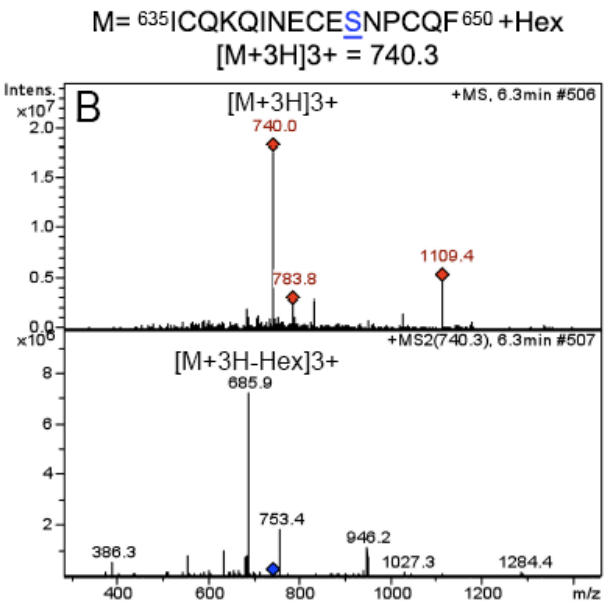


B

CTSVGCQNGGTCVTQLNGKTYCACDSHYVGDYCEHRNPCNSMRCQNGGTCQVTFRNG
RPGISCKCPLGFDESLCEIAVPNACDHVTCLNGGTCQLKTLEEYTCACANGYTGERCETKN
LCASSPCRNGATCTALAGSSSFTCSCPPGFTGDTCSYDIEECQSNPCKYGGTCVNTHGSYQ
CMCPTGYTGKDCDTKYKPCSPSPCQNGGICRSNGLSYECKCPKGFEGKNCEQNYDDCLG
HLCQNGGTCIDGISDYTCRCPPNFTGRFCQDDVDECAQRDHPVCQNGATCTNTHGSYSC
ICVNGWAGLDCSNNTDDCKQAACFYGATCIDGVGSFYCQCTKGKTGLLCHLDDACTSNP
CHADAICDTSPINGSYACSCATGYKGVDCSEDIDECDQGSPCEHNGICVNTPGSYRCNCS
QGFTGPRCETNINECESHPCQNEGSCLDDPGTFRCVCMPGFTGTQCEIDIDECQSNPCLND
GTCHDKINGFKCSCALGFTGARCQINIDDCQSQPCRNRGICHDSIAGYSCECPPGYTGTSC
EININDCDSNPCHRGKCIDDVNSFKCLCDPGYTGYICQKQINECESNPCQFDGHCQDRVGS
YYCQCQAGTSGKNCEVNVNECHSNPCNNGATCIDGINSYKCQCVPGFTGQHCEKNVDEC
ISSPCANNGVCIDQVNGYKCECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCP
PGYTGKRCELDIDECSSNPCQHGGTCYDKLNAFSCQCMPGYTGQKCETNIDDCVTNPCG
NGGTCIDKVNGYKCVCKVPFTGRDCESKMDPCASNRCKNEAKCTPSSNFLDFSCTCKLGY
TGRYCDEDIDECSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPCQNGG
TCLDGIGDYSCLCVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCPLGFSGINC
QTNDEDCTESSCLNGGSCIDGINGYNCSCLAGYSGANCQYKLNKCDSNPCLNGATCHEQ
NNEYTCHCPSGFTGKQCSEYVDWCGQSPCENGATCSQMKHQFSCKCSAGWTGKLCDVQ
TISCQDAADRKGLSLRQLCNNGTCKDYGNSHVCYCSQGYAGSYCQKEIDECQSQPCQNG
GTCRDLIGAYECQCRQGFQGQNCELNIDDCAPNPCQNGGTCHDRVMNFSCSCPPGTMGI
ICEINKDDCKPGACHNNGSCIDRVGGFECVCQPGFVGARCEGDINECLSNPCSNAGTLDC
VQLVNNYHCNCRPGHMGRHCEHKVDFCAQSPCQNGGNCNIRQSGHHCICNNGFYGKN
CELSGQDCDSNPCRVGNCVVADEGFGYRCECPRGTLGEHCEIDTLDECSPNPCAQGAACE
DLLGDYECLCPSKWKGKRCD

Figure 16. Predicted sites of *O*-glycosylation on N:EGF-FLAG.
A. Domain map of Notch-EGF:FLAG, which contains 36 EGF repeats of *Drosophila* Notch, and a C-terminal triple FLAG tag. Predicted sites of *O*-glycosylation, based on the presence of consensus sequences for *O*-glucosylation ($C^1xSx(P/A)C^2$) and *O*-fucosylation (C^2xxxxS/TC^3) within EGF repeats of *Drosophila* Notch. Blue ovals indicate predicted *O*-glucose sites, red ovals indicate predicted *O*-fucose sites, and hashed-line ovals designate EGF repeats containing both consensus sequences. B. Amino acid sequence of *Drosophila* Notch EGF repeats 1-36. Sequence highlighted in blue indicates portions of the protein identified by manual MS/MS data searches coupled with fragment ion confirmation, as well as database searching using MASCOT (sequence coverage is >70%).


M= 635ICQKQINECESNPCQF650 +Pen+Hex
[M+3H]3+ = 784.3
A
+MS, 6.3min #506
740.0
[M+3H]3+
783.8
598.2
663.3
947.3
1023.9
+MS2(784.3), 6.3min #509
[M+3H-Pen-Hex]3+
686.1
[M+3H-Pen]3+
551.3
740.3
945.9
m/z


M= 635ICQKQINECESNPCQF650 +Hex
[M+3H]3+ = 740.3
B
[M+3H]3+
+MS, 6.3min #506
740.0
1109.4
783.8
+MS2(740.3), 6.3min #507
[M+3H-Hex]3+
685.9
753.4
946.2
386.3
1027.3
1284.4
m/z

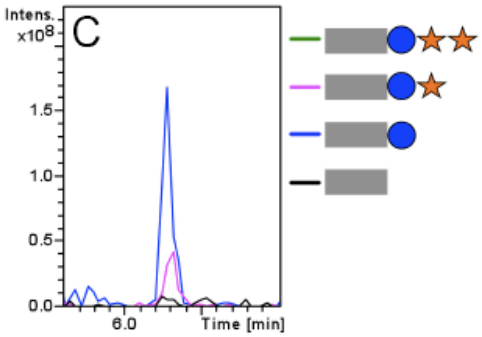

C
Time [min]

Figure 17. *Drosophila* Notch is modified with *O*-glucose at high stoichiometries. N-EGF:FLAG produced in S2 cells -/+Fng were subjected to reduction, alkylation, and chymotryptic digest as described (Experimental Procedures). **A.** MS spectrum of an ion from EGF 16. Top panel: M corresponds to the mass of the peptide (sequence shown above spectra) plus Pentose (xylose) and Hexose (glucose). The other ions present in the MS spectrum represent other peptides eluting from the LC at the same time. Red diamonds indicate ions selected for CID fragmentation. Bottom panel, MS/MS spectrum of m/z 784.3 ion ($[M+3H]^{3+}$) from the MS spectrum. Blue diamond shows the position of the parent ion prior to fragmentation. The major fragments show the sequential loss of a Pentose ($[M+3H\text{-}Pen]^{3+}$, m/z 740.3) and a Hexose ($[M+3H\text{-}Pen\text{-}Hex]^{3+}$, m/z 686.1) from the parent ion. $[M+3H\text{-}Pen\text{-}Hex]^{3+}$ corresponds to the triply charged mass of a chymotryptic peptide from EGF 16 of Notch that contains an *O*-glucose consensus site (highlighted in blue above spectra; see Table 5 for predicted masses). Further confirmation of this assignment comes from peptide fragment ions (b- and/or y-ions) in the MS/MS spectrum. **B.** MS spectrum of an ion from EGF 16. Top panel: M corresponds to the mass of the peptide (sequence shown above spectra) plus Hexose (Glucose). The other ions present in the MS spectrum represent other peptides eluting from the LC at the same time. Red diamonds indicate ions selected for CID fragmentation. Bottom panel, MS/MS spectrum of m/z 740.3 ion ($[M+3H]^{3+}$) from the MS spectrum. Blue diamond shows the position of the parent ion prior to fragmentation. The major fragment indicates the loss of a Hexose ($[M+3H\text{-}Hex]^{3+}$, m/z 686.1) from the parent ion. $[M+3H\text{-}Hex]^{3+}$ corresponds to the triply charged mass of a chymotryptic peptide from EGF 16 of Notch that contains an *O*-glucose consensus site as for A. Further confirmation of this assignment comes from peptide fragment ions (b- and/or y-ions) in the MS/MS spectrum. **C.** EIC searches for naked peptide (black), *O*-glucose monosaccharide (blue), *O*-glc-xyl disaccharide (pink), and *O*-glc-xyl-xyl trisaccharide (green) show no elongation beyond the disaccharide.

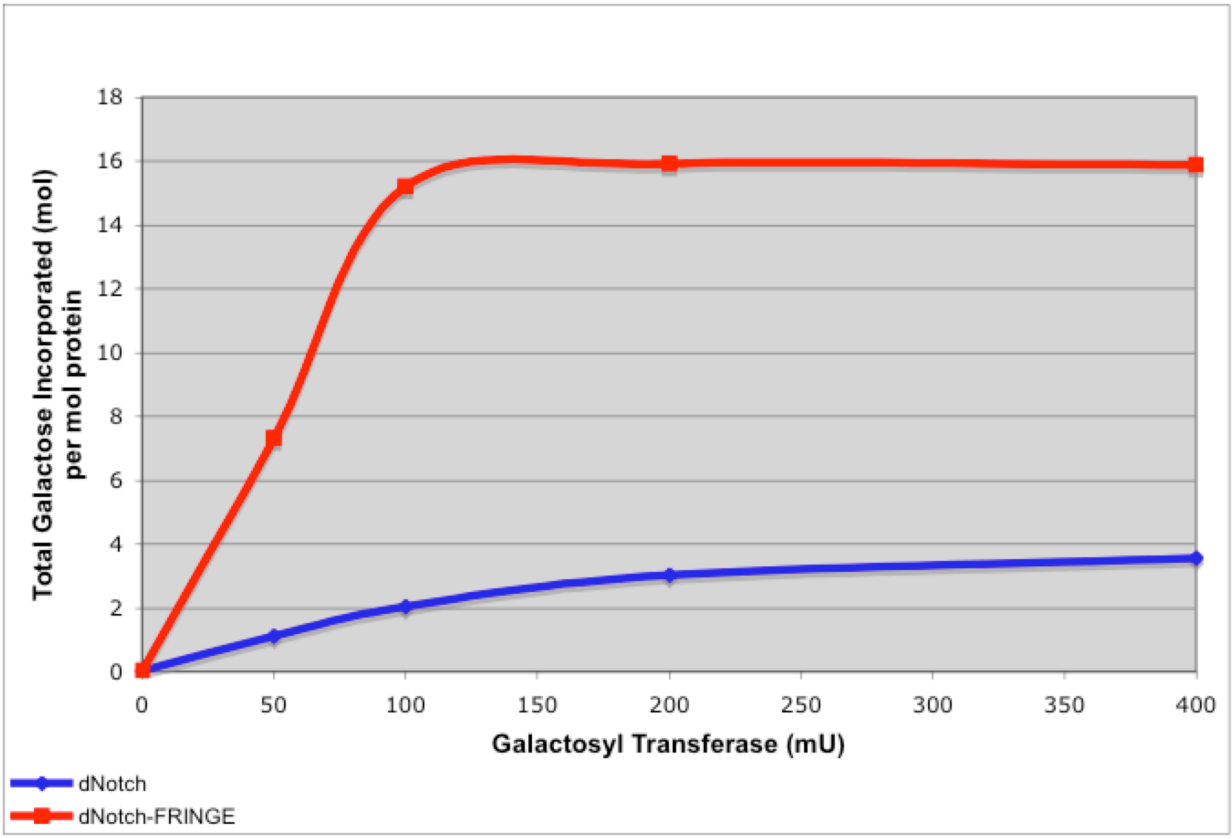


Figure 18. Quantitation of Fringe elongation of *O*-fucose on *Drosophila* Notch.
Purified N-EGF:FLAG produced in S2 cells in the absence or presence of Fng were glycosylated *in vitro* using saturating levels of β4GalT-1 and UDP-[^{3}H]galactose, in a standard galactosyltransferase assay. Samples were then separated on a Sephadex G-50 column and analyzed for radioactivity by scintillation counting. GlcNAc incorporation (Fng elongation) was quantified by measuring the amount of *O*-fucose-GlcNAc-[^{3}H]galactose trisaccharide generated. In the absence of Fng (blue), approximately 3.6 mol galactose were incorporated into 1 mol of protein. In contrast, Fng-modified N-EGF:FLAG incorporated 15.9 mol galactose per mol Notch, suggesting ~12 GlcNAc residues were added to each molecule of Notch by Fng.

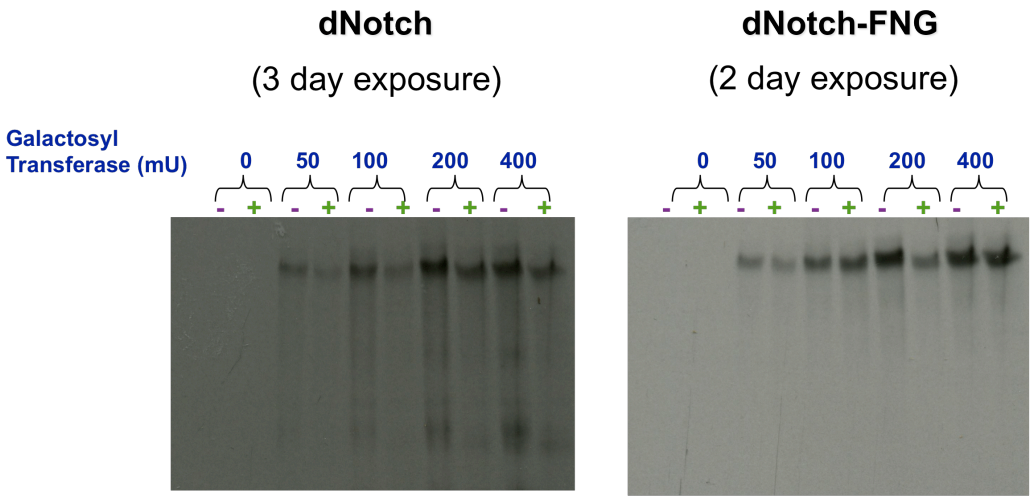


Figure 19. Removal of *N*-linked glycans by PNGaseF shows majority of radiolabel is incorporated into *O*-linked glycans.
N-EGF:FLAG produced in S2 cells in the absence (-) or presence (+) of Fng were glycosylated *in vitro* using saturating levels of β4GalT-1 and UDP-[^{3}H]galactose, in a standard galactosyltransferase assay (Figure 18). Samples were then separated on a Sephadex G-50 column, and treated with PNGaseF to remove *N*-glycans which could also incorporate the radiolabel. Removal of *N*-glycans was monitored by fluorography for – Fng (left) and +Fng (right) samples. Although a minor shift in molecular weight is detectable in both sample sets, indicative of the loss of some *N*-linked glycans, the intensities of the bands after PNGaseF treatment largely unaltered, indicating the majority of [^{3}H]galactose was incorporated into *O*-glycans.

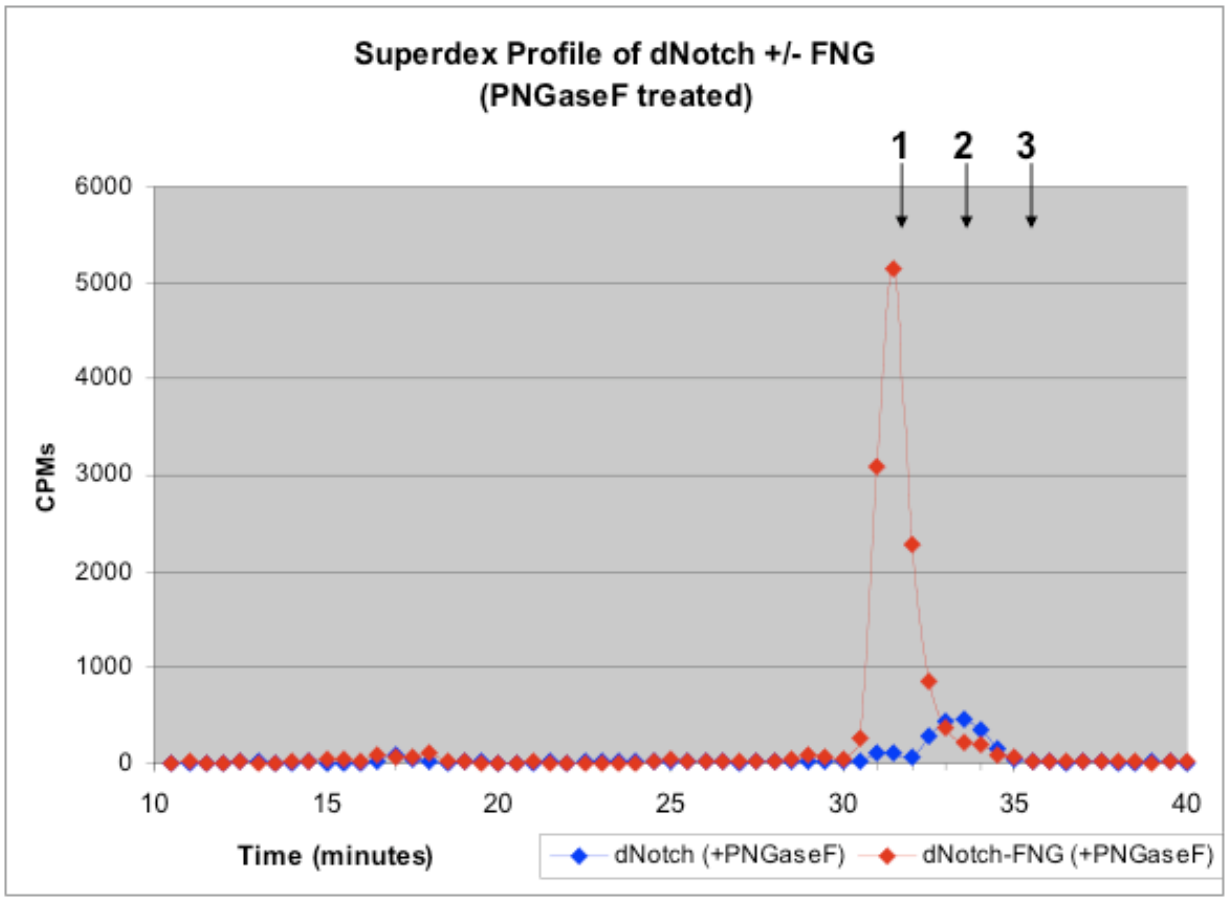


Figure 20. *O*-Fucose on Notch is highly modified by Fringe, and is not elongated beyond disaccharide in S2 cells.
Purified unmodified or Fng-modified N-EGF:FLAG generated in S2 cells was radiolabeled with saturating amounts of β4GalT-1 and UDP-[^{3}H]galactose. *O*-linked glycans were released by alkali-induced β-elimination following PNGaseF digestion to remove *N*-glycans (Figure 19) and were analyzed by gel filtration chromatography using a Superdex peptide column. The migration positions of standards are as follows: 1= Gal-β1,4-GlcNAc-β1,3-fucitol, 2=Glc-β1,3-fucitol, 3= fucitol. The radiolabeled *O*-glycans from a N-EGF:FLAG+Fng sample migrated at the size of Gal-β1,4-GlcNAc-β1,3-fucitol, the expected product of β4GalT radiolabeling of the GlcNAc-β1,3-fuc disaccharide. No trisaccharide was detected in the –Fng sample, consistent with absence of GlcNAc-β1,3-fuc. Disaccharide appears in minute quantities in this sample, likely the result of minor galactose incorporation into a monosaccharide such as *O*-GlcNAc.

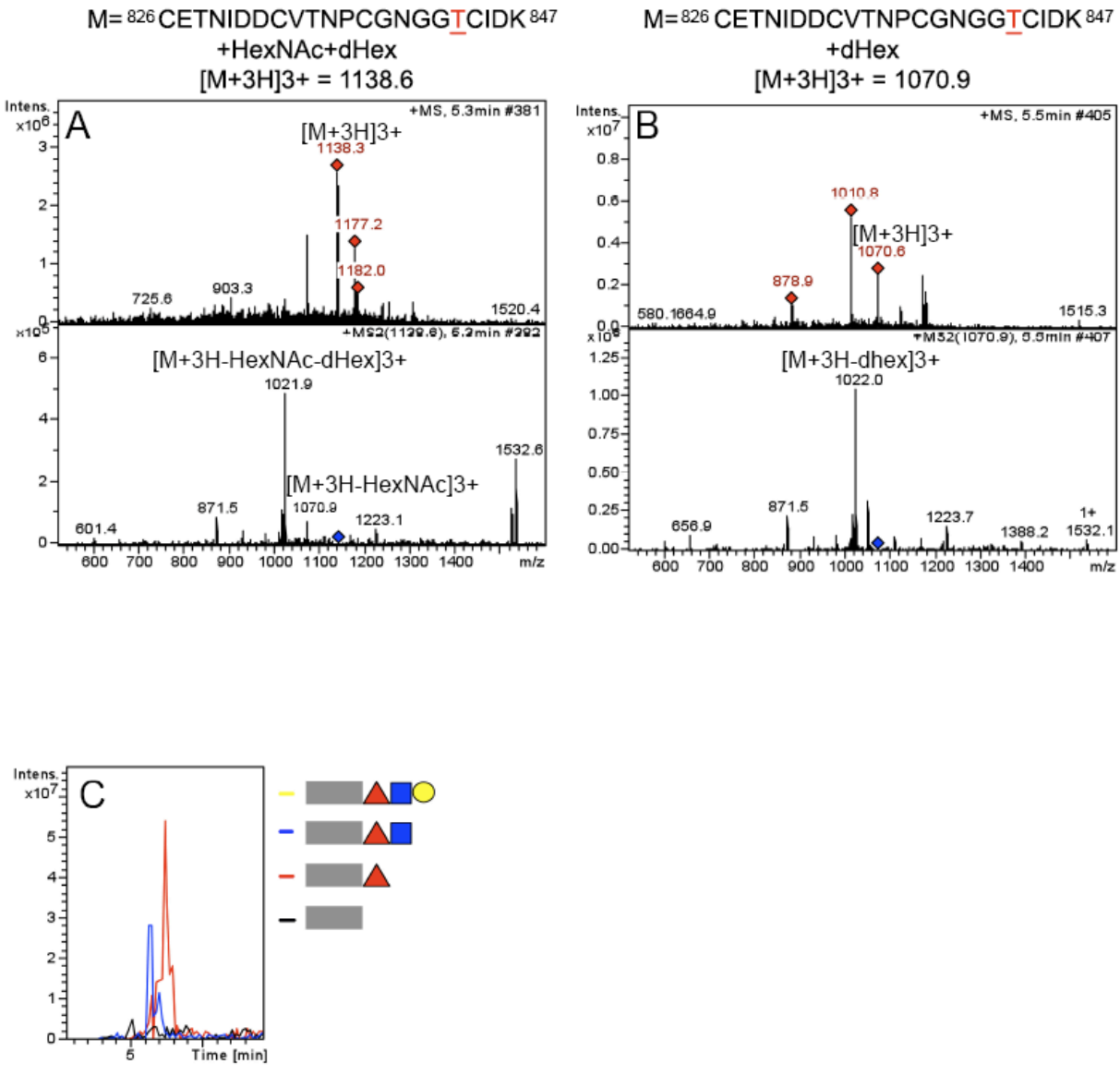

M= 826 CETNIDDCVTNPCGNGGTCIDK 847
+HexNAc+dHex
[M+3H]3+ = 1138.6
A
+MS, 5.3min #381
[M+3H]3+
1138.3
1177.2
1182.0
725.6
903.3
1520.4
+MS2(1138.6), 5.2min #392
[M+3H-HexNAc-dHex]3+
1021.9
1532.6
[M+3H-HexNAc]3+
871.5
1070.9
1223.1
601.4
m/z
M= 826 CETNIDDCVTNPCGNGGTCIDK 847
+dHex
[M+3H]3+ = 1070.9
B
+MS, 5.5min #405
1010.8
[M+3H]3+
1070.6
878.9
580.1664.9
1515.3
+MS2(1070.9), 5.5min #407
[M+3H-dhex]3+
1022.0
871.5
656.9
1223.7
1388.2
1+
1532.1
m/z
C
Time [min]

Figure 21. *Drosophila* Notch is modified with *O*-fucose at high stoichiometries.
N-EGF:FLAG produced in S2 cells -/+Fng were subjected to reduction, alkylation, and tryptic digest as described (Experimental Procedures). **A.** MS spectrum of an ion from EGF 21. Top panel: M corresponds to the mass of the peptide (sequence shown above spectra) plus HexNAc (GlcNAc) and dHexose (fucose). The other ions present in the MS spectrum represent other peptides eluting from the LC at the same time. Red diamonds indicate ions selected for CID fragmentation. Bottom panel, MS/MS spectrum of m/z 1138.6 ion ($[M+3H]^{3+}$) from the MS spectrum. Blue diamond shows the position of the parent ion prior to fragmentation. The major fragments show the sequential loss of a HexNAc ($[M+3H\text{-}HexNAc]^{3+}$, m/z 1070.9) and a dHexose ($[M+3H\text{-}HexNAc\text{-}dHex]^{3+}$, m/z 1021.9) from the parent ion. $[M+3H\text{-}HexNAc\text{-}dHex]^{3+}$ corresponds to the triply charged mass of a tryptic peptide from EGF 21 of Notch that contains an *O*-fucose consensus site (highlighted in red above spectra; see Table 5 for predicted masses). Further confirmation of this assignment comes from peptide fragment ions (b- and/or y-ions) in the MS/MS spectrum. **B.** MS spectrum of an ion from EGF 21. Top panel: M corresponds to the mass of the peptide (sequence shown above spectra) plus dHexose (fucose). The other ions present in the MS spectrum represent other peptides eluting from the LC at the same time. Red diamonds indicate ions selected for CID fragmentation. Bottom panel, MS/MS spectrum of m/z 1070.9 ion ($[M+3H]^{3+}$) from the MS spectrum. Blue diamond shows the position of the parent ion prior to fragmentation. The major fragment indicates the loss of a dHexose ($[M+3H\text{-}dHex]^{3+}$, m/z 1021.9) from the parent ion. $[M+3H\text{-}dHex]^{3+}$ corresponds to the triply charged mass of a tryptic peptide from EGF 21 of Notch that contains an *O*-fucose consensus site as for A. Further confirmation of this assignment comes from peptide fragment ions (b- and/or y-ions) in the MS/MS spectrum **C.** EIC searches for naked peptide (black), *O*-fucose monosaccharide (red), *O*-fuc-GlcNAc disaccharide (blue), and *O*-fuc-GlcNAc-GlcA trisaccharide (yellow) show no elongation beyond the disaccharide.

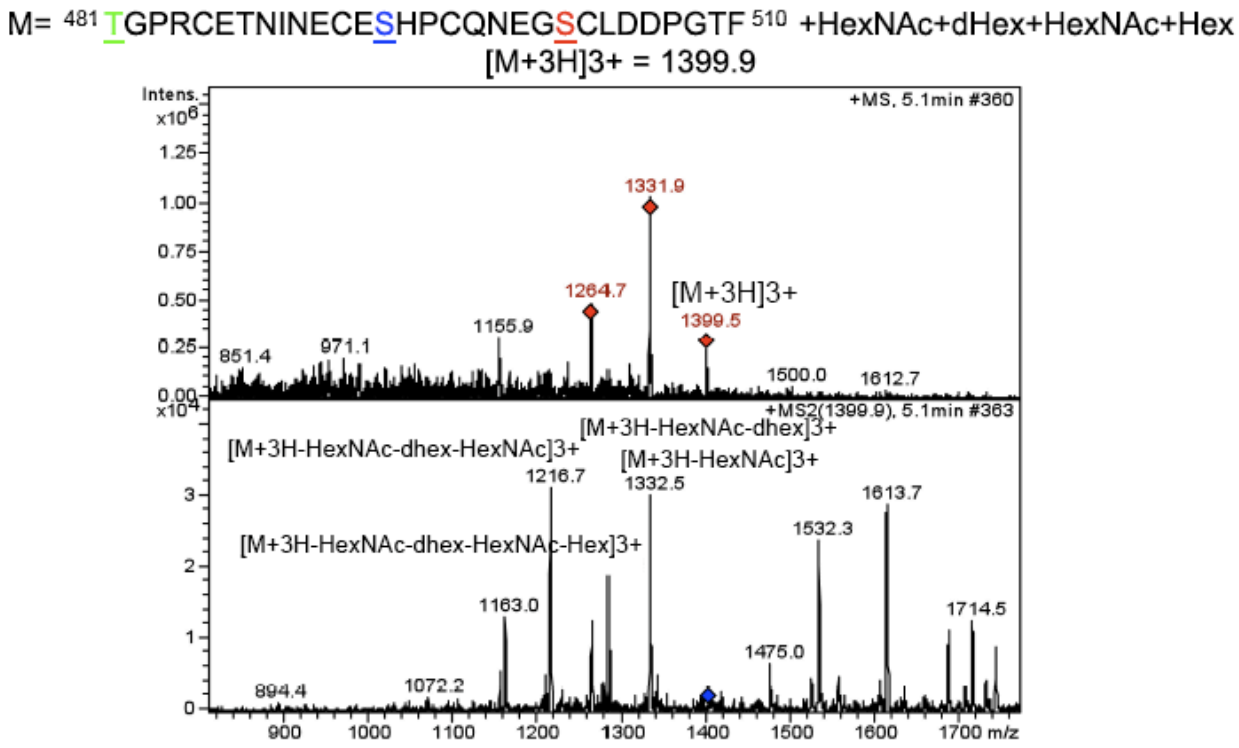


Figure 22. EGF 11 is *O*-GlcNAcylated, and EGF 12 is modified with *O*-fucose disaccharide and *O*-glucose monosaccharide.
MS spectrum of an ion from EGF 11-12. Top panel: M corresponds to the mass of the peptide (sequence shown above spectra) plus HexNAc (GlcNAc), dHexose (fucose), HexNAc (GlcNAc), and Hexose (glucose). The other ions present in the MS spectrum represent other peptides eluting from the LC at the same time. Red diamonds indicate ions selected for CID fragmentation. Bottom panel, MS/MS spectrum of m/z 1399.9 ion ($[M+3H]^{3+}$) from the MS spectrum. Blue diamond shows the position of the parent ion prior to fragmentation. The major fragments show the sequential loss of a HexNAc ($[M+3H\text{-}HexNAc]^{3+}$, m/z 1332.5), dHexose ($[M+3H\text{-}HexNAc\text{-}dHex]^{3+}$, m/z 1264.8), a second HexNAc ($[M+3H\text{-}HexNAc\text{-}dHex\text{-}HexNAc]^{3+}$, m/z 1216.7), and a Hexose ($[M+3H\text{-}HexNAc\text{-}dHex\text{-}HexNAc\text{-}Hex]^{3+}$, m/z 1163.0), from the parent ion. $[M+3H\text{-}HexNAc\text{-}dHex\text{-}HexNAc\text{-}Hex]^{3+}$ corresponds to the triply charged mass of a chymotryptic peptide spanning EGF 11-12 of Notch that contains *O*-glucosylation, *O*-fucosylation, and *O*-GlcNAcylation sites (highlighted within the sequence in blue, red, and green respectively; see Table 5 for predicted masses). Further confirmation of this assignment comes from peptide fragment ions (b- and/or y-ions) in the MS/MS spectrum.

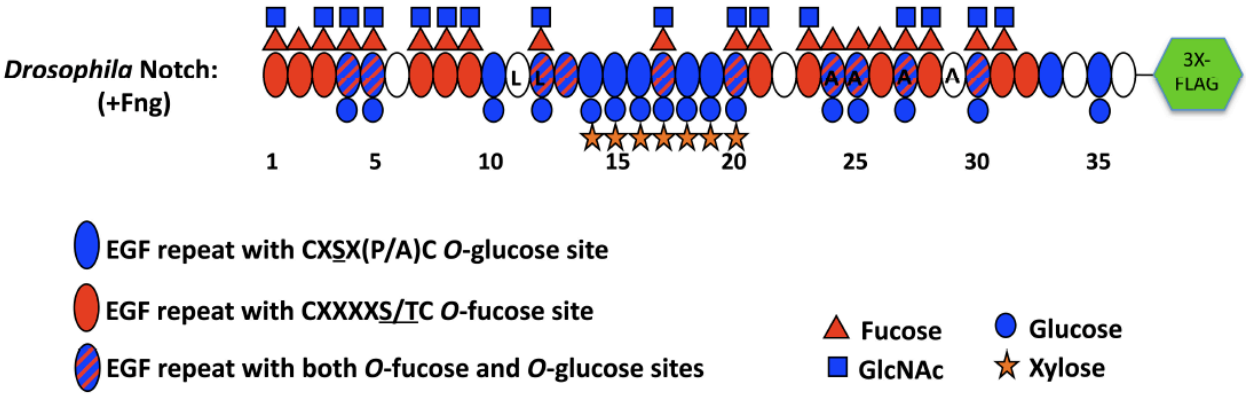


Figure 23. Summary of *O*-glycosylation site-mapping on N-EGF:FLAG.
Domain map of N-EGF:FLAG with observed *O*-fucose and *O*-glucose structures using nano-LC-ESI-MS/MS. This schematic was compiled using a composite of CNL searches, EIC searches, and MRM to confirm sites were modified with *O*-glycans and to denote the extent to which each site was elongated. Red triangles, fucose; blue squares, GlcNAc; blue circles, glucose; orange stars, xylose.

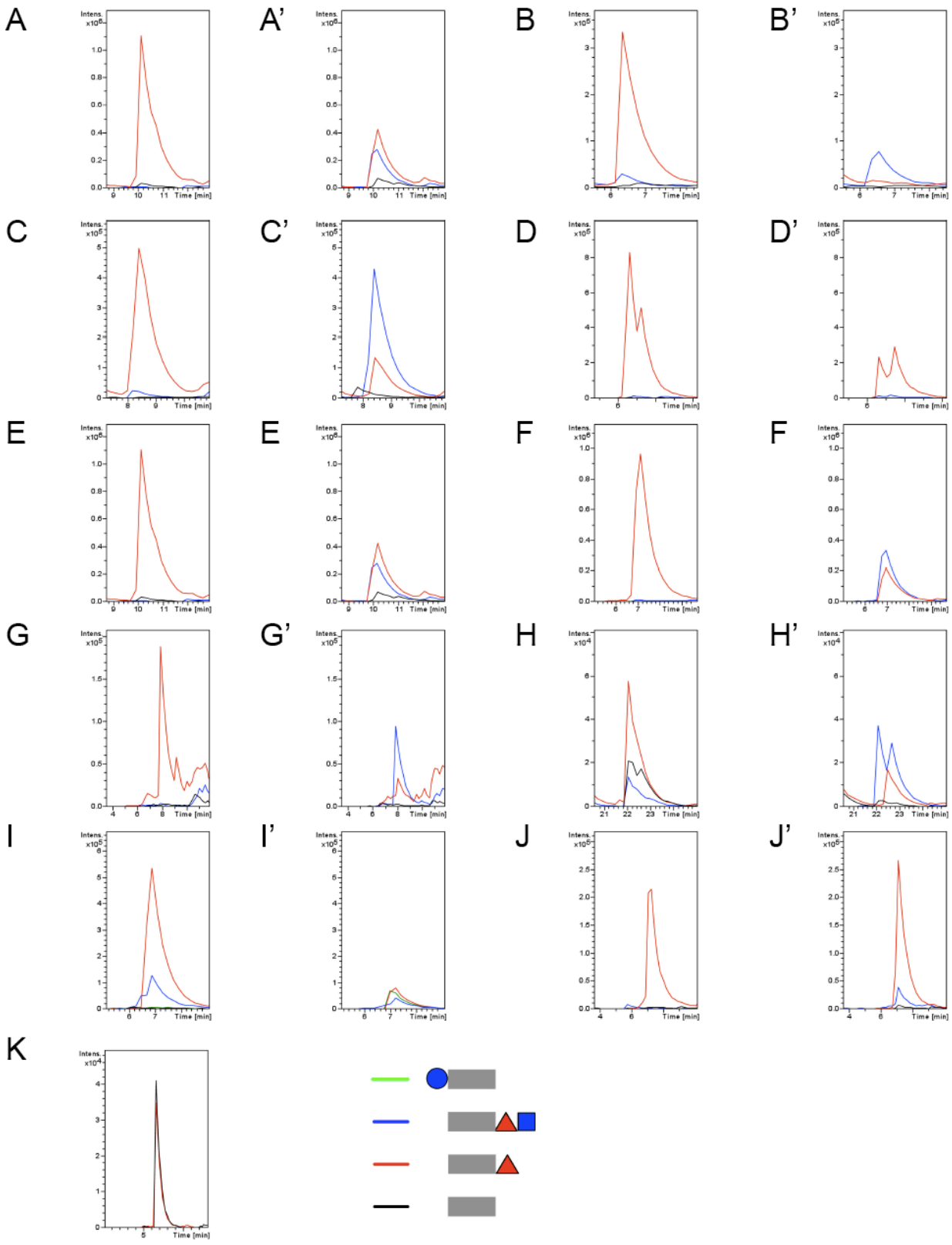

A
A'
B
B'
C
C'
D
D'
E
E
F
F
G
G'
H
H'
I
I'
J
J'
K
Intens.
Time [min]

Figure 24. *O*-fucosylation sites are elongated by Fringe to varying degrees.
Multiple Reaction Monitoring (MRM) analyses of *O*-fucosylated peptides from *Drosophila* Notch. A-J show N-EGF:FLAG, and A'-J' show an equivalent amount of N-EGF:FLAG + Fng. Black trace is naked peptide; red is peptide bearing *O*-fucose monosaccharide; blue is peptide with *O*-fucose disaccharide; green is peptide with *O*-glucose monosaccharide (only panels I, I'). K shows a representative internal standard control peptide from N-EGF:FLAG – (black) and + Fng (red) samples. **A, A'** (EGF 3); **B, B'** (EGF 5); **C, C'** (EGF 7); **D, D'** (EGF 8); **E, E'**(EGF 9); **F, F'** (EGF 21); **G, G'** (EGF 23); **H, H'** (EGF 28); **I, I'** (EGF 30); **J, J'** (EGF 31).

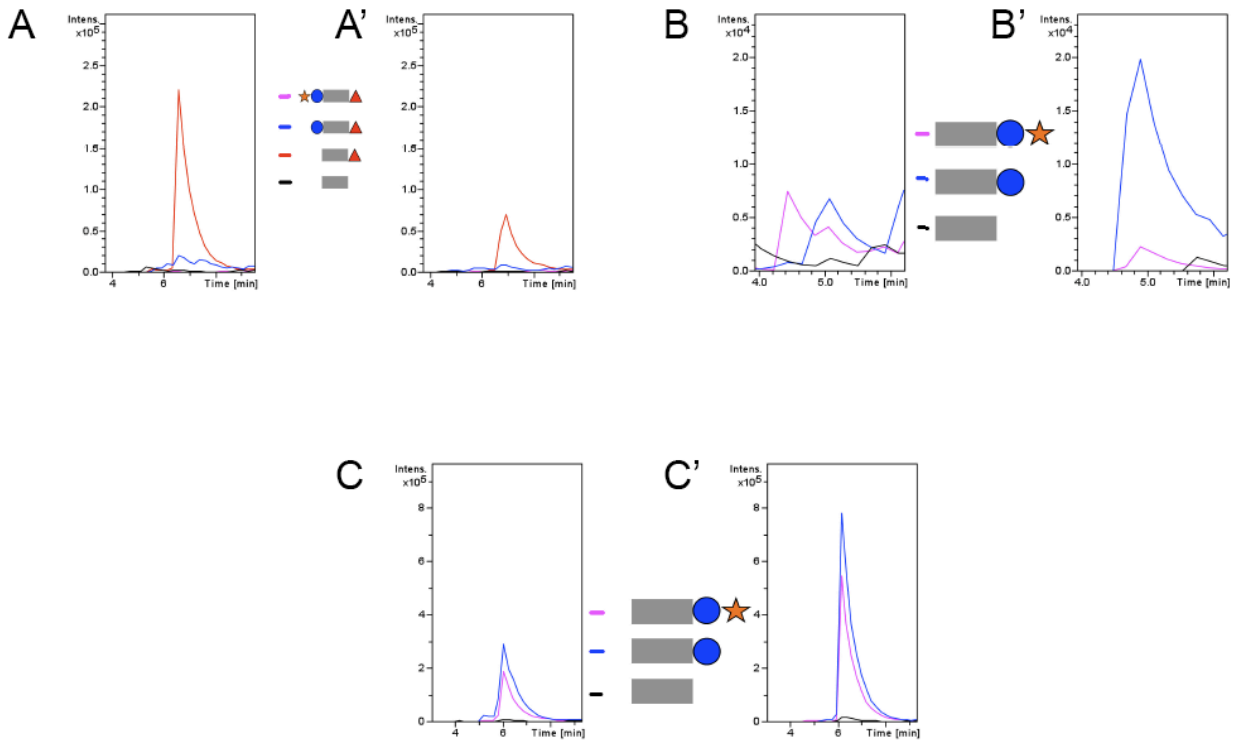


Figure 25. Elongation to disaccharide at *O*-glucose sites occurs substoichiometrically. Multiple Reaction Monitoring (MRM) analyses of *O*-glucosylated peptides from *Drosophila* Notch. A and A' show N-EGF:FLAG -/+Fng. Black trace is naked peptide; red is peptide with *O*-fucose monosaccharide; blue is peptide with *O*-fucose and *O*-glucose monosaccharides; pink trace is peptide with *O*-fucose monosaccharide and *O*-glucose disaccharide. B,C show N-EGF:FLAG, and B',C' show N-EGF:FLAG + Fng. Black trace is naked peptide; blue is peptide with *O*-glucose monosaccharide; pink is peptide with *O*-glucose disaccharide. **A, A'** (EGF 4); **B, B'** (EGF 5); **C, C'** (EGF 16).

Figure 26. **Fringe shows preference for some *O*-fucose sites over others.**
A. Sequence alignment of highly (>50%) and less highly (<50%) Fringe-elongated *O*-fucose sites, based upon decreases in relative levels of monosaccharide observed during MRM between – and + Fng samples, using MAFFT with CLUSTALW output. (*) denotes six conserved Cysteines. (*) indicates completely conserved residues. (**:**) indicates highly conserved residues. (**.**) indicates remotely conserved residues. **B.** Summary of conservation, where positions of interest are colored red (acidic) and blue (neutral/non-polar), relative to the six conserved Cysteines of EGF repeats.

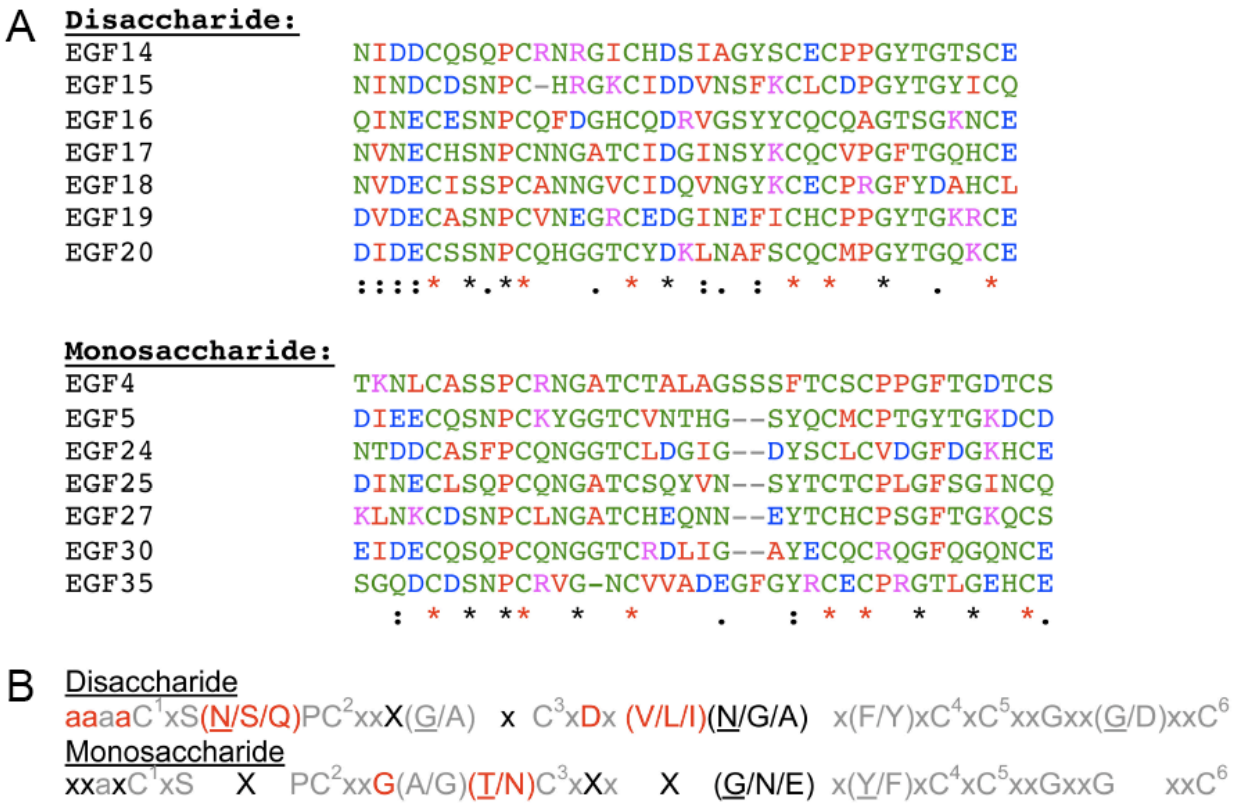
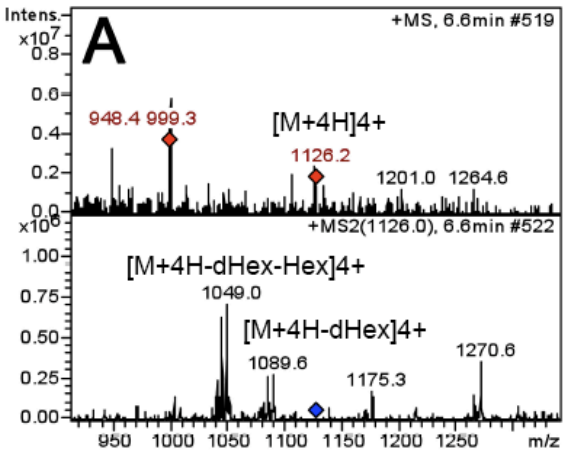


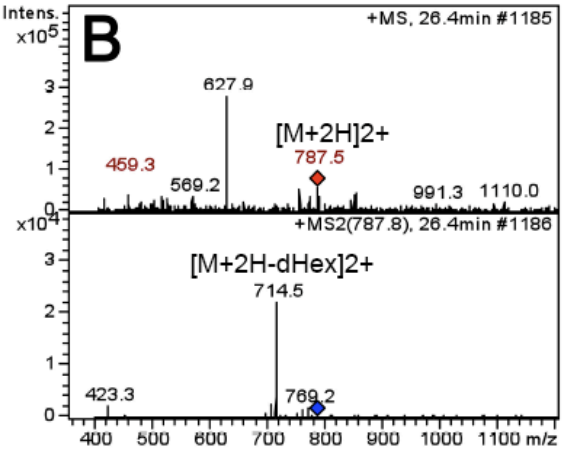
Figure 27. ***O*-glucose sites are elongated at substoichiometrically.**
A. Sequence alignment of *O*-glucose sites elongated by GXYLT, based upon detection of *O*-glc-xyl disaccharide by CNL and EIC searches, using MAFFT in CLUSTALW alignment format. (*) denotes six conserved Cysteines. (*) indicates completely conserved residues. (**:**) indicates moderately conserved residues. (**.**) indicates remotely conserved residues. **B.** Summary of conservation is seen in sequences below, where amino acids of interest are highlighted in red (corresponding residues in the opposite group are highlighted in black. Grey indicates common residues. a=acidic, underline=highest frequency of appearance).

Figure 28. Identification of *O*-fucosylated and *O*-glucosylated peptides from multiple EGF repeats of *Drosophila* Notch by LC-MS/MS. *O*-Glycosylated peptides were identified using the methods described in Figures 17 and 21. For each peptide, an MS spectrum showing the selection of the parent ion for fragmentation (top) and an MS/MS spectrum showing the resulting CID fragmentation (bottom) is shown. Predicted masses of singly charged ions are shown in Table 5. Ions in the MS/MS spectrum showing losses of the modifications are indicated, and EGF repeat from which the peptide is derived is labeled above each MS spectrum.

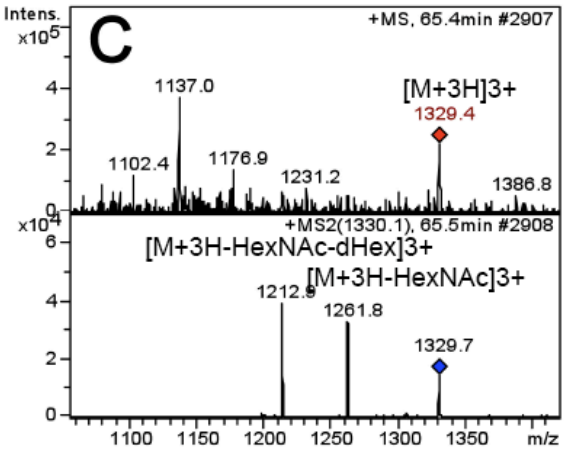
EGF 1



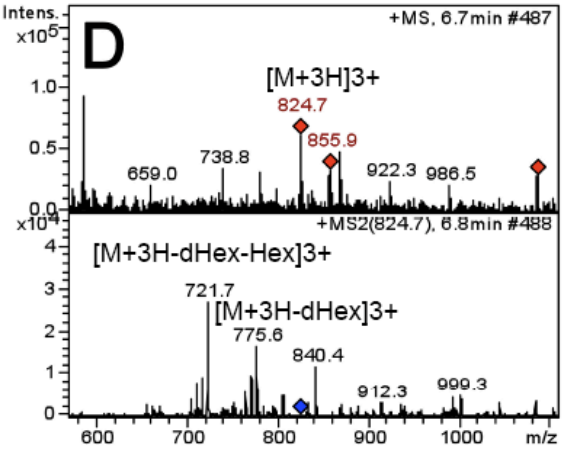
EGF 2



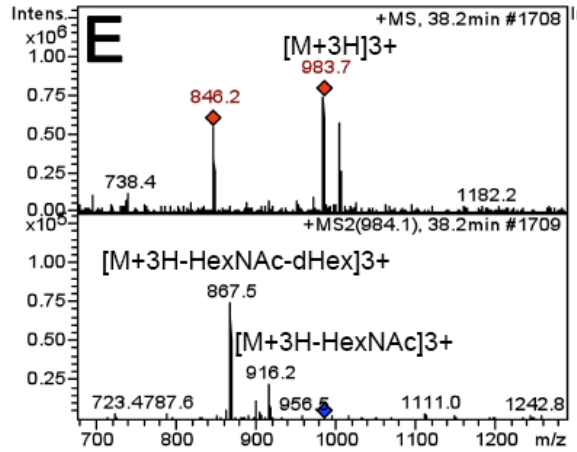
EGF 3



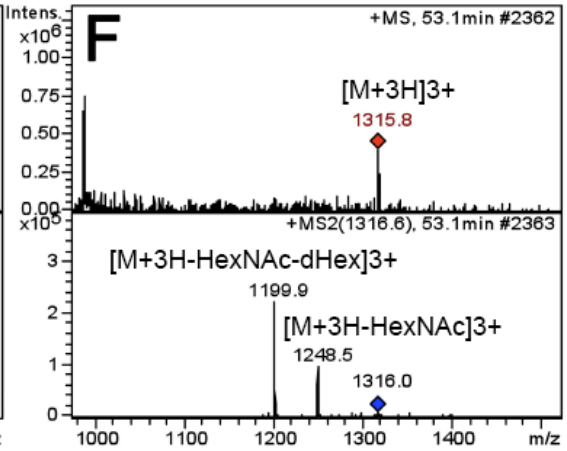
EGF 4



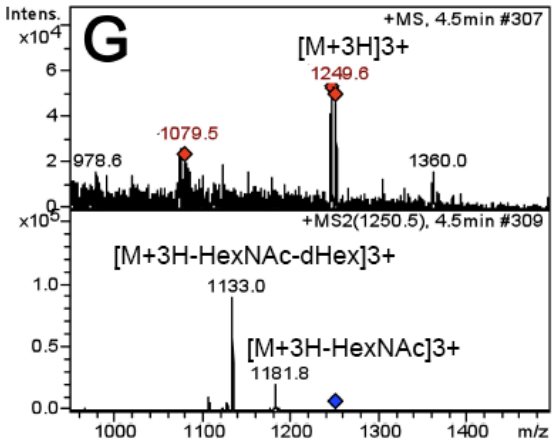
EGF 5



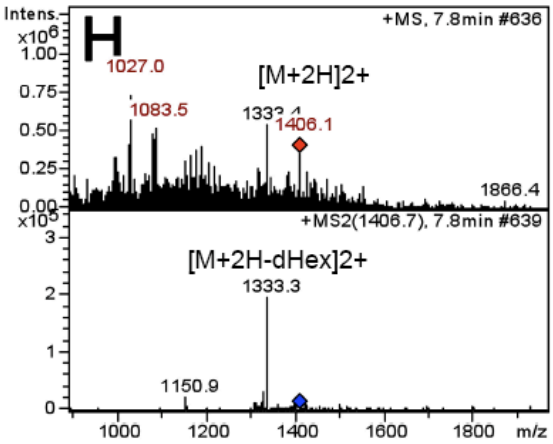
EGF 7



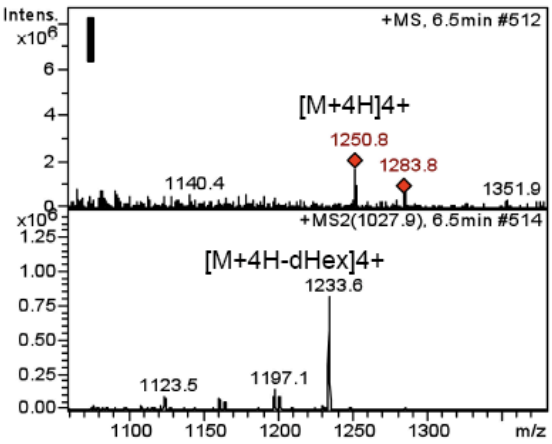
EGF 8



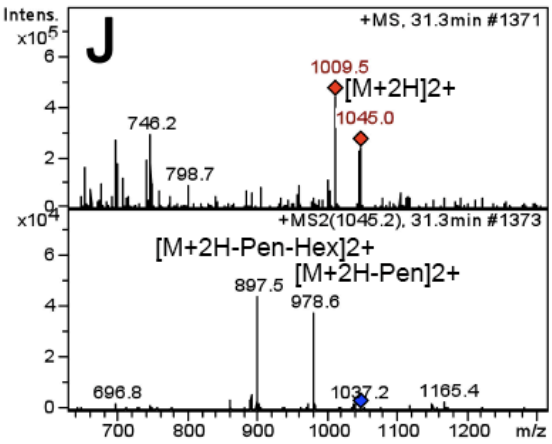
EGF 9



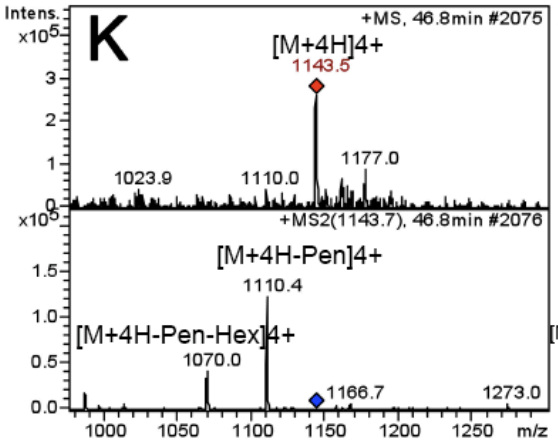
EGF 10



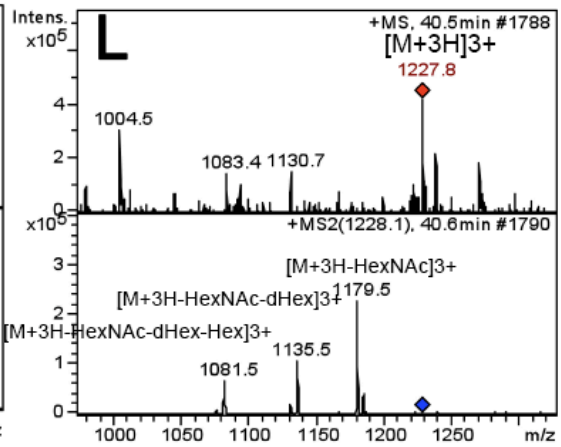
EGF 14



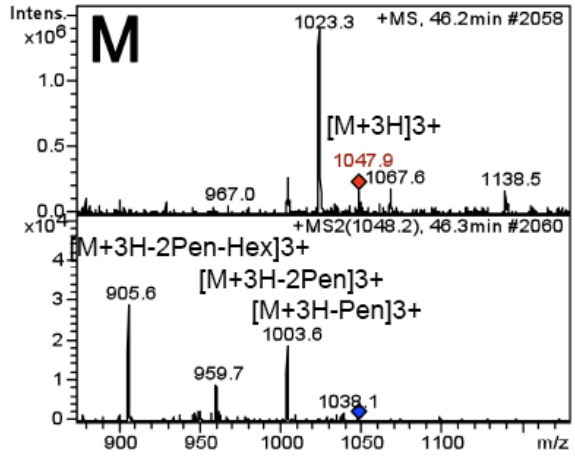
EGF 15



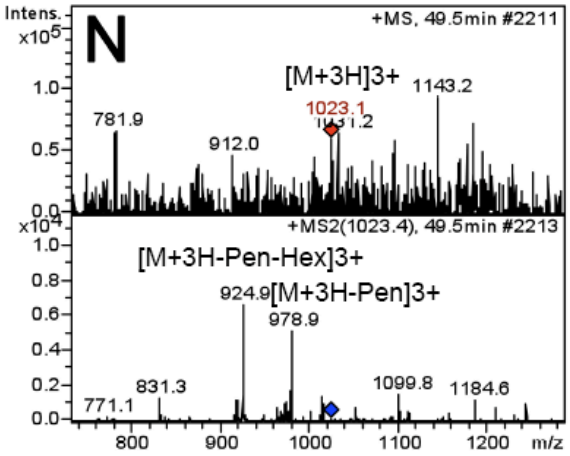
EGF 17



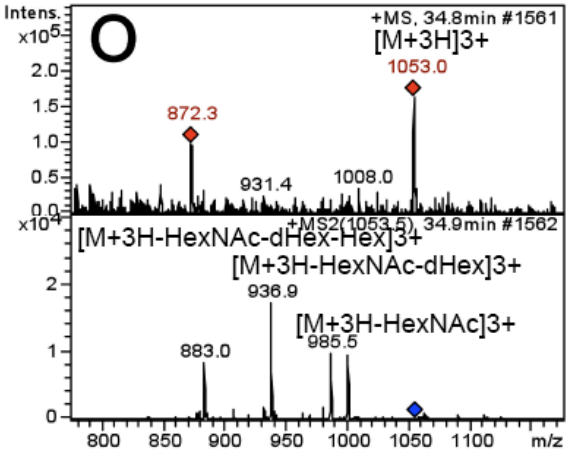
EGF 18



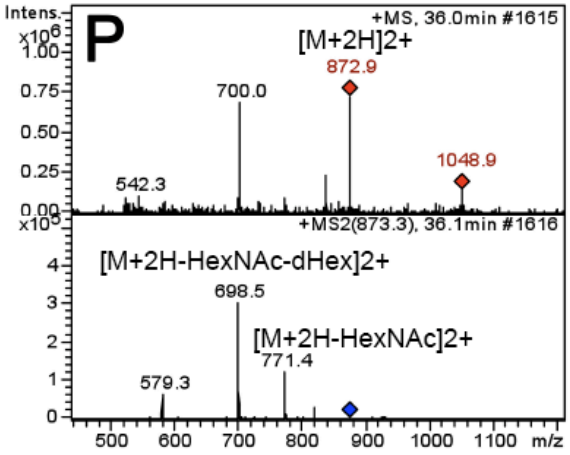
EGF 19



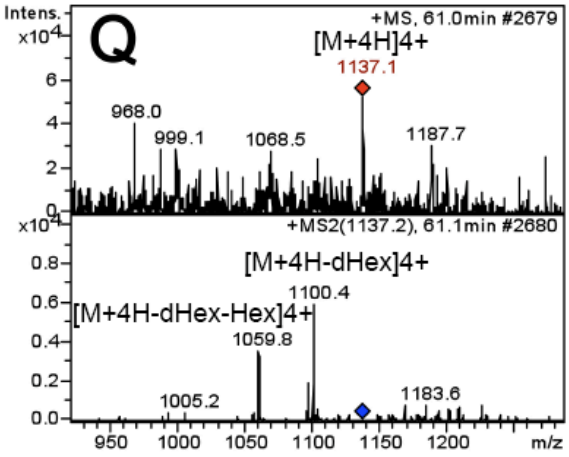
EGF 20



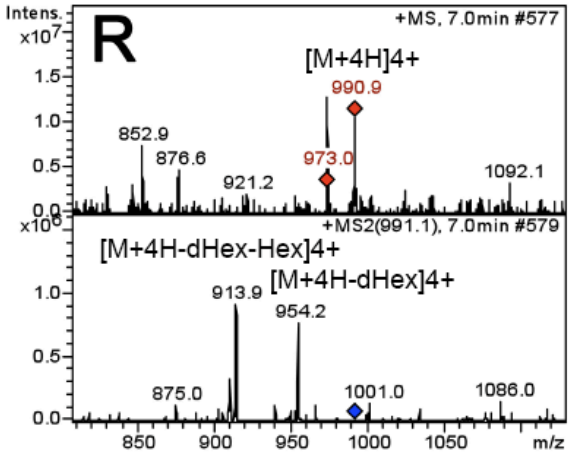
EGF 23



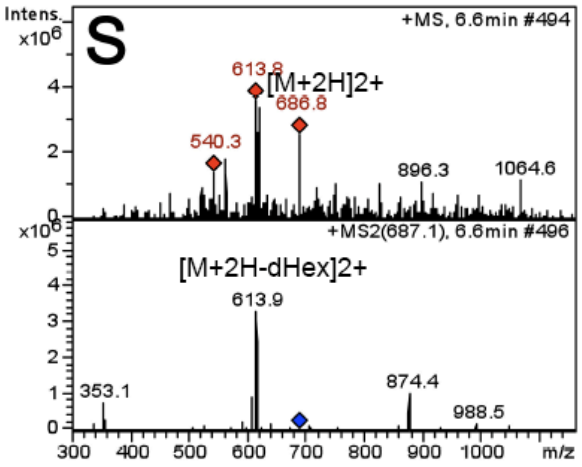
EGF 24



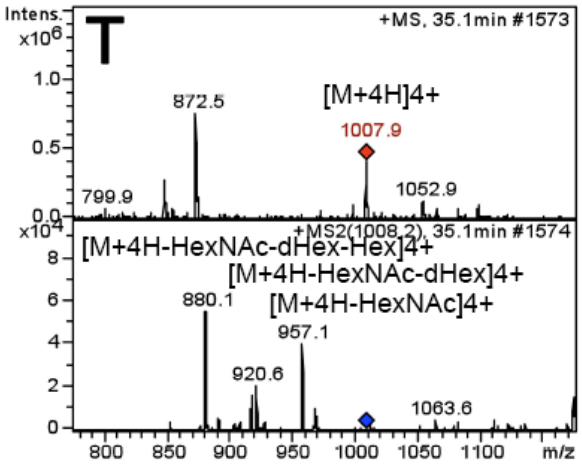
EGF 25



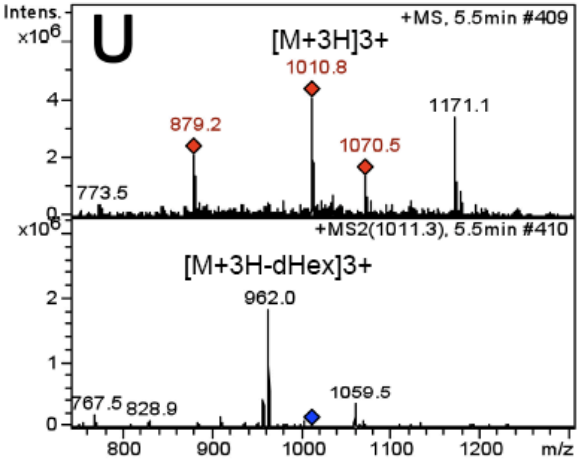
EGF 26



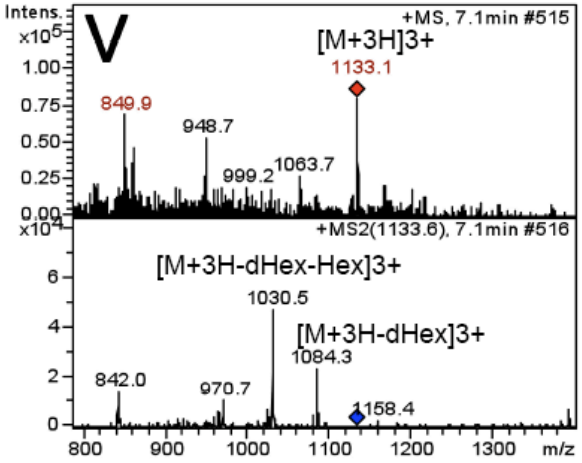
EGF 27



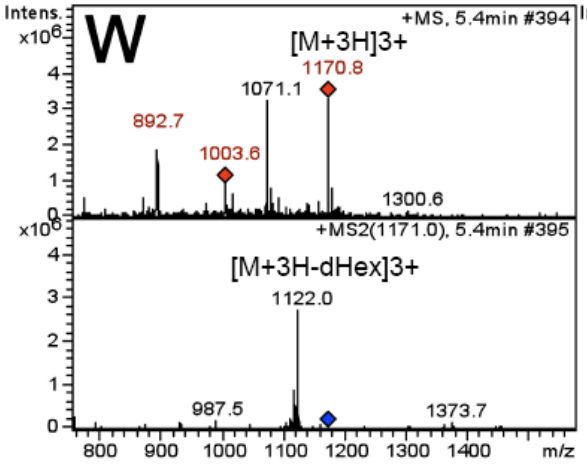
EGF 28



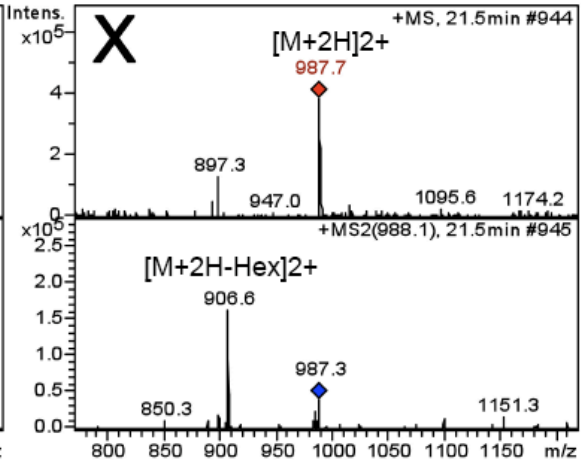
EGF 30



EGF 31



EGF 35



Chapter IV: Conclusions

The vast majority of EGF repeats on Notch contain the consensus sequence for *O*-glucosylation or *O*-fucosylation, and sometimes both within a single EGF repeat. The latest site-mapping work has focused on the occupancy of the predicted sites of *O*-glycosylation on mouse Notch1 and *Drosophila* Notch. Using the consensus sequence C^1xSxPC^2 as a prediction tool, a total of 16 EGF repeats are predicted to be *O*-glucosylated on mouse Notch1. Building on mass spectrometry strategies used previously to map selected sites of *O*-glycosylation, similar approaches were used to confirm all 16 predicted sites were modified. The efficiency of this modification was determined to be cell type-dependent though, as MS searches for monosaccharide compared to unmodified peptide for some sites revealed relative underglucosylation. An example of this can be seen in the partial glucosylation of EGF 27. Xylosylation levels at all sites did not appear to be affected, as *O*-glucose was efficiently elongated to the trisaccharide Xyl-Xyl-Glc. While performing MS searches for *O*-glucosylated peptides, a 17th glycopeptide was discovered that did not correspond to any of the masses of peptides predicted to be *O*-glucosylated based on the consensus sequence C^1xSxPC^2. When searches were broadened to include masses of "non-glycopeptides" produced by chymotryptic digest of mouse Notch1, the glycopeptide was identified as including the C1-C2 region of EGF 9, a novel *O*-glucosylated site that is efficiently elongated to trisaccharide. Identification of this non-traditional consensus site at EGF 9 led to the broadening of the consensus sequence for *O*-glucosylation to allow Alanine N-terminal to Cysteine 2: $C^1xSx(A/P)C^2$. A comprehensive summary of this site-mapping can be seen in Figure 12.

Similar studies were conducted using *Drosophila* Notch over-expressed in S2 cells in the absence or presence of exogenous Fringe. Site-mapping confirmed that this receptor is also *O*-fucosylated and *O*-glucosylated at high stoichiometries, and consistent with previous observations, these modifications do not appear to be elongated beyond the disaccharide in S2 cells. A summation of these results is seen in Figure 23, which is based on a composite of MS/MS data, EIC searches to detect potentially elongated forms of glycans at each site, and sensitive MRM studies that allowed detection of even low levels of elongation at assorted EGF repeats. We further examined sites of Fringe using MRM mass spectrometry, a semi-quantitative technique, and compared relative levels of Fringe elongation at specific sites to determine if there is a hierarchy for preferential elongation among *O*-fucose sites. Fringe elongation to the GlcNAc-β1,3-Fucose disaccharide was observed at sub-stoichiometric levels at numerous sites, indicating Fringe may show a preference for elongation depending on context. The utility of MRM for these studies lay in the provision of relative quantitation of *O*-glycan structures both within a single site (as was the case of – and + Fng samples) and between sites, allowing for comparative studies of relative site-to-site elongation. Sequence alignment analyses comparing groups of sites described as highly-Fringe modified (>50% decrease in monosaccharide signal), less-Fringe-modified (<50% decrease in monosaccharide signal), and Fringe-unmodified revealed two residues of interest that may be acting as specificity determinants for Fringe modification. An acidic residue in position 18 and a non-polar residue in position 20 of EGF repeats that are highly Fng elongated, are less highly conserved in EGF repeats modified by Fng to a lesser degree. In the future, experiments involving site-directed mutagenesis of singly-expressed EGF repeats to substitute neutral amino acids at these positions would be interesting tools for *in vitro* Fng assays to elucidate their contributions as specificity

determinants in substrate recognition in *Drosophila*. Feasibility of these assays will of course be dependent upon the ability to express properly folded individual EGF repeat mutants, as mis-folding will prevent requisite *O*-fucosylation.

O-Glucosylation of EGF repeats in *Drosophila* Notch was also observed to occur stoichiometrically, with the elongated disaccharide modifications appearing to cluster to the middle region of the ECD, unlike the more punctate pattern of Fringe elongation observed on *O*-fucose sites. Interestingly, this is in stark contrast to *O*-glucose elongation in mammalian systems, where we observe elongation at high stoichiometries at all *O*-glucosylated sites. On mouse Notch1, we observe *O*-glucosylation at high stoichiometries across the ECD at all predicted sites, and elongation to trisaccharide is also found at high stoichiometries (Figure 12). *Drosophila* Notch also shares in *O*-glucosylation at high stoichiometries, but in contrast to the mammalian system, we see substoichiometric xylosylation cluster to the middle region of the receptor, an area rich in cbEGF repeats. In this case, multiple residues appear to be involved in distinguishing *O*-glucosylated EGF repeat substrates for xylosylation, and may be related to specificity of the *Drosophila* GXYLT homolog to a subset of sites recognized by either GXYLT1 or GXYLT2. To determine whether this is the case, future experiments could include expression of individual EGF repeats of mouse Notch1 that are not predicted to be xylosylated by *Drosophila* GXYLT, and use them in *in vitro* assays to test their ability to act as substrates for GXYLT1 and GXYLT2.

Several groups have proposed models of how *O*-glycans may be affecting the global conformation state of Notch. The most recent models incorporate the flexibility imparted by non-cbEGFs, as well as the rigidity imposed by cbEGFs. Xu *et al.* proposed a "triple-strand" conformation of the receptor, with flexible regions at and around EGF repeats 10 and 28, which

are highly conserved non-cbEGFs (Figure 29) (56). In this model, the flexible non-cbEGFs would allow a folding over onto themselves, leading to packing against each other and forming an almost pretzel-like shape. Upon ligand binding, a strand "displacement" is proposed to cause a conformational change. A more simplified version of this model was proposed by Rampal *et al.*, based on conserved flexible "hinge"-like regions, and defined by the conserved non-cbEGFs (82). These hinge regions would then undergo conformational changes imposed by *O*-glycosylation at EGF repeats in these regions, leading to extension of the ECD into a more rod-like linear structure. Support for this concept can be found in mouse Notch1 cell-based signaling studies where individual *O*-fucose site mutations at EGF repeats 26 and 27, and *O*-glucose site mutation at EGF repeat 28 (Chapter II, Figure 11), all located in putative flexible regions of the ECD, showed significant effects on Notch activation in the absence of *O*-glycosylation at these sties. It is possible that glycosylation in these regions alters the flexibility in the ECD in these regions, allowing for appropriate interaction with ligands (Figure 29C). Based on EM studies by Kelly *et al*., it appears that this model may prove to be a highly accurate depiction of Notch structure/conformation.

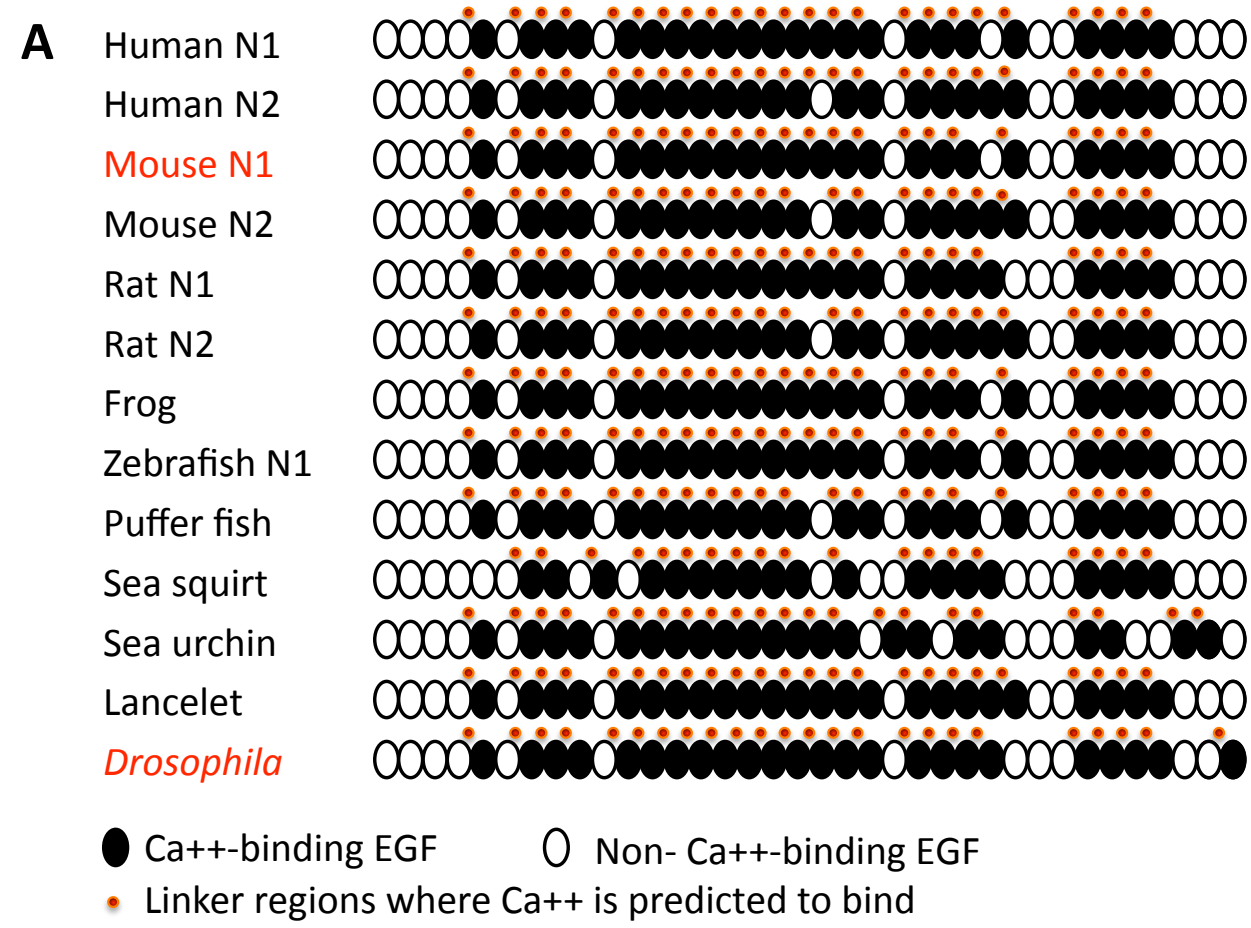

A
Human N1
Human N2
Mouse N1
Mouse N2
Rat N1
Rat N2
Frog
Zebrafish N1
Puffer fish
Sea squirt
Sea urchin
Lancelet
Drosophila
Ca++-binding EGF
Non- Ca++-binding EGF
Linker regions where Ca++ is predicted to bind
B

Figure 29. Notch conformation: flexibility models.
A. Notch is comprised of cbEGFs (black ovals) and non-cbEGFs (white ovals) that are highly conserved across species. Coordination of calcium is predicted to impart structural rigidity based on NMR studies of EGF repeats 11-13 of human Notch1. Red spheres denote Ca++. Using structural information from this fragment, Hambleton and co-workers proposed flexible and rod-like regions exist within the Notch ECD. Adapted from (56). **B.** Mouse Notch1 ECD is heavily *O*-glycosylated, including in predicted flexible regions that are indicated with arrow-flexibility spring (Chapter II, Figure 12).

References:

1. Haltiwanger, R. S., and Lowe, J. B. (2004) *Annu Rev Biochem* **73**, 491-537
2. Rampal, R., Luther, K. B., and Haltiwanger, R. S. (2007) *Current molecular medicine* **7**, 427-445
3. Kornfeld, R., and Kornfeld, S. (1985) *Annu Rev Biochem* **54**, 631-664
4. Herscovics, A. (1999) *Biochimica et biophysica acta* **1473**, 96-107
5. Martin, P. T. (2007) *Current molecular medicine* **7**, 417-425
6. Shao, L., and Haltiwanger, R. S. (2003) *Cell Mol Life Sci* **60**, 241-250
7. Jafar-Nejad, H., Leonardi, J., and Fernandez-Valdivia, R. (2010) *Glycobiology* **20**, 931-949
8. Acar, M., Jafar-Nejad, H., Takeuchi, H., Rajan, A., Ibrani, D., Rana, N. A., Pan, H., Haltiwanger, R. S., and Bellen, H. J. (2008) *Cell* **132**, 247-258
9. Fernandez-Valdivia, R., Takeuchi, H., Smarghandi, A., Lopez, M., Leonardi, J., Haltiwanger, R. S., and Jafar-Nejad, H. (In Press) *Development (Cambridge, England)*
10. Takeuchi, H., and Haltiwanger, R. S. (2010) *Semin Cell Dev Biol* **21**, 638-645
11. Hase, S., Kawabata, S., Nishimura, H., Takeya, H., Sueyoshi, T., Miyata, T., Iwanaga, S., Takao, T., Shimonishi, Y., and Ikenaka, T. (1988) *Journal of biochemistry* **104**, 867-868
12. Hase, S., Nishimura, H., Kawabata, S., Iwanaga, S., and Ikenaka, T. (1990) *The Journal of biological chemistry* **265**, 1858-1861
13. Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., and Iwanaga, S. (1989) *The Journal of biological chemistry* **264**, 20320-20325
14. Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., and Iwanaga, S. (1992) *The Journal of biological chemistry* **267**, 17520-17525
15. Krogh, T. N., Bachmann, E., Teisner, B., Skjodt, K., and Hojrup, P. (1997) *Eur J Biochem* **244**, 334-342
16. Harris, R. J., and Spellman, M. W. (1993) *Glycobiology* **3**, 219-224
17. Matsuura, A., Ito, M., Sakaidani, Y., Kondo, T., Murakami, K., Furukawa, K., Nadano, D., Matsuda, T., and Okajima, T. (2008) *The Journal of biological chemistry* **283**, 35486-35495
18. Moloney, D. J., Shair, L. H., Lu, F. M., Xia, J., Locke, R., Matta, K. L., and Haltiwanger, R. S. (2000) *The Journal of biological chemistry* **275**, 9604-9611
19. Shao, L., Moloney, D. J., and Haltiwanger, R. (2003) *The Journal of biological chemistry* **278**, 7775-7782
20. Aoki, K., Porterfield, M., Lee, S. S., Dong, B., Nguyen, K., McGlamry, K. H., and Tiemeyer, M. (2008) *The Journal of biological chemistry* **283**, 30385-30400
21. Whitworth, G. E., Zandberg, W. F., Clark, T., and Vocadlo, D. J. (2010) *Glycobiology* **20**, 287-299
22. Bakker, H., Oka, T., Ashikov, A., Yadav, A., Berger, M., Rana, N. A., Bai, X., Jigami, Y., Haltiwanger, R. S., Esko, J. D., and Gerardy-Schahn, R. (2009) *The Journal of biological chemistry* **284**, 2576-2583
23. Wang, Y., Shao, L., Shi, S., Harris, R. J., Spellman, M. W., Stanley, P., and Haltiwanger, R. S. (2001) *The Journal of biological chemistry* **276**, 40338-40345

24. Shi, S., and Stanley, P. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5234-5239
25. Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K. D., Haltiwanger, R. S., and Vogt, T. F. (2000) *Nature* **406**, 369-375
26. Sethi, M. K., Buettner, F. F., Krylov, V. B., Takeuchi, H., Nifantiev, N. E., Haltiwanger, R. S., Gerardy-Schahn, R., and Bakker, H. (2010) *The Journal of biological chemistry* **285**, 1582-1586
27. Sakaidani, Y., Furukawa, K., and Okajima, T. (2010) *Methods Enzymol* **480**, 355-373
28. Mumm, J. S., and Kopan, R. (2000) *Developmental biology* **228**, 151-165
29. Artavanis-Tsakonas, S., Rand, M. D., and Lake, R. J. (1999) *Science (New York, N.Y* **284**, 770-776
30. Morgan, T. H. (1917) *The American Naturalist* **51**, 513-544
31. Swiatek, P. J., Lindsell, C. E., del Amo, F. F., Weinmaster, G., and Gridley, T. (1994) *Genes & development* **8**, 707-719
32. Hamada, Y., Kadokawa, Y., Okabe, M., Ikawa, M., Coleman, J. R., and Tsujimoto, Y. (1999) *Development (Cambridge, England)* **126**, 3415-3424
33. Weng, A. P., Ferrando, A. A., Lee, W., Morris, J. P. t., Silverman, L. B., Sanchez-Irizarry, C., Blacklow, S. C., Look, A. T., and Aster, J. C. (2004) *Science (New York, N.Y* **306**, 269-271
34. Joutel, A., and Tournier-Lasserve, E. (1998) *Semin Cell Dev Biol* **9**, 619-625
35. McDaniell, R., Warthen, D. M., Sanchez-Lara, P. A., Pai, A., Krantz, I. D., Piccoli, D. A., and Spinner, N. B. (2006) *American journal of human genetics* **79**, 169-173
36. Kusumi, K., Dunwoodie, S. L., and Krumlauf, R. (2001) *Mechanisms of development* **100**, 141-144
37. Zhang, Y., Argaw, A. T., Gurfein, B. T., Zameer, A., Snyder, B. J., Ge, C., Lu, Q. R., Rowitch, D. H., Raine, C. S., Brosnan, C. F., and John, G. R. (2009) *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19162-19167
38. Mazzone, M., Selfors, L. M., Albeck, J., Overholtzer, M., Sale, S., Carroll, D. L., Pandya, D., Lu, Y., Mills, G. B., Aster, J. C., Artavanis-Tsakonas, S., and Brugge, J. S. (2010) *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5012-5017
39. Aster, J. C., Pear, W. S., and Blacklow, S. C. (2008) *Annual review of pathology* **3**, 587-613
40. Kopan, R., and Ilagan, M. X. (2009) *Cell* **137**, 216-233
41. Okajima, T., and Irvine, K. D. (2002) *Cell* **111**, 893-904
42. Sasamura, T., Sasaki, N., Miyashita, F., Nakao, S., Ishikawa, H. O., Ito, M., Kitagawa, M., Harigaya, K., Spana, E., Bilder, D., Perrimon, N., and Matsuno, K. (2003) *Development (Cambridge, England)* **130**, 4785-4795
43. Stahl, M., Uemura, K., Ge, C., Shi, S., Tashima, Y., and Stanley, P. (2008) *The Journal of biological chemistry* **283**, 13638-13651
44. Zhou, L., Li, L. W., Yan, Q., Petryniak, B., Man, Y., Su, C., Shim, J., Chervin, S., and Lowe, J. B. (2008) *Blood* **112**, 308-319
45. Yao, D., Huang, Y., Huang, X., Wang, W., Yan, Q., Wei, L., Xin, W., Gerson, S., Stanley, P., Lowe, J. B., and Zhou, L. (2011) *Blood*
46. Zhang, N., and Gridley, T. (1998) *Nature* **394**, 374-377
47. Zhang, N., Norton, C. R., and Gridley, T. (2002) *Genesis* **33**, 21-28

48. Visan, I., Yuan, J. S., Liu, Y., Stanley, P., and Guidos, C. J. (2010) *J Immunol* **185**, 4609-4617
49. Stanley, P., and Guidos, C. J. (2009) *Immunological reviews* **230**, 201-215
50. Wilson, A., MacDonald, H. R., and Radtke, F. (2001) *The Journal of experimental medicine* **194**, 1003-1012
51. Koch, U., Fiorini, E., Benedito, R., Besseyrias, V., Schuster-Gossler, K., Pierres, M., Manley, N. R., Duarte, A., Macdonald, H. R., and Radtke, F. (2008) *The Journal of experimental medicine* **205**, 2515-2523
52. Tan, J. B., Xu, K., Cretegny, K., Visan, I., Yuan, J. S., Egan, S. E., and Guidos, C. J. (2009) *Immunity* **30**, 254-263
53. Shao, L., Luo, Y., Moloney, D. J., and Haltiwanger, R. (2002) *Glycobiology* **12**, 763-770
54. Takeuchi, H., Fernandez-Valdivia, R., Caswell, D. S., Nita-Lazar, A., Kakuda, S., Rana, N. A., Macnaughtan, M., Jafar-Nejad, H., and Haltiwanger, R. S. (In Preparation)
55. Sethi, M. K., Buettner, F. F., Krylov, V. B., Takeuchi, H., Nifantiev, N. E., Haltiwanger, R. S., Gerardy-Schahn, R., and Bakker, H. *The Journal of biological chemistry* **285**, 1582-1586
56. Xu, A., Lei, L., and Irvine, K. D. (2005) *The Journal of biological chemistry* **280**, 30158-30165
57. Ge, C., and Stanley, P. (2008) *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1539-1544
58. Okajima, T., Xu, A., Lei, L., and Irvine, K. D. (2005) *Science (New York, N.Y* **307**, 1599-1603
59. Okajima, T., Reddy, B., Matsuda, T., and Irvine, K. D. (2008) *BMC biology* **6**, 1
60. Bruckner, K., Perez, L., Clausen, H., and Cohen, S. (2000) *Nature* **406**, 411-415
61. Okajima, T., Xu, A., and Irvine, K. D. (2003) *The Journal of biological chemistry* **278**, 42340-42345
62. Xu, A., Haines, N., Dlugosz, M., Rana, N. A., Takeuchi, H., Haltiwanger, R. S., and Irvine, K. D. (2007) *J Biol Chem (accepted)*
63. Hicks, C., Johnston, S. H., diSibio, G., Collazo, A., Vogt, T. F., and Weinmaster, G. (2000) *Nature cell biology* **2**, 515-520
64. Shimizu, K., Chiba, S., Saito, T., Kumano, K., Takahashi, T., and Hirai, H. (2001) *The Journal of biological chemistry* **276**, 25753-25758
65. Chen, J., Moloney, D. J., and Stanley, P. (2001) *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13716-13721
66. Chen, J., Lu, L., Shi, S., and Stanley, P. (2006) *Gene Expr Patterns* **6**, 376-382
67. Arnett, K. L., Hass, M., McArthur, D. G., Ilagan, M. X., Aster, J. C., Kopan, R., and Blacklow, S. C. *Nature structural & molecular biology* **17**, 1312-1317
68. Aste-Amezaga, M., Zhang, N., Lineberger, J. E., Arnold, B. A., Toner, T. J., Gu, M., Huang, L., Vitelli, S., Vo, K. T., Haytko, P., Zhao, J. Z., Baleydier, F., L'Heureux, S., Wang, H., Gordon, W. R., Thoryk, E., Andrawes, M. B., Tiyanont, K., Stegmaier, K., Roti, G., Ross, K. N., Franlin, L. L., Wang, H., Wang, F., Chastain, M., Bett, A. J., Audoly, L. P., Aster, J. C., Blacklow, S. C., and Huber, H. E. *PloS one* **5**, e9094
69. Gordon, W. R., Roy, M., Vardar-Ulu, D., Garfinkel, M., Mansour, M. R., Aster, J. C., and Blacklow, S. C. (2009) *Blood* **113**, 4381-4390
70. Sanchez-Irizarry, C., Carpenter, A. C., Weng, A. P., Pear, W. S., Aster, J. C., and Blacklow, S. C. (2004) *Molecular and cellular biology* **24**, 9265-9273

71. Rebay, I., Fleming, R. J., Fehon, R. G., Cherbas, L., Cherbas, P., and Artavanis-Tsakonas, S. (1991) *Cell* **67**, 687-699
72. Cordle, J., Johnson, S., Tay, J. Z., Roversi, P., Wilkin, M. B., de Madrid, B. H., Shimizu, H., Jensen, S., Whiteman, P., Jin, B., Redfield, C., Baron, M., Lea, S. M., and Handford, P. A. (2008) *Nature structural & molecular biology* **15**, 849-857
73. Hambleton, S., Valeyev, N. V., Muranyi, A., Knott, V., Werner, J. M., McMichael, A. J., Handford, P. A., and Downing, A. K. (2004) *Structure* **12**, 2173-2183
74. Artavanis-Tsakonas, S., Rand, M. D., and Lake, R. J. (1999) *Science (New York, N.Y* **284**, 770-776
75. Campbell, I. D., and Bork, P. (1993) *Curr.Opin.Struct.Biol.* **3**, 385-392
76. Moloney, D. J., Shair, L., Lu, F. M., Xia, J., Locke, R., Matta, K. L., and Haltiwanger, R. S. (2000) *J.Biol.Chem.* **275**, 9604-9611
77. Shao, L., Moloney, D. J., and Haltiwanger, R. S. (2003) *The Journal of biological chemistry* **278**, 7775-7782
78. Wang, Y., and Spellman, M. W. (1998) *J.Biol.Chem.* **273**, 8112-8118
79. Wang, Y., Shao, L., Shi, S., Harris, R. J., Spellman, M. W., Stanley, P., and Haltiwanger, R. S. (2001) *J.Biol.Chem.* **276**, 40338-40345
80. Luo, Y., and Haltiwanger, R. S. (2005) *The Journal of biological chemistry* **280**, 11289-11294
81. Shi, S., and Stanley, P. (2003) *Proc.Natl.Acad.Sci.USA* **100**, 5234-5239
82. Rampal, R., Arboleda-Velasquez, J. F., Nita-Lazar, A., Kosik, K. S., and Haltiwanger, R. S. (2005) *The Journal of biological chemistry* **280**, 32133-32140
83. Shi, S., Ge, C., Luo, Y., Hou, X., Haltiwanger, R. S., and Stanley, P. (2007) *The Journal of biological chemistry* **282**, 20133-20141
84. Hase, S., Kawabata, S., Nishimura, H., Takeya, H., Sueyoshi, T., Miyata, T., Iwanaga, S., Takao, T., Shimonishi, Y., and Ikenaka, T. (1988) *J.Biochem.(Tokyo)* **104**, 867-868
85. Hase, S., Nishimura, H., Kawabata, S., Iwanaga, S., and Ikenaka, T. (1990) *J.Biol.Chem.* **265**, 1858-1861
86. Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., and Iwanaga, S. (1989) *J.Biol.Chem.* **264**, 20320-20325
87. Nishimura, H., Yamashita, S., Zeng, Z., Walz, D. A., and Iwanaga, S. (1992) *J.Biochem.(Tokyo)* **111**, 460-464
88. Krogh, T. N., Bachmann, E., Teisner, B., Skjodt, K., and Hojrup, P. (1997) *Eur.J.Biochem.* **244**, 334-342
89. Harris, R. J., and Spellman, M. W. (1993) *Glycobiology* **3**, 219-224
90. Stanley, P., and Siminovitch, L. (1977) *Somatic Cell Genetics* **3**, 391-405
91. Moloney, D. J., Lin, A. I., and Haltiwanger, R. S. (1997) *J.Biol.Chem.* **272**, 19046-19050
92. Nita-Lazar, A., and Haltiwanger, R. S. (2006) *Methods Enzymol* **417**, 93-111
93. Rampal, R., Li, A. S., Moloney, D. J., Georgiou, S. A., Luther, K. B., Nita-Lazar, A., and Haltiwanger, R. S. (2005) *The Journal of biological chemistry* **280**, 42454-42463
94. Frey, P. A. (1996) *FASEB J* **10**, 461-470
95. Kumar, R., and Stanley, P. (1989) *Molec.Cell.Biol.* **9**, 5713-5717
96. Xu, A., Haines, N., Dlugosz, M., Rana, N. A., Takeuchi, H., Haltiwanger, R. S., and Irvine, K. D. (2007) *The Journal of biological chemistry* **282**, 35153-35162
97. Ge, C., Liu, T., Hou, X., and Stanley, P. (2008) *BMC developmental biology* **8**, 48

98. de Celis, J. F., and Bray, S. J. (2000) *Development (Cambridge, England)* **127**, 1291-1302
99. van den Brandt, J., Kwon, S. H., Hunig, T., McPherson, K. G., and Reichardt, H. M. (2005) *J Immunol* **174**, 7845-7852
100. Haines, N., and Irvine, K. D. (2003) *Nat. Rev. Mol. Cell Biol.* **4**, 786-797
101. Rebay, I., Fleming, R. J., Fehon, R. G., Cherbas, L., Cherbas, P., and Artavanis-Tsakonas, S. (1991) *Cell* **67**, 687-699
102. Rampal, R., Arboleda-Velasquez, J., Nita-Lazar, A., Kosik, K. S., and Haltiwanger, R. S. (2005) *The Journal of biological chemistry* **280**, 32133-32140
103. Hiruma-Shimizu, K., Hosoguchi, K., Liu, Y., Fujitani, N., Ohta, T., Hinou, H., Matsushita, T., Shimizu, H., Feizi, T., and Nishimura, S. *Journal of the American Chemical Society* **132**, 14857-14865
104. Haltiwanger, R. S., Luo, Y., and Luther, K. B. (2007) O-Fucosylation of Glycoproteins. in *Comprehensive Glycoscience* (Kamerling, J. P. ed.), Elsevier. pp
105. Haines, N., and Irvine, K. D. (2003) *Nat Rev Mol Cell Biol* **4**, 786-797
106. Cho, K. O., and Choi, K. W. (1998) *Nature* **396**, 272-276
107. Dominguez, M., and de Celis, J. F. (1998) *Nature* **396**, 276-278
108. Papayannopoulos, V., Tomlinson, A., Panin, V. M., Rauskolb, C., and Irvine, K. D. (1998) *Science (New York, N.Y* **281**, 2031-2034
109. de Celis, J. F., Tyler, D. M., de Celis, J., and Bray, S. J. (1998) *Development (Cambridge, England)* **125**, 4617-4626
110. Rauskolb, C., and Irvine, K. D. (1999) *Developmental biology* **210**, 339-350
111. Grammont, M., and Irvine, K. D. (2001) *Development (Cambridge, England)* **128**, 2243-2253
112. Haines, N., and Irvine, K. D. (2005) *Glycobiology* **15**, 335-346
113. Unwin, R. D., Evans, C. A., and Whetton, A. D. (2006) *Trends in biochemical sciences* **31**, 473-484
114. Yocum, A. K., and Chinnaiyan, A. M. (2009) *Brief Funct Genomic Proteomic* **8**, 145-157
115. Han, B., and Higgs, R. E. (2008) *Brief Funct Genomic Proteomic* **7**, 340-354