Stony Brook University

The Versatile Applications of 6-Alkynyl Fucose and Other Fucose Analogs in Studying Protein Fucosylation

A Dissertation Presented

by

Esam Al-Shareffi

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

August 2014

Stony Brook University

The Graduate School

Esam Al-Shareffi

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. Isaac Carrico – Chairperson of the Defense
Associate Professor, Department of Chemistry

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

Dr. Bernadette Holdener
Associate Professor, Department of Biochemistry and Cell Biology

Dr. M. Raafat El-Maghrabi
Associate Professor, Department of Physiology and Biophysics
Stony Brook University Medical Center, Stony Brook, NY

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

The Versatile Applications of 6-Alkynyl Fucose and Other Fucose Analogs in Studying Protein Fucosylation

by

Esam Al-Shareffi

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2014

Glycosylation is the modification of a target molecule with carbohydrate. It is a common and complex co- or post-translational protein modification that is responsible for diverse and essential biological roles including cellular communication, modulation of function, and the provision of specific ligands for carbohydrate-binding proteins. Fucosylation is a specific type of glycosylation that results from the transfer of fucose, a unique monosaccharide with the L-configuration, to its target proteins or glycans through the action of fucosyltransferases. Thirteen fucosyltransferases have been identified, including protein *O*-fucosyltransferases 1 and 2 (POFUT1 and POFUT2), responsible for addition of fucose to the hydroxyl group of serine or threonine residues within specific sequences of Epidermal Growth Factor-like (EGF) repeats and Thrombospondin-1 Type Repeats (TSRs) in proteins, respectively, as well as Fucosyltransferase 8 (FUT8), the fucosyltransferase responsible for the addition of fucose to the core *N*-acetylglucosamine on complex-type *N*-linked glycans.

Unlike protein synthesis, glycosylation is not template driven. The resulting glycosylated proteins can be quite complicated, and robust tools are needed to analyze their structures and biology. The use of sugar analogs has revolutionized the study of glycosylation. Modified forms of naturally occurring sugars synthesized with an artificial functional group, which can be incorporated and metabolized by cells to reach their end-target proteins, glycans, or lipids. The functional group then allows for probing, visualizing, or potentially purifying the end-target molecules modified with the sugar analog.

In this dissertation, I first studied the properties of 6-alkynylfucose (6AF), a fucose analog with an alkyne functional group in the C6 position, and its *in vivo* incorporation into glycoproteins expressed in Human Embyronic Kidney 293T (HEK293T) cells. I directly tracked the 6AF by mass spectrometry, taking advantage of the added mass of the alkyne functional group to distinguish 6AF from fucose. Additionally, I utilized the alkyne by reacting it with biotin-azide using the copper-catalyzed azide-alkyne cycloaddition or "click" reaction. Addition of biotin to glycoproteins which incorporated 6AF allowed direct detection using streptavidin probes by Western blot. Using these methods I concluded that 6AF was efficiently incorporated onto EGF repeats and TSRs by POFUT1 and POFUT2, respectively, in a "bioorthogonal" or minimally perturbing manner. I then attempted to use 6AF as part of a proteomics protocol to metabolically label HEK293T cells, carrying out the "click" reaction to biotinylate all fucosylated proteins for purification with streptavidin beads and subsequent identification. While I was able to carry out "proof-of-concepts" experiments where I was able to pulldown and identify overexpressed proteins containing EGF repeats or TSRs known to be fucosylated, I was unable to use the protocol to identify any endogenously expressed fucosylated proteins from HEK293T cells primarily due to their low abundance.

During the initial the 6AF experiments described above, I noticed that 6AF was poorly incorporated into *N*-linked glycans. Using human Immunoglobulin G (IgG) heavy chain, which has a single N-linked glycan known to be efficiently core fucosylated by FUT8, I confirmed that 6AF inhibits core fucosylation. This result suggests that in contrast to the results above for POFUT1 and POFUT2, 6AF is not a substrate but an inhibitor of FUT8. Since inhibitors of fucosyltransferases are valuable as potential therapeutics, I then examined whether two other fucose analogs (5-thiofucose (5TF) and 2-fluorofucose (2FF)) are substrates or inhibitors of POFUT1, POFUT2, and FUT8. All three fucose analogs were at least partially incorporated as substrates by POFUT1 and POFUT2 onto EGF repeats and TSRs. Additionally, however, a mild decrease in overall fucosylation was also observed on these targets, suggesting that all three fucose analogs had simultaneous inhibitory properties on POFUT1 and POFUT2. Only trace amounts of all three fucose analogs were incorporated as substrates by FUT8 on IgG. Moreover, 6AF caused a profound decrease in core fucosylation of IgG, confirming its role as a potent inhibitor of FUT8. A smaller decrease in core fucosylation of IgG observed for 5TF and 2FF, suggesting that these fucose analogs are also inhibitors of FUT8, but less potent than 6AF.

I dedicate this dissertation to my family – my father, mother, brothers, and sister. I love you with all my being.

Table of Contents

List of Figures

Chapter 4

List of Abbreviations

2FF	2-fluorofucose
5TF	5-thiofucose
6AF	6-alkynylfucose
ADCC	Antibody Dependent Cell-mediated Cytotoxicity
α-MEM	Alpha Minimum Essential Medium
ADAMTSL	A Disintegrin-like And Metalloprotease Domain with Thrombospondin-type-I repeats-Like
Asn	Asparagine
AU	Arbitrary unit
AUC	Area under the curve
BCA	Bicinchoninic assay
cax	Compact axial skeleton
CC	Click chemistry
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
CuAAC	Copper(I) catalyzed azide alkyne cycloaddition
Da	Dalton
DMEM	Dulbecco's Modified Eagle's Medium
EGF	Epidermal Growth Factor-like
EIC	Extracted Ion Chromatogram
ER	Endoplasmic Reticulum
ERGIC	Endoplasmic Reticulum-Golgi Intermediate Compartment
EV	Empty vector
FCS	Fetal Calf Serum
Fuc	Fucose
FUT	Fucosyltransferase
FX	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase
GalNAc	*N*-Acetylgalactosamine
GlcNAc	*N*-Acetylglucosamine
GMD	GDP-mannose 4,6-dehydratase
HEK293T	Human Embryonic Kidney 293T
HPLC	High Performance Liquid Chromatography
hT1	human Thrombospondin-1
hTSP	human Thrombospondin-1
IgG	Immunoglobulin G
LADII	Leukocyte Adhesion Deficiency type II
LC/MS	Liquid Chromatography / Mass Spectroscopy
Lfng	Lunatic fringe
Man	Mannose
m/z	Mass to charge ratio
MMP	Matrix Metalloprotease
mN1	mouse Notch 1 (protein)
MS	Mass Spectrum / Mass Spectra
Ni-NTA	Nickel-nitrilotriacetic acid

OptiMEM	Optimum Minimum Essential Medium
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline with Tween
PNGaseF	Peptide *N*-glycosidase F
POFUT	Protein *O*-fucosyltransferase
RIPA	Radioimmunoprecipitation assay
SLC	Solute Carrier (protein family)
TBS	Tris-buffered saline
TBTA	Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TFA	Trifluoroacetic acid
TSR	Thrombospondin type-1 repeat
X	Any amino acid

Acknowledgments

My most sincere and profound thanks go to Dr. Haltiwanger, for his tireless dedication to the lab, the department, and to everyone who is blessed by his presence. Bob (I shudder even now to address him by his first name!) is a mentor of unparalleled scientific and personal greatness. I have been deeply honored by the experience of working in your lab, by the innumerable opportunities and chances you have given me, and by your personal devotion to my success and the welfare of everyone in the lab. Words are inadequate to express my gratitude.

I would also like to thank the members of my committee: Dr. Isaac Carrico, Dr. Deborah Brown, and Dr. Bernadette Holdener, for your kind, thoughtful, and insightful guidance inside and outside of committee meetings, and Dr. Raafat El-Maghrabi, for kindly agreeing to join the committee and for your enthusiastic support which means a great deal to me. My special thanks go to Dr. Erwin London and everyone in the BSB program, for running and being part of a tightly knit and wonderful program; I am also deeply grateful to Dr. Nicole Sampson, Dr. Peter Tonge, and Lauretta Passanant for their incredible efforts in the Chemical Biology Training Program. I am deeply grateful to Dr. Mike Frohman, for his continuing support as head of the Medical Scientist Training Program and Ms. Carron Kaufman who is instrumental in keeping the MSTP running.

I would like to acknowledge members of the Haltiwanger lab, past and present, for their kind and selfless support. We are a family far more than we are work colleagues, and I am deeply indebted to Shinako, Deepika, and Hideyuki for teaching me all that I know in the lab. My thanks also to Beth, Mike, Rich, Megumi, Nadia, Eugene, Devin, and so many others!

Most especially, I am profoundly indebted to my family. My mother, Yasmin; my father, Kadum; my wonderful brothers Hyder and Hassan; my wise and beautiful sister Nadine. I am nothing without the support and care of my family and my greatest aspiration in life is to make them proud.

Chapter 1: Introduction

1.1. Fucose

Fucose is an unusual carbohydrate that is present in a variety of glycolipids as well as a component of *O*-linked and *N*-linked modifications on mammalian proteins. It is unique in having the L-configuration, whereas all other sugars in mammals exist in the D-configuration, and is also structurally distinct in lacking a hydroxyl group on its C-6 carbon. (Figure 1.1)[1] A study of 3299 mammalian oligosaccharides revealed that fucose is found in 7.2% of the oligosaccharides studied, and is a common terminal monosaccharide moiety, second only to sialic acid, making fucose a commonplace component of carbohydrate modifications of proteins and lipids.[2]
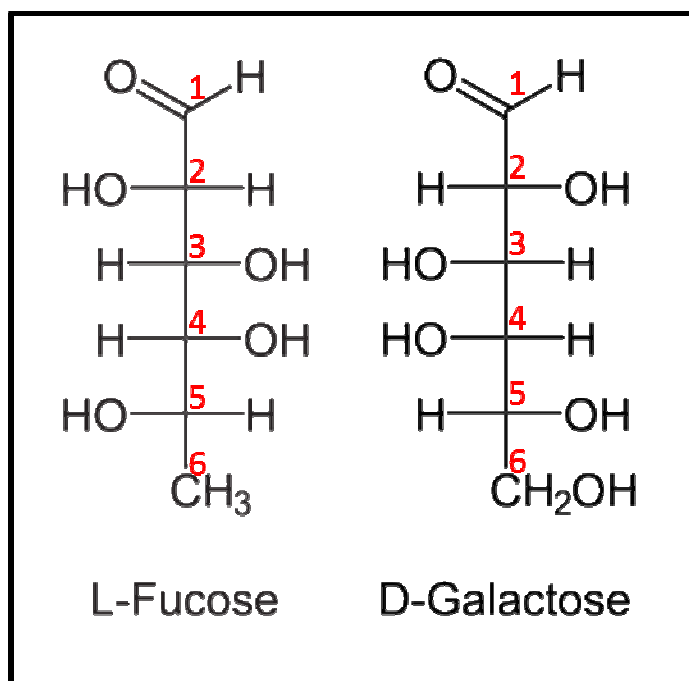
Figure 1.1: Fischer projection of the structure of L-Fucose. The six carbons of fucose are labeled. Note that most naturally occurring sugars, such as galactose, are present in the D-configuration, as can be determined by the arrangement of the hydroxyl-group bound to C-5 carbon. Note further that the C-6 carbon lacks a hydroxyl group. L-Fucose can also be described as 6-deoxy L-galactose. This figure was adapted from structures available in the public domain.

The scientific literature is teeming with published studies implicating fucose in a variety of physiological and pathophysiological processes. For instance, fucose modifies the terminal galactose on the lactosamine-type precursor on a variety of glycoproteins and glycolipids, synthesizing H Antigen (Figure 1.2), which is expressed in "O" blood group individuals, highlighting the role of fucose in the ABO blood group setting.[1,3] Additionally, sialyl-Lewis[b] (Figure 1.2) antigen, of which fucose is a major component, allows attachment of *H. pylori* to the gastric epithelium, which can lead to peptic ulcer disease and gastric cancer. Additionally, fucose is a big part of sialyl-Lewis[x] (Figure 1.2) antigen, which plays an essential role in selectin-dependent leukocyte adhesion and lymphocyte homing. Sialyl-Lewis[x] has also been implicated in mediating human sperm binding to the zona pellucida in fertilization,[4] and overexpression of sialyl-Lewis[x] has been implicated in a variety of cancers.[5] Core fucosylation of *N*-linked glycans (Figure 1.2) affects proper functioning of transmembrane signaling receptors. Disruption of core fucosylation of *N*-linked glycans leads to neonatal lethality and emphysema-like changes in the lungs of mice,[6] while midline patterning and eye development defects are observed in zebrafish mutants.[7] The role of fucose in inflammation and cancer is widely recognized[8-10] and is the subject of numerous ongoing investigations, particularly those concerning the tracking of serum fucose levels or aberrantly fucosylated glycoproteins as biomarkers for cancer diagnosis and prognosis, particularly those of breast, liver, and pancreas.[11-14] Additionally, fucose as well as the enzymes responsible for fucose synthesis, transport, and transfer onto target glycans and proteins, are increasingly the subject of intense study for the purpose of developing therapeutic inhibitors and antibodies in an effort to treat a variety of malignancies.[9,15-17]

In order for fucose to be used as a substrate to modify target proteins or lipids, it must be coupled to GDP. In mammals, GDP-fucose is synthesized through two pathways, known as the

de novo and salvage pathways (Figure 1.3).[1,18] In the *de novo* pathway, GDP-mannose is first

converted to GDP-4-keto-6-deoxymannose, through the enzyme GDP-mannose dehydrogenase

(GMD).[1,19-21] Next, GDP-4-keto-6-deoxymannose is converted to GDP-fucose through the action

of FX, a dual functional epimerase/reductase enzyme.[1,21,22] It is estimated that ~ 90% of GDP-

fucose in mammals is obtained through the de novo pathway under ordinary circumstances,[23]

although this often quoted figure is derived exclusively from one study in which fucose

metabolism was studied in HeLa cells with radiolabeled fucose. In that study, the authors

calculated that of the 11.1 nmol GDP-fucose / 23 hours / 10^7 HeLa cells in the GDP-fucose pool,

10.2 nmol / 23 hours / 10^7 HeLa cells came from an endogenous source (GDP-mannose), while

0.9 nmol / 23 hours / 10^7 HeLa cells, or roughly 8%, came from the exogenous radiolabeled

fucose.

In the salvage pathway, free fucose derived from dietary sources or added to culture

medium is transported into the cytosol by a poorly studied mechanism, presumed to be through

facilitated diffusion, and is then converted to GDP-fucose through a two-step mechanism.[1] First,

fucose kinase converts fucose to fucose-1-phosphate in a reaction that consumes ATP. Next,

GDP-fucose pyrophosphorylase converts fucose-1-phosphate to GDP-fucose in a reaction that

consumes GTP.[1,24]
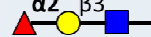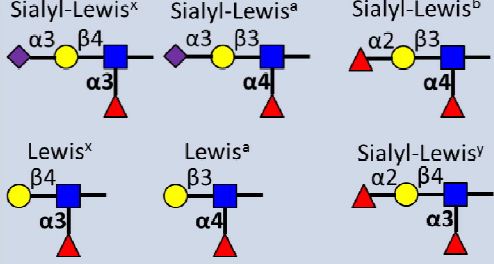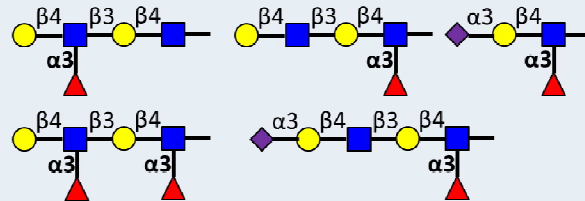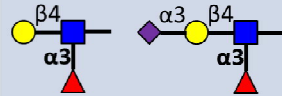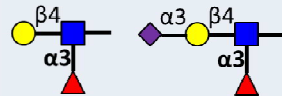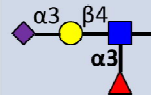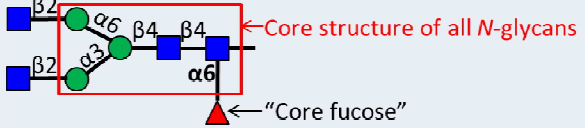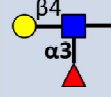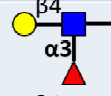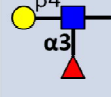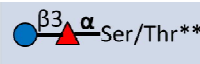
Once the GDP-fucose has been synthesized in the cytosol, it requires the action of a

GDP-fucose transporter to ferry the GDP-fucose to the endoplasmic reticulum or the Golgi

apparatus, the sites where fucosyltransferases use GDP-fucose as a substrate to modify targets of

fucosylation.[25] In humans, the Golgi-localized GDP-fucose transporter is known as Slc35c1, and

defects in the SLC35C1 gene are directly implicated in a rare but devastating human disease

known as leukocyte adhesion deficiency II (LAD2).[26-28] LAD2 is characterized by mental

retardation, growth retardation, immunodeficiency, and leukocytosis without pus formation, directly attributed to the absence of neutrophil sialyl-Lewis[X],[29] which cannot be produced due to the defect in fucose metabolism (Figure 1.2). In *Drosophila*, the homolog of Scl35c1 is called Golgi GDP-fucose transporter (Gfr).[30] Gfr has been demonstrated to transport GDP-fucose in vitro, although *Gfr[-/-] Drosophila*[30] and *Slc35c1[-/-]* mice[31] showed only mild defects that are far less significant than flies or mice unable to synthesize GDP-fucose, suggesting the presence of alternate mechanisms for GDP-fucose transport. The search for such alternate mechanisms in *Drosophila* led to the discovery of the ER GDP-fucose transporter (Efr), which localizes specifically to the endoplasmic reticulum, and seems to function redundantly with Gfr in the *O*-fucosylation of Notch, a transmembrane protein that is crucial for local cell-cell communication, particularly in development.[32] Further work in the field revealed that Slc35c2 might be the mammalian version of *Drosophila* Efr. Overexpression of Slc35c2 was shown to compete with Slc35c1 for GDP-fucose, promote the *O*-fucosylation of EGF repeats 29-36 in Notch 1, and localize to the cis-Golgi network and endoplasmic reticulum-Golgi intermediate compartment (ERGIC), although further work on a targeted mutation in Slc35c2 is required before confirming that Slc35c2 is definitely a GDP-fucose transporter.[33]

Once transported into the lumen of the ER or Golgi, GDP-fucose serves as donor substrate for one of the thirteen known fucosyltransferases in humans.[1] Fucosyltransferases 1-11 (FUT1-11) are localized in the Golgi apparatus, while Protein *O*-fucosyltransferases 1 and 2 (POFUT1 and POFUT2) are localized to the ER.[25] Each fucosyltransferase recognizes a particular glycan or protein site for modification. A list of the known fucosyltransferases and the representative major products they modify is listed in Figure 1.2. A prior review focused on FUT1-11 (Becker and Lowe, 2003)[1], but the last decade has seen the discovery of POFUT1 and

POFUT2 and significant advances in our understanding of FUT8. While all of the fucosyltransferases play important roles in biology, only elimination of FUT8, POFUT1, or POFUT2 results in lethality in mice.[34-37] The remainder of this chapter focuses on summarizing what has been learned recently about these biologically essential enzymes: FUT8, POFUT1, and POFUT2.

Common Name(s)	Abbreviation	Representative Major Product(s)
H blood group α2fucosyltransferase	FUT1	H antigen, type 2
Secretor (Se) blood group α2fucosyltransferase	FUT2	H antigen, type 1
Fuc-TIII α3/4fucosyltransferase Lewis blood group fucosyltransferase	FUT3	Sialyl-Lewis^x Sialyl-Lewis^a Sialyl-Lewis^b Lewis^x Lewis^a Sialyl-Lewis^y
Fuc-TIV α3fucosyltransferase ELAM-1 ligand fucosyl transferase	FUT4	
Fuc-TV α3fucosyltransferase	FUT5	
Fuc-TVI α3fucosyltransferase	FUT6	
Fuc-TVII α3fucosyltransferase	FUT7	
Fuc-TVIII α6fucosyltransferase	FUT8	← Core structure of all *N*-glycans ← "Core fucose"
Fuc-TIX α3fucosyltransferase	FUT9	
Fuc-TX α3fucosyltransferase	FUT10	
Fuc-TXI α3fucosyltransferase	FUT11	
Protein *O*-fucosyltransferase 1	POFUT1	Ser/Thr* Ser/Thr*
Protein *O*-fucosyltransferase 2	POFUT2	Ser/Thr**

Figure 1.2: List of the thirteen known fucosyltransferases in humans.[1,25,38,39] Major representative products of each fucosyltransferase are listed. The linkage of the fucose added by

each enzyme appears in **bold**. *This modification has only been observed on serine or threonine residues that fit the consensus sequence $C^2XXXX(\underline{Ser/Thr})C^3$, where C^2 and C^3 are the second and third conserved cysteines of properly folded Epidermal Growth Factor-like (EGF) repeats. **This modification has only been observed on serine or threonine residues that fit the consensus sequence $C^1XX(X)(\underline{Ser/Thr})C^2XXG$, where C^1 and C^2 are conserved cysteines on properly folded Thrombospondin-1 type repeats (TSRs). Red triangle, fucose; yellow circle, galactose; blue square, *N*-acetylglucosamine (GlcNAc); purple diamond, sialic acid; green circle, mannose; blue circle, glucose.
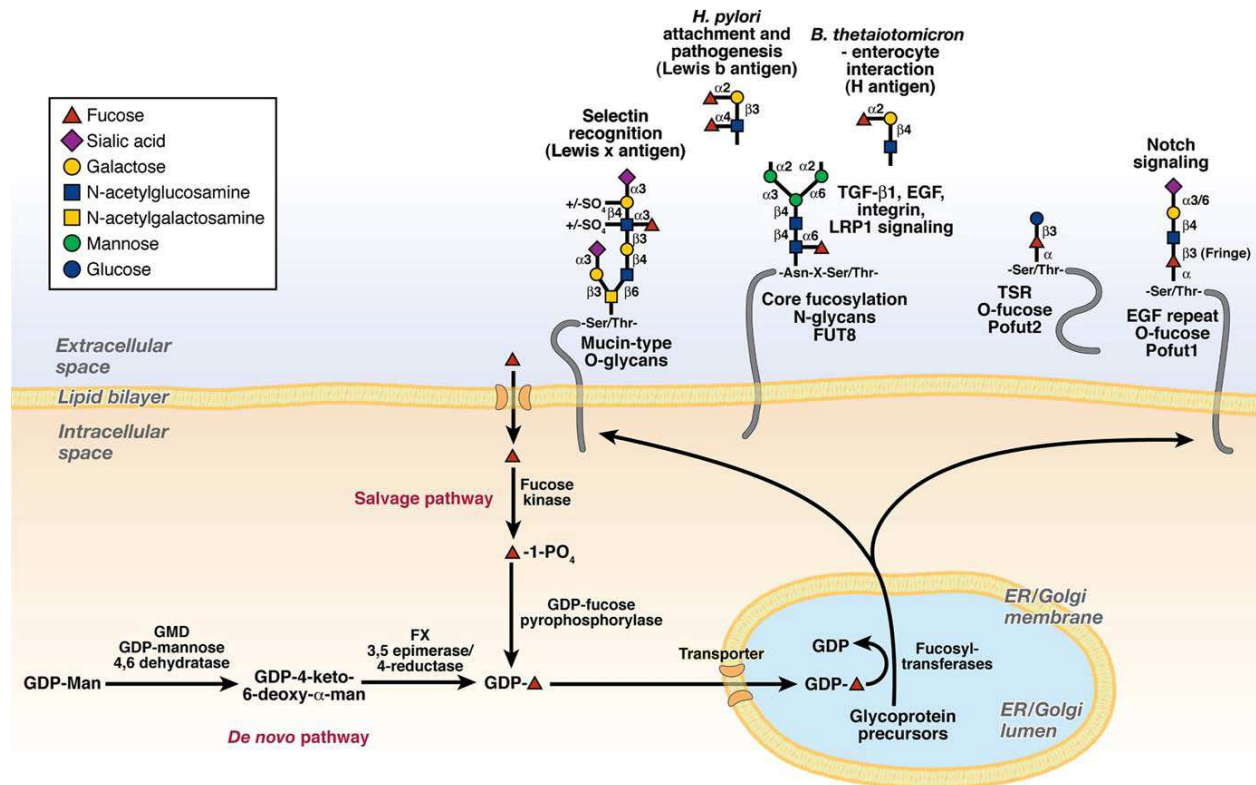
Figure 1.3: *De novo* and salvage pathways for GDP-fucose synthesis. The figure illustrates the *de novo* pathway which converts GDP-Mannose to GDP-Fucose through the enzymes GMD and FX, as well as the salvage pathway, which directly utilizes fucose and converts it to GDP-Fucose through the action of fucose kinase and GDP-fucose pyrophosphorylase enzymes. This figure is presented here with permission from the author.[25]

1.2. Fucosyltransferase 8 (FUT8)

Fucosyltransferase 8 (FUT8) is the only enzyme capable of catalyzing α1,6 fucosylation in mammals.[40] As Figure 1.2 illustrates, many of the other fucosyltransferases catalyze formation of redundant linkages, but FUT8 is unique in targeting the core GlcNAc on complex-type *N*-linked glycans for fucosylation.[41] The generation of *Fut8*[-/-] mice revealed the essential role FUT8 plays in survival and growth. *Fut8*[-/-] mice are born with no apparent anomalies and their appearance is indistinguishable from *Fut8*[+/+] mice, but approximately 70% of the mice die within 3 days of age. Survivors display severe growth retardation, while analysis of *Fut8*[-/-] mice reveals progressive emphysema-like changes in their lungs, likely due to pathologically enhanced expression of matrix metalloproteases (MMPs), especially of McolB (the mouse orthologue of MMP-1), MMP-12, and MMP-13. This enhanced matrix metalloprotease expression was traced to down-regulation of the transforming growth factor (TGF-β1) receptor mediated signaling pathways, which in turn is due to the lack of core fucose on the TGF-β1 receptor.[34] Disruptions in the epidermal growth factor (EGF) receptor and the vascular endothelial growth factor receptor (VEGF), E-cadherin, and α3β1 integrin, caused by lack of core fucosylation of these receptors, was also implicated.[42-44]

The crystal structure of FUT8 has been previously solved.[41] It contains 15 β-strands and 16 α-helices with several distinct features, including a coiled-coil structure formed by two long antiparallel α-helices at the N-terminus, a catalytic domain consisting of a Rossman fold typically seen in nucleotide binding proteins, and the SH3 domain typical of cytosolic proteins located at the C-terminus.[41,45]

The processing and maturation of an *N*-glycan is quite complex and carried out by the concerted action of a staggering number of enzymes. Of particular relevance to FUT8 is that core fucosylation only takes place on "complex-type" *N*-linked glycans, in which "antennae" initiated by GlcNAc are attached to both terminal mannose sugars in the "core structure" common to all *N*-glycans (Figure 1.2). Core fucosylation of complex-type *N*-glycans by FUT8 takes place in the *medial*-Golgi (Figure 1.4).[46]
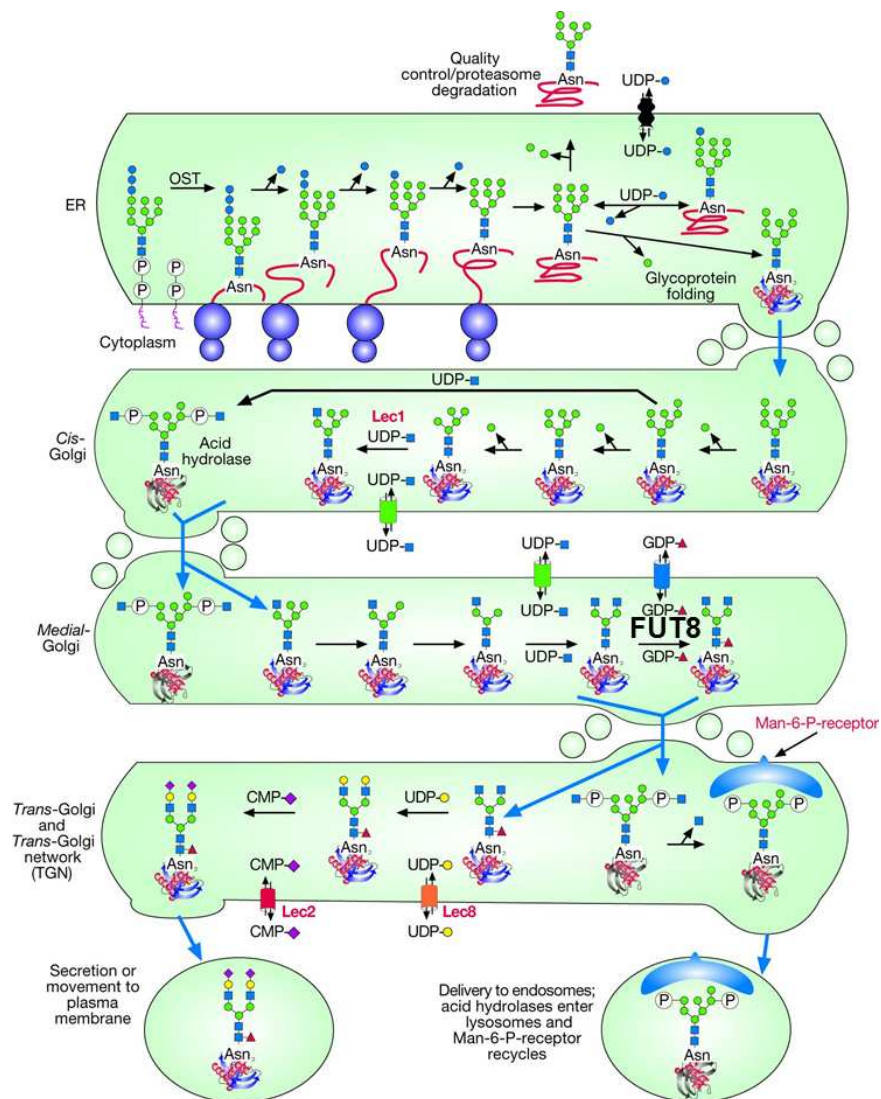
Figure 1.4: Maturation of an *N*-glycan. Following the initial steps of the biosynthesis of the *N*-glycan on the cytoplasmic face of the ER membrane (not shown in this figure), the maturing

N-glycan is further processed in the ER lumen. Complex-type *N*-glycans can be modified by FUT8 (as indicated in the figure) in the *medial*-Golgi. This is an adaptation of Figure 8.4 published in The Essentials of Glycobiology, 2nd edition.[46]

FUT8 has been implicated in a wide variety of physiological and pathological functions. Core-fucosylated alpha-fetoprotein is a well-studied tumor marker for hepatocellular carcinoma – while levels of alpha-fetoprotein are elevated in all forms of liver disease (e.g. cirrhosis, hepatitis, chronic liver disease), only in hepatocellular carcinoma is the fraction of core-fucosylated alpha-fetoprotein elevated.[47,48] Core fucosylation on the *N*-linked glycan of E-cadherin plays a role in the regulation of this important mediator of cell-cell adhesion. Increased levels of FUT8 expression are correlated with increased levels of E-cadherin expression in primary colorectal cancer samples.[49] Similarly, core fucosylation is known to modulate integrin activation and plays an essential role for the functioning of α3β1 integrin in cell migration and signaling, while aberrant glycosylation in the form of increased core fucosylation on α5β1 integrins plays a role in altered cell adhesion seen in cancer invasion and metastasis.[49]

Perhaps of greatest interest is the observation that antibody dependent cellular (or cell-mediated) cytotoxicity (ADCC), achieved through therapeutic antibodies, is enhanced 50-100 fold in core fucose depleted antibodies.[50,51] ADCC is induced by antibody binding to antigens on target cells, including tumor cells. Natural Killer cells (or other immune cells) express a receptor which recognizes and binds to the Fc portion of antibodies, typically Immunoglobulin G, leading to crosslinking of the immune effector cell with the antibody-bound tumor cell and degranulation, causing lysis of the tumor.[52] The interaction of the antibody with the effector cells is reduced greatly in core-fucosylated antibody, and it has been shown that strategies to produce non-core fucosylated antibodies significantly enhance the potency of therapeutic antibodies.[17]

One strategy to achieve this goal has been the testing of various sugar analogs to act as inhibitors of FUT8 (Chapter 4).[53]

1.3. Protein *O*-fucosyltransferase 1 (POFUT1)

O-Fucosylation by POFUT1 occurs at consensus sequences on small cysteine-rich domains known as Epidermal growth factor-like (EGF) repeats, at the consensus, C^2xxxx(S/T)C^3, where C^2 and C^3 are the second and third conserved cysteines of the EGF repeat.[54] While over 100 proteins are predicted to be modified by *O*-fucose based on the presence of the consensus sequence within EGF repeats,[55] the Notch family of receptors have more predicted *O*-fucose sites than any other protein in the databases.[56,57] Results from many groups reveal that *O*-fucosylation is essential for Notch function.[58-61] POFUT1 is critical for Notch signaling activation. *Pofut1* knockout mice show growth retardation by embryonic day 9.5 and die by embryonic day 10 displaying severe morphological defects, particularly in somite formation, as well as in neural tube, cardiac, and blood vessel development.[35,36] POFUT1 also regulates lymphoid and myeloid homeostasis by modulating Notch receptor ligand interactions. Inducible inactivation of Pofut1 results in a reduction in T lymphopoiesis, reduced production of marginal-zone B cells, and myeloid hyperplasia.[62] Outside of POFUT1's role in regulating Notch signaling, heterozygous mutations in POFUT1 have been implicated in a rare disorder, Dowling-Degos Disease (DDD), characterized by dermatological changes which typically include variable hyperpigmentation and hypopigmentation of the flexures.[63] Additional evidence for the sensitivity of certain Notch-dependent processes to POFUT1 levels emerged with the characterization of the "compact axial skeleton" or *cax* mutation that arose spontaneously in a breeding colony, which "carries an insertion of an intracisternal A particle retrotransposon" into an intron of the Pofut1 gene causing a Pofut1 hypomorphic allele that results in the reduction of transcription and consequently reduced Notch expression.[64] The *cax* mutant embryos show specific defects in anterior-posterior somite patterning and axial skeletal development[64], but do

13

not show any other obvious defects, suggesting that even a 50% reduction in POFUT1 activity has phenotypes in some genetic backgrounds.

O-Fucose also appears to play an important functional role in the ability of agrin to cluster acetylcholine receptors,[65] but the function of *O*-fucose on the vast majority of predicted modified proteins, including notable examples like fibrillin-1 and fibulin-1, both essential for elastic fiber formation in connective tissues,[66,67] are unknown. *O*-Fucose on EGF repeats in several proteins can be extended by the Fringe family of β3-N-acetylglucosaminyltransferases[68-70] and other enzymes, ultimately to the tetrasaccharide Siaα2-3/6Galβ1-4GlcNAcβ1-3Fucα1-*O*-Ser/Thr (Figure 1.2).[56,68] There exist three Fringe homologs in mammals, known as Lunatic, Manic, and Radical Fringe, while only one Fringe protein is known to exist in *Drosophila*. In flies, Fringe regulates Notch by inhibiting Notch activation from the "Serrate" ligand, while enhancing Notch activation from the "Delta" ligand. In this way, changing the glycosylation state of the receptor by Fringe elongation of the *O*-fucose monosaccharide (which was added by POFUT1) can regulate Notch signaling. This strategy is used by *Drosophila* to ensure that Notch is only activated at the dorsal/ventral boundary in wing development, even though Notch, Serrate, and Delta are expressed throughout the imaginal wing disc, by tightly regulating Notch activation in only the desired area through Fringe.[71-74]

The structure of POFUT1 derived from *Caenorhabditis elegans* (*Ce*POFUT1), has been solved, both in apo-form (without GDP-fucose bound) and in complex with its ligand, GDP-fucose.[75] *Ce*POFUT1 is a monomeric protein, as confirmed by analytical ultracentrifugation and gel filtration chromatography. It adopts an overall GT-B (Glycosyltransferase-B) fold, characterized by two main "Rossmann-like" domains at the N- and C-termini with a connecting linker region and the catalytic site between these two domains.[75,76] POFUT1 is unique among

fucosyltransferases, in that only POFUT1 and POFUT2 are localized to the endoplasmic reticulum, whereas the other eleven fucosyltransferases (FUT1-FUT11) are all Golgi-resident proteins.[25,77] Moreover, POFUT1 *O*-fucosylates only properly folded EGF repeats, as only one of several differently folded substrates with the same molecular weight (as determined by LC/MS) was modified by POFUT1.[78] Additionally, other than POFUT1's catalytic role in *O*-fucosylation, POFUT1 has been shown to have a distinct chaperone activity in *Drosophila*. This was shown by experiments where depleting POFUT1 by RNAi resulted in the accumulation of Notch in the ER of those cells. Nevertheless, Notch secretion was rescued in those cells with a POFUT1 mutant lacking fucosyltransferase activity.[77]

Recently, POFUT1 was recognized as one of a few potential prognostic marker for breast cancer in a study that examined over 200 glycosylation genes in breast cancer tumor cells compared to normal controls.[79] Of perhaps greater significance, a recent study showed a statistically significant upregulation of POFUT1 expression in oral squamous cell carcinoma-cells as compared to normal oral keratinocytes, raising the prospect that POFUT1 inhibition could be used as a strategy for treating human oral cancer.[80]

1.4. Protein *O*-fucosyltransferase 2 (POFUT2)

O-Fucosylation by protein *O*-fucosyltransferase 2 (POFUT2) takes place on Thrombospondin type-1 repeats (TSRs) at the consensus sequence $C^1X_{2-3}(\underline{S/T})C^2X_2G$.[81-85] Approximately 50 proteins are predicted to be *O*-fucosylated based on the presence of this consensus within TSRs, including all 19 members of the ADAMTS family of matricellular proteases.[37,86] Disruption of *Pofut2* in mice results in embryonic lethality by E10.5 indicating that *O*-fucosylation of TSRs, like that of EGF repeats, is essential for proper development.[37] Nonetheless, the specific targets responsible for the lethality are not yet known. Several reports suggest that *O*-fucosylation of TSRs is required for proper folding and secretion, suggesting that the lethality may be due to loss of secretion of members of the ADAMTS family.[83,84] *O*-Fucose on TSRs is typically extended by a β3-glucosyltransferase,[87,88] and mutations in this enzyme lead to Peters'-plus syndrome, an autosomal recessive congenital disorder of glycosylation characterized by a specific malformation of the eye, and associated with developmental defects, cleft lip/palate, and mental retardation.[89-91] Nonetheless, the biological function of extending *O*-fucose on any of the proteins predicted to be modified is not known.

Like POFUT1, POFUT2 is an ER-localized protein[82] and only *O*-fucosylates properly folded TSRs.[81] Of great significance, Chen et al. solved the 3D structure of human POFUT2 in 2012.[92] The authors were particularly interested in understanding how POFUT2 was able to accommodate over 50 predicted protein targets for *O*-fucosylation and determined that the critical component in TSRs was the presence of a unique three dimensional "AB-loop" conformation. In fact, the authors designed an artificial "mini-TSR" that comprised only 29 amino acids as compared to 55 amino acids in a wild-type TSR, and the authors showed that the mini-TSR was fucosylated successfully by POFUT2. Although the artificial mini-TSR retained

the predicted consensus sequence for *O*-fucosylation by POFUT2, there was still a real prospect

based on those results that POFUT2 could in fact have a broader list of targets than originally

anticipated, since it would at least be theoretically possible that the "AB-loop" confirmation

could be a feature of some proteins that did not have the currently defined POFUT2 consensus

sequence. Therefore, an endeavor to map and define the *O*-fucose proteome could potentially

discover proteins not currently predicted to be *O*-fucosylated, or at a minimum provide

confirmation to the currently generated list of predicted targets, many of which have not been

independently confirmed to be *O*-fucosylated by biochemical or mass spectrometric methods.

1.5. Sugar analogs

The development of sugar analogs has revolutionized the study of glycans by providing a powerful tool for tracking or modifying glycosylation in a number of ways. One strategy involves the metabolic labeling of cells with a monosaccharide analog bearing an azide (or alkyne) chemical handle. The metabolically labeled sugar analog is incorporated onto target glycoproteins and/or glycolipids, allowing visualization after reacting with a reagent bearing the complementary alkyne (or azide) conjugated to an imaging probe. The reaction that covalently couples the azide and alkyne groups is a variation of the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) (informally referred to as the "click" reaction (also see Figure 1.5)).[93,94] This strategy has allowed successful imaging of membrane-associated glycans in developing zebrafish[95] and *C. elegans* cell-surface glycans.[96] Similar strategies have been employed to detect a variety of sugar modifications, including sialylated and fucosylated glycans.[97-99] The same strategy could be expanded by reacting the metabolically labeled sugar analog to a complementary reagent bearing biotin (instead of an imaging probe), as this would allow detection of proteins modified with the labeled sugar analog through anti-biotin or streptavidin probes by Western blot, or by potentially isolating and purifying those proteins by passing them over streptavidin beads. Alternatively, one may use sugar analogs as part of a strategy to inhibit glycosyltransferases, which could be achieved if the sugar analog's structure binds to a target glycosyltransferase but cannot be transferred to the acceptor.
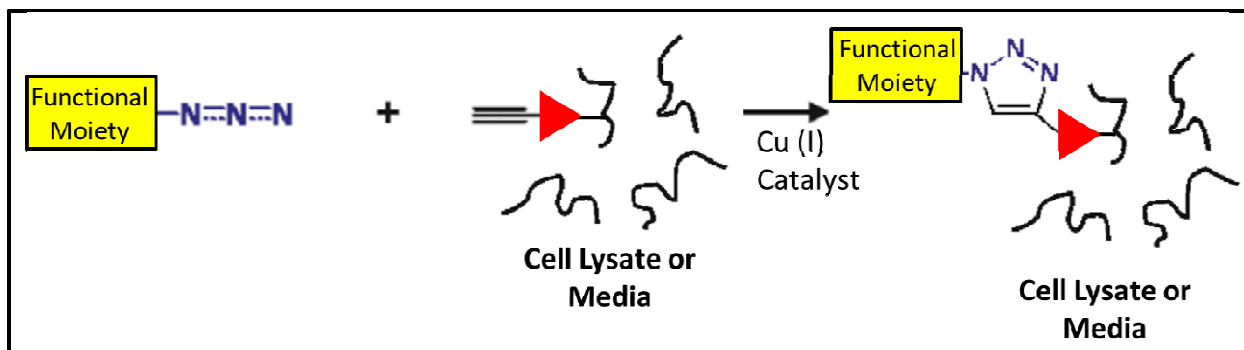
Figure 1.5: Cu(I)-catalyzed azide-alkyne cycloaddition or "click" reaction. The azide group bound to some desirable functional moiety (e.g. fluorescent probe or biotin) is reacted with the alkyne-tagged sugar analog that has already been incorporated into target proteins that are harvested from cell lysate or cell medium. The "click" reaction conditions are generally mild, taking place at room temperature and physiological pH. Even the Cu(I) catalyst can be avoided using alternative reagents for in vivo reactions.

In this dissertation, I utilized three fucose analogs: 6-alkynylfucose (6AF), 5-thiofucose (5TF), and 2-fluorofucose (2FF). 6AF resembles fucose but has an alkyne group incorporated on the C-6 carbon (Figure 1.6A). 6AF was favored over the related azido-fucose analog because the latter has reported toxicity issues in some cells and is thus less useful compared to the alkyne analog[100]. 6AF is taken up through the fucose salvage pathway, eventually leading to addition of 6AF onto proteins that normally bear fucose.[101,102] Of the thirteen known fucosyltransferases (FUT1-FUT11; POFUT1; POFUT2)[1], the tolerance for fucose analogs has already been demonstrated with FUT2-FUT7[102] and FUT9.[101] While most of these studies were done in vitro[102] and with other fucosyltransferases, the demonstration that 6AF was incorporated in those settings gave us confidence that 6AF might be successfully incorporated in HEK293T cells. Additionally, no direct mass spectrometric analysis of proteins where 6AF had been metabolically incorporated was previously reported, raising the prospects of publishing the results for this novel approach. Finally, though fairly expensive, 6AF is commercially available, negating the need to perform a special synthesis to acquire the reagent.

While 6AF was selected for the purpose of incorporation into HEK293T cells by POFUT1 and POFUT2, the other two sugar analogs were mainly selected for their ability to serve as potential inhibitors of fucosylation. 5-thiofucose (5TF) (Figure 1.6B) has been demonstrated to be an inhibitor of sialyl-Lewis[x] biosynthesis, as GDP-5TF was not transferred by either FUT3 or FUT7 in cancer cells, as part of work to develop sialyl-Lewis[x] inhibitors for anticancer therapeutics.[103] 2-fluorofucose (2FF) (Figure 1.6C) was selected because it was shown to have inhibited core fucosylation by FUT8 of IgG in Chinese Hamster Ovary (CHO) cells, as part of work to develop non-core fucosylated antibodies that have enhanced potency in triggering antibody dependent cellular cytotoxicity in the context of anticancer therapeutics.[53]
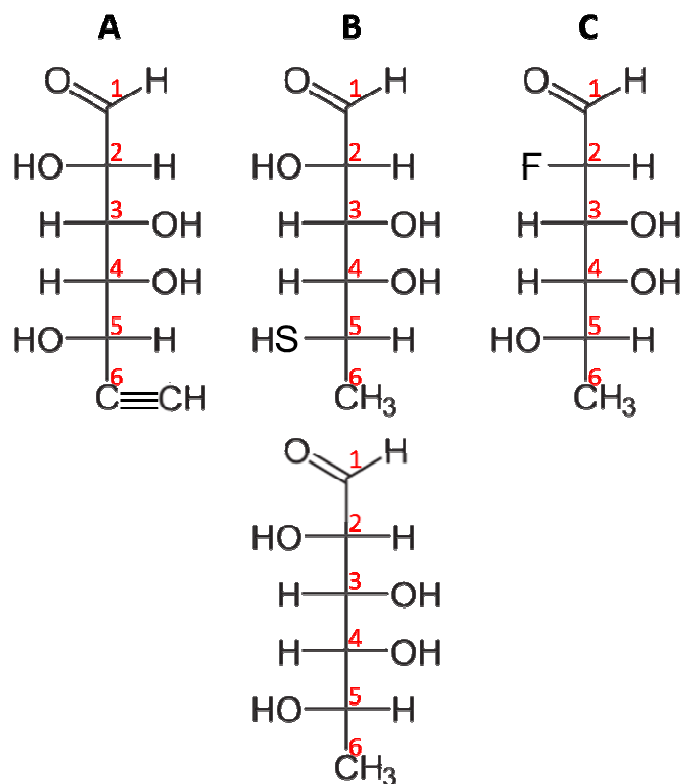
Figure 1.6: Structures of the fucose analogs. A. 6-alkynylfucose (6AF). **B.** 5-thiofucose (5TF). **C.** 2-fluorofucose (2FF). Naturally occurring L-Fucose is shown at the bottom for comparison.

1.6. Aims of this Dissertation

The overall aim of this dissertation is to use sugar analogs to study protein fucosylation. My hypothesis is that 6-alkynylfucose (6AF) is a bioorthogonal analog of fucose in the context of *O*-fucosylation by Protein *O*-fucosyltransferase 1 and 2 (POFUT1 and POFUT2) and can be used as part of a proteomics protocol to determine the *O*-fucose proteome. I further hypothesize that 6AF, as well as other fucose analogs 5-thiofucose (5TF) and 2-fluorofucose (2FF) can potentially inhibit the core fucosylation of *N*-glycans by Fucosyltransferase 8 (FUT8).

Each of these questions is addressed in the ensuing sections. The first focuses on experiments probing the non-perturbing qualities of 6AF in the context of *O*-fucosylation. Next, work on utilizing 6AF as a tool to elucidate the *O*-fucose proteome is described. Finally, the use of 6AF, 5TF, and 2FF as potential inhibitors of FUT8, as well as their impact on POFUT1 and POFUT2, is explored.

Chapter 2: 6-Alkynyl Fucose is a bioorthogonal analogue for *O*-fucosylation of Epidermal Growth Factor-like repeats and Thrombospondin Type-1 repeats by Protein *O*-Fucosyltransferases 1 and 2*

[1]Department of Biochemistry and Cell Biology, Stony Brook University, NY 11794-5215; [2]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125; [3]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; [4]Genomics Research Center, Academia Sinica, 128 Section 2, Academia Road, Nankang, Taipei 115, Taiwan

[*]These authors contributed equally to this work.

2.1. Introduction

Over 150 proteins are predicted to be *O*-fucosylated by POFUT1 or POFUT2 and mutations in these proteins cause embryonic lethality in mice [37,55,86], presumably in part due to changes caused by the abrogation of *O*-fucosylation in these proteins. The only known targets of *O*-fucosylation are proteins which contain properly folded EGF repeats, which are *O*-fucosylated by POFUT1 and proteins containing TSRs which are *O*-fucosylated by POFUT2. Of note, only some of these predicted proteins have actually been confirmed to be *O*-fucosylated by biochemical or mass spectral approaches, while the remaining proteins are only presumed to be *O*-fucosylated based on their containing the currently designated consensus sequence observed for *O*-fucosylation. Moreover, the prospect that additional *O*-fucosylated proteins may exist was indirectly raised by the work of Chen et al. in 2012, when these scientists showed that a "mini-TSR" containing only some of the essential features of a TSR was still *O*-fucosylated by POFUT2.[92] As such, a convenient method that can elucidate all *O*-fucosylated proteins would, in one fell swoop, be capable of confirming that all 150 or so predicted *O*-fucosylation targets are indeed modified. More tantalizingly, such a method would also enable refining the consensus sequence for *O*-fucosylation, by expanding the sequence to fit any new targets found to be modified, and/or by restricting the sequence if some predicted *O*-fucosylation targets are found not to be modified.

The use of a fucose analog, 6-alkynylfucose (6AF), was considered ideal to approaching this problem. The general concept would be to metabolically label a desired cell line with 6AF. The 6AF would then be incorporated onto *O*-fucosylated proteins. These proteins would become biotinylated after reacting the proteins with azido-biotin, utilizing the Cu(I)-catalyzed azide-alkyne cycloaddition, or "click" reaction, and then purified by passage over streptavidin beads

for subsequent proteomics analysis, to determine the *O*-fucose proteome. Beyond the use of 6AF in proteomics, the fucose analog could be used as a means to conveniently track fucosylation in a cell. For instance, in the Haltiwanger lab, several investigators are interested in whether or not *O*-fucosylation plays a role in the secretion of a variety of proteins, like ADAMTS13. Incorporating 6AF into target *O*-fucosylated proteins, "clicking" with azido-biotin, and performing a Western blot probing with streptavidin or anti-biotin, thus provides a convenient assay for assessing the *O*-fucosylation of a protein, which might otherwise require far more complicated, time-consuming, and costly analysis requiring more starting sample (e.g. by proteomics.) Other investigators might be interested in visualizing *O*-fucosylation by fluorescent tagging, which would also be possible by metabolically labeling with 6AF and "clicking" with an azide bound to a fluorophore.

In any case, before utilizing 6AF for any of these purposes, it must be proven that 6AF can be efficiently incorporated into known targets of *O*-fucosylation, namely proteins containing EGF repeats and TSRs. Moreover, as the fucose monosaccharide can be further elongated by other enzymes to a tetrasaccharide (Siaα2-3Galβ1-4GlcNAcβ1-3Fucα1-*O*-Ser/Thr)[56,68] on EGF repeats or to the Glcβ1-3Fuc disaccharide by a β3-glucosyltransferase on TSRs[87,88], it is also important to know whether or not 6AF incorporation perturbs the addition of these additional sugars. If 6AF can be efficiently incorporated by POFUT1 and POFUT2 and does not appreciably alter further glycosylation on these targets, then it could be a viable tool in studying *O*-fucosylation.

2.2. 6-Alkynyl fucose is successfully incorporated onto EGF repeats and TSRs

To determine whether 6AF is tolerated by Pofut1 and Pofut2, we examined metabolic incorporation into several protein fragments: EGF repeats 1-5 from mouse Notch 1 (mN1 EGF1-5), which contains three known Pofut1 modification sites [57,104], and TSRs 1-3 from human Thrombospondin1 (hT1 TSR1-3), which contains three known Pofut2 modification sites [85] (Figure 2.1B). Constructs encoding secreted versions of mN1 EGF1-5 or hT1 TSR1-3 (or untransfected control) were transfected into wild type Human Embyronic Kidney (HEK293T) cells, and incubated with 200 µM peracetylated 6AF, or 200 µM peracetylated fucose, for 72 hours. Cell lysates and culture media were subjected to CuAAC with azido-biotin (+ click chemistry) or without azido-biotin (- click chemistry), separated by SDS-PAGE, and transferred to nitrocellulose. Both mN1 EGF1-5 and hT1 TSR1-3 could be detected by streptavidin peroxidase indicating the presence of 6AF tagged with biotin on the proteins (Figure 2.2A). Reactivity with streptavidin peroxidase required both 6AF and click chemistry (CC). The anti-myc blot confirmed the expression of mN1 EGF1-5 and hT1 TSR1-3 in transfected media samples. Control media, obtained under identical conditions from untransfected cells, showed no signal with streptavidin or anti-myc. Cell lysate samples, obtained from control cells incubated with 6AF and subjected to the click reaction, showed efficient labeling of a variety of fucosylated glycoproteins. Most of these protein species disappeared in samples without click chemistry or without 6AF incubation, with the exception of two distinct bands, presumably corresponding to endogenously biotinylated proteins (Figure 2.2A). These results demonstrate 6AF is successfully incorporated into EGF repeats and TSRs and selective detection is achieved.
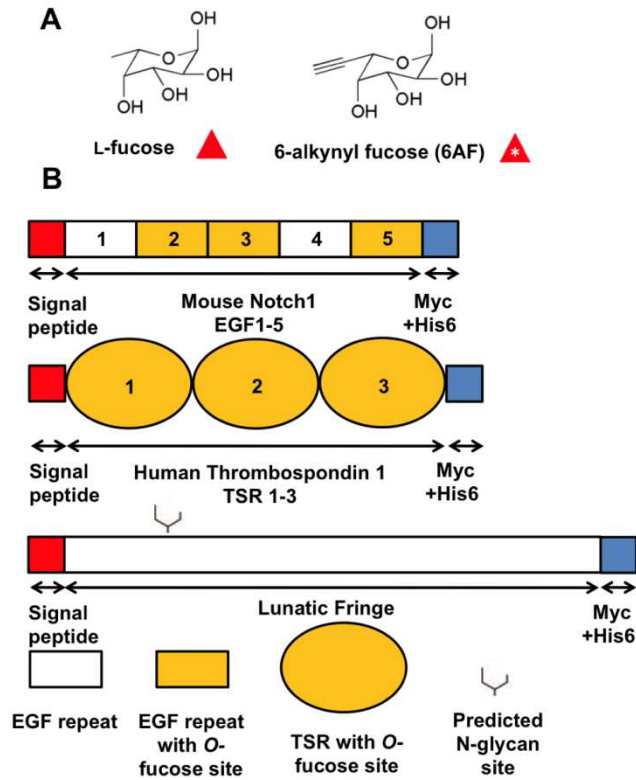
Figure 2.1: Structure of 6AF and domain map of the constructs used. **A.** Structures of **L**-fucose and 6AF. **B.** All expression constructs contained an N-terminal Igk signal peptide (red) and C-terminal Myc-His$_6$ tags (blue) for detection by Western blot and purification. Mouse Notch 1 EGF 1-5 contains three *O*-fucosylated EGF repeats (indicated by yellow shading) [54], and all three TSRs in Human Thrombospondin 1 TSR 1-3 are *O*-fucosylated [85], while neither construct contains an N-glycsoylation site. Lunatic Fringe is not predicted to be *O*-fucosylated and contains one *N*-glycosylation site at Asn-167.

2.3. 6-Alkynyl fucose is specifically incorporated onto fucosylated glycoproteins

To confirm that 6AF is specific to fucosylation within the cell and is not metabolized into other sugars, we examined whether the 6AF signal could be eliminated by excess **L**-fucose (Figure 2.2B). Incorporation of 6AF onto endogenous glycoproteins in HEK293T cells decreased in a dose-dependent fashion as the concentration of **L**-fucose was increased. In contrast, incubation of control cells with 6AF and **D**-galactose had no effect on 6AF incorporation (Figure 2.2C). These results establish that 6AF is specifically utilized by fucosylation pathways within cells.
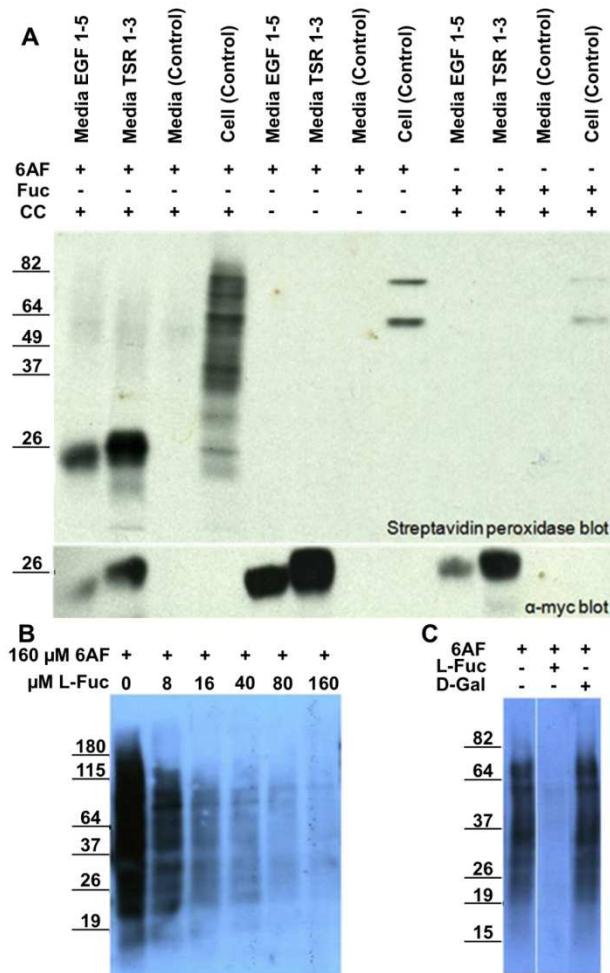
Figure 2.2: 6AF is incorporated into EGF repeats and TSRs in HEK293T cells. A. HEK293T cells were either untransfected (Control) or transfected with plasmids encoding mN1

EGF1-5 or hT1 TSR1-3, and incubated with 200 µM peracetylated 6AF (6AF) or peracetylated **L**-fucose (Fuc) for 3 days. Media and lysates were subjected to CuAAC with (+CC) or without (-CC) azido-biotin. The samples were then subjected to SDS-PAGE and Western blotting using Streptavidin-HRP and anti-myc. **B.** Cell lysates were prepared from untransfected HEK293T cells incubated with 160 µM peracetylated 6AF and increasing amounts of **L**-Fucose, subjected to CuAAC with azido-biotin, and analyzed as described above. **C.** Cell lysates were prepared from untransfected HEK293T cells incubated with 200 uM peracetylated 6AF in the presence or absence of 200 µM **L**-fucose or **D**-galactose, subjected to CuAAC with azido-biotin, and analyzed as described above.

2.4. 6-Alkynyl fucose is incorporated into both *N*-linked and *O*-linked glycans

The results from Figure 2.2 suggest that 6AF is incorporated onto EGF repeats and TSRs as well as a large number of unidentified proteins in untransfected cell lysates. As an initial means of characterizing whether the 6AF is incorporated into *N*- or *O*-linked glycans in the crude lysates, we took advantage of Pro5 (wild type) and Lec1 Chinese hamster ovary cells. Lec1 cells are unable to synthesize complex *N*-linked glycans [105] and are predicted to have little or no fucosylated *N*-glycans [106]. To confirm that 6AF is incorporated onto EGF repeats, TSRs, and *N*-glycans in this system, we transfected constructs encoding mN1 EGF1-5, hT1 TSR1-3, and Lfng (which contains a single consensus site for *N*-linked glycan modification on Asn-167 and no sites of *O*-fucosylation (Figure 2.1), into Pro5 cells grown in the presence of 6AF. All three proteins were detected in the media of transfected Pro5 cells (anti-myc; red) and incorporated 6AF as measured by streptavidin (green) in a click chemistry dependent manner (Figure 2.3A). Although the streptavidin signal for the +CC Lfng is clearly above background (-CC), it is much lower than that for either EGF1-5 or TSR1-3 (Figure 2.3B). Since Lfng does not contain EGF or TSRs, we predicted that the 6AF signal on Lfng resulted from modification of its *N*-glycan. To test this, we performed PNGase F digestion on all three proteins. PNGase F completely removed the streptavidin signal (green) from Lfng and caused a mass shift of several kDa, characteristic of the loss of an *N*-glycan, but had no effect on the 6AF signal on EGF1-5 or TSR1-3 (Figure 2.3C). Moreover, when these proteins were expressed in Lec1 cells, 6AF incorporation was lost on Lfng, but not for EGF1-5 or TSR1-3, consistent with 6AF incorporation into the *N*-glycan on Lfng (Figure 2.3D).

Comparison of 6AF incorporation into endogenous proteins in Pro5 versus Lec1 cells (Figure 2.3E) revealed that while many proteins are labeled in both CHO cell lines, several are

lost in Lec1 cells. These results indicate that much of the 6AF is incorporated into *N*-glycans on proteins in Pro5 cells. The proteins labeled in Lec1 cells lysates are presumably modified with 6AF on EGF repeats or TSRs, or possibly on mucin-type *O*-glycans.
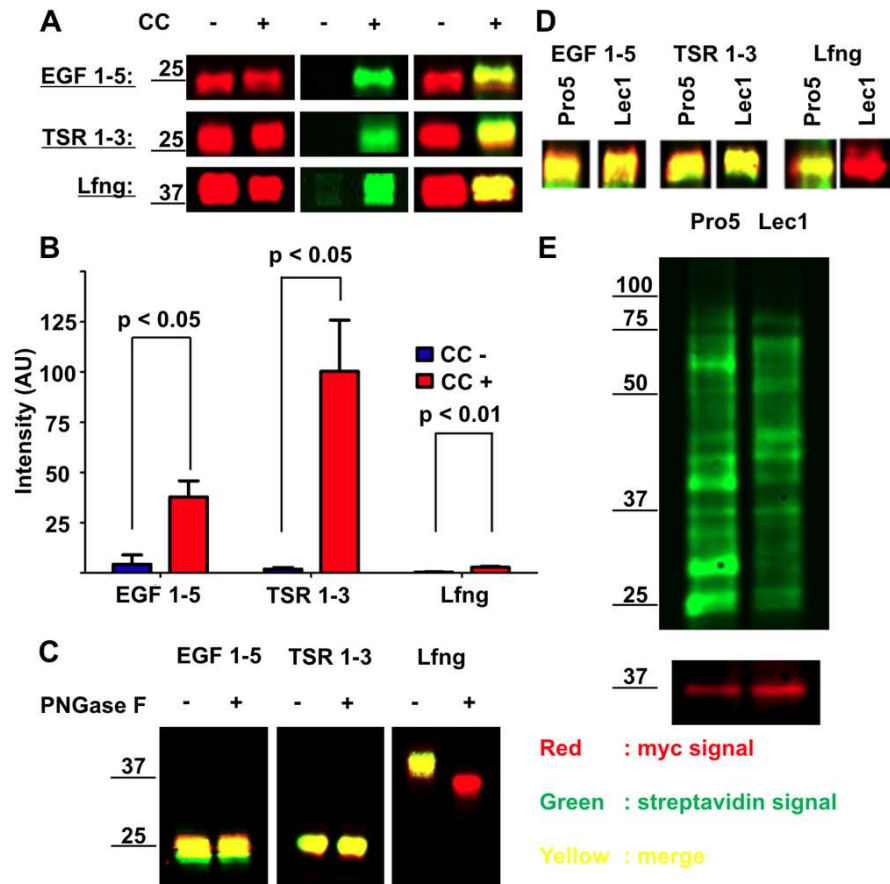
Figure 2.3: 6AF is incorporated into N-glycans found on Lunatic fringe and numerous proteins in crude extracts of CHO cells. A. Pro5 cells were transfected with plasmids encoding mN1 EGF1-5, hT1 TSR1-3, or Lfng (Figure 2.1) and grown in the presence of 200 µM peracetylated 6AF. Media samples were subjected to CuAAC with (+ Click Chemistry) or without (- Click Chemistry) azido-biotin and analyzed as described in Materials and Methods. Protein was monitored using the anti-myc probe (red), and biotin using the streptavidin probe (green), while merging of the two channels results in yellow. **B**. Bar graph illustrating the intensity (in arbitrary units or AU) of streptavidin signal in EGF 1-5, TSR 1-3, and Lfng samples, with (+CC) and without (-CC) performing the click chemistry reaction. The experiment was performed in triplicate and the streptavidin intensity values were normalized to anti-myc intensity values. Paired student t-test was performed to determine statistical significance. **C**. Samples prepared as in A were subjected to PNGase F digestion as described in Materials and Methods. **D.** Plasmids encoding mN1 EGF1-5, hT1 TSR1-3, and Lfng were transfected into either Pro5 or Lec1 cells in the presence of 200 µM peracetylated 6AF, and media samples were analyzed as in A. **E**. Pro5 and Lec1 cells transfected with Lfng were grown in the presence of 200 µM peracetylated 6AF and cell lysates were biotinylated and analyzed as in A. Many proteins are labeled in both cell lines, but the signal is lost on several proteins in Lec1 cells. The bottom panel shows an anti-Myc blot for the Lfng-Myc-His$_6$ to show that similar amounts of extract were loaded in each lane.

2.5. 6-Alkynyl fucose is efficiently incorporated onto *O*-fucose sites on EGF repeats and does not perturb glycan elongation by Fringe

Although Figure 2.2 demonstrates incorporation of 6AF into EGF repeats and TSRs, it does not address whether the POFUT1 and 2 consensus sites are modified, measure the efficiency with which the fucose analogue is incorporated onto the site, nor determine whether the presence of the alkyne affects glycan elongation. Specifically, *O*-fucose on EGF repeats can be elongated by the Fringe family of β3-N-acetylglucosaminyltransferases to the disaccharide GlcNAcβ1-3Fucose, and further elongated by other enzymes to the mature tetrasaccharide Siaα2-3-Galβ1-4GlcNAcβ1-3Fuc[68]. To address the incorporation efficiency and determine whether the presence of the alkyne perturbs glycan elongation beyond the *O*-fucose on EGF repeats, we used a mass spectral approach. The mass of 6AF is 10 Da greater than that of fucose (accounting for one additional carbon and a net subtraction of two hydrogen atoms, Figure 2.1A) and thus the mass-to-charge (m/z) ratio will differ based on incorporation of the analogue.

To examine 6AF incorporation and glycan elongation, Pro5 cells were co-transfected with the plasmids encoding mN1 EGF1-5 and Lfng and grown in the presence of either 6AF or **L**-fucose. Lfng elongates the *O*-fucose beyond a monosaccharide[68]. The mN1 EGF1-5 protein was purified from medium, digested with trypsin, and the resulting glycopeptides were subjected to nano-LC-MS/MS. Mass spectra of a tryptic glycopeptide containing the consensus site for *O*-fucosylation from EGF3 of mN1 EGF1-5 show ions at m/z of 866.4, 968.0, and 1048.7 (Figure 2.4A; top panel). These ions correspond to the predicted masses of doubly-charged forms of the mono-, di-, and trisaccharide glycoforms of this peptide with natural fucose (note that the tetrasaccharide form of this glycopeptide is not observed under these conditions). The MS/MS spectra of each glycoform (bottom panels) reveal the sequential neutral losses of the modifying

glycans as well as several "b" and "y" ions confirming the identity of the peptide. Mass spectra of the same glycopeptide obtained from a sample generated with 6AF show ions at m/z of 871.5, 972.9, and 1054.1, which correspond to the above masses plus five mass units (Figure 2.4B). An m/z difference of 5 in the 2+ charged state is consistent with incorporation of 6AF on the peptide containing the *O*-fucosylation site of EGF3. Fragmentation of each glycoform (Figure 2.4B, bottom panels) shows that the extra mass is lost at the appropriate position for the fucose. Surprisingly, very little glycopeptide with m/z corresponding to fucose (e.g. at 866.4, 968.0, or 1048.7) is observed in the MS spectra from the samples generated with 6AF.

To evaluate the relative amounts of each glycoform in these samples, we performed a semiquantitative mass spectral analysis. Extracted Ion Chromatograms (EIC) were generated for each glycoform of the EGF3 peptide present in the samples, searching for fucose or 6AF forms using the corresponding masses as determined in Figures 2.4A and 2.4B (Figure 2.4C Monosaccharide glycoform; Figure 2.4D Disaccharide glycoform; Figure 2.4E Trisaccharide glycoform). As expected, only the fucose form of the glycopeptides was found in the samples prepared in the presence of **L**-fucose, while the 6AF form of each glycopeptide was the predominant signal detected in the 6AF samples. These results indicate that 6AF is incorporated at very high efficiencies onto the *O*-fucose site of EGF repeat 3 under these conditions, and that elongation by Fringe is not affected by 6AF.

A EGF 3 incubated in L-fuc

MS

866.4

968.0

1048.7

= ^{113}NGG$\underline{T}$CDLLTLTEYK^{126}
= Fucose
= GlcNAc
= Galactose

MS/MS (866.7)

793.2

y_3 439.4 y_4 540.4 y_5 653.5 y_{10} 1256.0

MS/MS (968.3)

793.2

y_3 439.4 y_4 540.4 y_5 653.5 y_{10} 1256.7

MS/MS (1049.4)

793.2

y_3 439.5 y_4 540.4 y_5 653.5 y_{10} 1255.9

B EGF 3 incubated in 6AF

MS

871.5

972.9

1054.1

= ^{113}NGG$\underline{T}$CDLLTLTEYK^{126}
= 6AF
= GlcNAc
= Galactose

MS/MS (871.6)

793.2

y_3 439.3 y_4 540.4 y_5 653.5 y_{10} 1255.8

MS/MS (973.3)

793.2   871.1

y_3 439.4 y_4 540.4 y_5 653.5 y_{10} 1255.9

MS/MS (1054.1)

793.2   872.0

y_3 439.4 y_4 540.4 y_5 653.5 y_{10} 1255.9

C Monosaccharide   **D** Disaccharide   **E** Trisaccharide

L-fuc   6AF   L-fuc   6AF   L-fuc   6AF

33

Figure 2.4: 6AF is efficiently incorporated into the *O*-fucosylation site in EGF3 of mouse Notch1 and elongated by Lfng. A. Pro5 cells were co-transfected with plasmids encoding mN1 EGF1-5 and Lfng grown in the presence of 200 µM peracetylated **L**-fucose. mN1 EGF1-5 was purified from the medium, digested with trypsin, and subjected to nano-LC-MS/MS analysis as described in Materials and Methods. MS (top panel) and MS/MS (lower panels) spectra of a peptide from EGF3 containing the *O*-fucose consensus sequence (represented by a gray bar). The ions at m/z 866.4, 968.0, and 1048.7 correspond to the doubly charged forms of the mono-, di-, and trisaccharide glycoforms of this peptide. The MS/MS spectra of each glycoform shows loss of the modifying glycans to the unglycosylated form of the peptide (m/z 793.2), as well as several "b" and "y" fragmentation ions (indicated in red). **B.** mN1 EGF1-5 was prepared as in A except that the cells were grown in 200 µM peracetylated 6AF. MS (top panel) and MS/MS (lower panels) spectra of the same peptide from EGF 3. The ions at m/z 871.5, 972.9, and 1054.1 correspond to the doubly charged forms of the mono-, di-, and trisaccharide glycoforms of this peptide with 6AF. These masses closely match the expected increase of m/z of 5 as compared to the **L**-fuc samples, as the peptides are in the 2+ charge state and as 6AF is 10 Da heavier than **L**-fucose. The MS/MS spectra confirm loss of the modifying glycans to the unglycosylated form of the peptide (m/z 793.2), and show several "b" and "y" fragmentation ions (in red). Extracted Ion Chromatograms (EIC) for samples obtained from cells incubated in **L**-fucose (+**L**-Fuc) and for samples obtained from cells incubated in 6AF (+6AF), for the monosaccharide glycoform (**C**), the disaccharide glycoform (**D**), and the trisaccharide glycoform (**E**) of the peptide from EGF3. Gray rectangle, peptide; red triangle, fucose; red triangle with *, 6AF; blue square, GlcNAc; yellow circle, galactose.

2.6. 6-Alkynyl fucose is efficiently incorporated onto *O*-fucose sites on TSRs and can be readily elongated to the disaccharide form

As with the incorporation of 6AF onto EGF repeats, we examined the efficiency of 6AF incorporation onto the glycopeptide containing the *O*-fucosylation site from TSR2 of TSR 1-3, and whether the β3-glucosyltransferase responsible for elongating *O*-fucose on TSRs tolerates the analogue. Figure 2.5A shows an MS spectrum (top panel) of the tryptic glycopeptide containing the consensus site for *O*-fucosylation of TSR 2, obtained from a sample generated in HEK293T cells incubated in **L**-fucose. The MS spectra contain an ion at m/z of 933.7, corresponding to the quadruply charged form of the glycopeptide bearing the disaccharide Glcβ1-3Fuc. The MS/MS spectra (Figure 2.5A, bottom panel) indicate ions at m/z 893.2 and 856.8, corresponding respectively to the monosaccharide and unmodified forms of the peptide. We observed a similar MS spectrum, obtained from a sample generated as described above, but from cells cultured with 6AF (Figure 2.5B). The MS spectrum shows a peak at 936.1, which closely corresponds to the mass of the glycopeptide analyzed in Figure 2.5A plus 2.5 mass units. An m/z difference of 2.5 in the quadruply charged state is consistent with the incorporation of 6AF in place of **L**-fucose into the peptide containing the *O*-fucosylation site from TSR 2. Fragmentation of the glycopeptide (Figure 2.5B, bottom panel) demonstrates that the extra mass is lost at the appropriate position for the fucose. The EIC for the 6AF sample shows that the glycopeptide containing 6AF was the principal form detected (Figure 2.5C). These results reveal that 6AF is efficiently incorporated into the *O*-fucosylation site of TSR2, and demonstrate that 6AF can be efficiently elongated by the β3-glucosyltransferase.

A TSR 2 incubated in L-fuc

MS

x10⁵

Intensity

933.7

=
⁴³²FKQDGGWSHW#SPW#
SSCSV<u>T</u>CGDGVITR⁴⁵⁸

▲ = Fucose

● = Glucose

MS/MS (933.8)

x10⁵

856.8

893.2

933.8

y_3 389.4

y_9^{+2} 489.9

y_5 545.6

y_7 717.7

b_{18}^{+2} 1223.0

400 600 800 1000 1200 m/z

B TSR 2 incubated in 6AF

MS

x10⁵

Intensity

936.1

=
⁴³²FKQDGGWSHW#SPW#
SSCSV<u>T</u>CGDGVITR⁴⁵⁸

⊕* = 6AF

● = Glucose

MS/MS (936.2)

x10⁴

856.6

895.8

936.2

y_3 389.3

y_9^{+2} 489.9

y_5 545.5

y_6 660.7

y_7 717.6

b_{18}^{+2} 1223.2

400 600 800 1000 1200 m/z

C Disaccharide

x10⁷ L-fuc

Intensity

x10⁵ 6AF

8 9 (min)

8 (min)

36

Figure 2.5: 6AF is efficiently incorporated into TSR2 of Thromospondin1 and elongated by β3-glucosyltransferase. **A.** Pro5 cells were transfected with the plasmid encoding hT1 TSR1-3 and grown in the presence of 200 µM peracetylated fucose. hT1 TSR1-3 was purified from the medium, digested with trypsin, and subjected to nano-LC-MS/MS analysis as described in Materials and Methods. MS (top panel) and MS/MS (bottom panel) spectra of a peptide from TSR 2 containing the *O*-fucose consensus sequence (represented by gray bar). The ion at m/z 933.7 corresponds to the quadruply charged form of this peptide modified with the *O*-fucose disaccharide. The MS/MS spectra indicate loss of the modifying glycans, first to the monosaccharide form (m/z 893.2), and then to the unglycosylated form of the peptide, (m/z 856.8). Several "b" and "y" fragmentation ions confirming the identity of the peptide are also indicated in black. **B.** hT1 TSR1-3 was prepared as in A except that the cells were grown in the presence of 200 µM peracetylated 6AF. MS (top panel) and MS/MS (bottom panel) spectra of the same peptide from TSR 2 is shown. The ion at m/z 936.1 corresponds to the quadruply charged form of this peptide modified with the *O*-6AF disaccharide. The MS/MS spectra indicate sequential loss of glycans, first to the monosaccharide (m/z 895.8) and then to the unmodified form (m/z 856.6). These masses closely match the expected increase of m/z of 2.5, as compared to the **L**-fucose labeled sample, as the peptides are in the 4+ charge state and as 6AF is 10 Da heavier than **L**-fucose. Several "b" and "y" ions are also indicated in black. **C.** EIC for sample obtained from cells incubated in **L**-fucose (+**L**-Fuc) and for sample obtained from cells incubated in 6AF (+6AF), for the disaccharide glycoform of TSR 2. Gray rectangle, peptide; red triangle, fucose; red triangle with *, 6AF; blue circle, glucose.

2.7. Discussion and Conclusions

Here we demonstrated that 6AF is efficiently incorporated into EGF repeats, TSRs, an *N*-glycan on Lfng, and *N*-glycans on a number of proteins in crude lysates of CHO cells. 6AF is incorporated onto the predicted *O*-fucosylation sites of the EGF repeats and TSRs, and metabolic labeling at the concentrations used (200 µM) leads to highly efficient incorporation. Glycan elongation of the *O*-fucose on either EGF repeats or TSRs does not appear to be affected by the presence of 6AF. 6AF can also be incorporated onto the *N*-linked glycan of Lfng, though apparently with less efficiency as compared to the *O*-fucosylation of EGF repeats and TSRs. Based on these results, 6AF appears to be well tolerated by POFUT1, POFUT2, Lunatic fringe and β3-glucosyltransferase during biosynthesis of *O*-fucose glycans on EGF repeats and TSRs. These data suggest that 6AF is a non-perturbing bioorthogonal analogue to **L**-fucose for the study of *O*-fucosylation.

The mass spectral data provide strong evidence in support of 6AF incorporation into the correct *O*-fucosylation sites of EGF3 and TSR2, as indicated by the increased mass of 10 Da (reflected by an increased m/z of 5 for the EGF3 peptide in the doubly charged state and an increased m/z of 2.5 for the TSR2 peptide in the quadruply charged state), versus the same peptides with **L**-fucose instead. Interestingly, virtually all of the glycopeptides incorporated 6AF at the concentration we used for metabolic labeling (200 µM). Normally, the fucose salvage pathway is believed to account for less than 10% of the fucose incorporated into proteins in cells [107]. The nearly stoichiometric incorporation of 6AF into the *O*-fucose sites of EGF repeats and TSRs indicates that a concentration of 200 µM 6AF is sufficient to shift almost all GDP-fucose production from the *de novo* pathway to the salvage pathway, as has been suggested elsewhere [108]. These results also suggest that POFUT1 and POFUT2 utilize GDP-6AF quite efficiently.

While no crystal structure of GDP-6AF bound to these enzymes exists, the data suggest that the alkynyl group does not sterically inhibit binding of the nucleotide sugar to the enzyme active site. The recently published structures for POFUT1 from *C. elegans* and human POFUT2 in complex with GDP-fucose suggest that there is sufficient space in the active site for the alkyne group[75,109]. Future structural studies on both of these enzymes should help us to better understand how GDP-6AF binds to the enzyme active sites.

In contrast to the EGF repeats and TSRs, incorporation of 6AF into N-glycans appears to be sub-stoichiometric under the conditions used in this study. The data in Figure 2.3A clearly demonstrates some incorporation of 6AF into the N-glycan on Lfng, albeit much lower than its incorporation into the *O*-fucose sites of EGF or TSR repeats (Figure 2.3B). There are several potential explanations for this observation. It is possible that only a small fraction of the *N*-linked glycan on Lfng is normally fucosylated. So that, even if 6AF was incorporated at high efficiency, the signal detected by Western blot would be quite low. Alternatively, fucosylation of the *N*-linked glycan may occur with high frequency, but 6AF itself might be a poorly tolerated substrate for the appropriate fucosyltransferase. Finally, the fucose on the *N*-glycan may be less accessible for biotinylation during the click reaction. Interestingly, there is a slight difference in the streptavidin intensity between normalize amounts of the EGF repeats and TSRs (Figure 2.3B), even though the mass spectral analysis shows that 6AF is incorporated at high stoichiometries on both (Figures 2.4 and 2.5). This suggests that there may be a slight difference in the reactivity of the 6AF on EGF repeats versus TSRs during the click reaction.

In observing the crude lysates of Pro5 and Lec1 cells (Figure 2.3E), several protein species appear in the Pro5 cell lysates but not in the Lec1 cell lysates, presumably because they bear fucosylated complex-type *N*-linked glycans, which would not be expected to be synthesized

in Lec1 cells. Exhaustive genomic sequencing of CHO cell lines conducted by Xu and coworkers indicates that the major *N*-glycan modifying fucosyltransferase found to be expressed in CHO cells is FUT8[110]. Assuming this is the case, our results suggest that 6AF is utilized by FUT8, although we cannot yet comment on how efficiently it is used.

Building on the reports that GDP-6AF is tolerated by FUT2-FUT7 and FUT9, we can now add POFUT1, POFUT2, and most likely FUT8 (though possibly with a lower efficiency) to the growing list of fucosyltransferases that can utilize 6AF. This observation raises the prospect of using 6AF as a means to track fucosylation in a myriad of settings. A recent study used 6AF, after genetically engineering the cells, in a proteomic study to identify proteins modified on *N*-glycans by FUT9 [101]. Similar studies could be done to identify proteins modified with several other fucosyltransferases. Because 6AF does not appear to interfere with removal of *N*-glycans by PNGase F (Figure 2.3), proteins modified with *N*-linked versus *O*-linked 6AF modified glycans could be separated prior to analysis. 6AF can be a powerful tool in tracking fucosylation, but the very versatility and non-perturbing nature of the fucose analogue also demands that caution be exercised in its use. Particularly, imaging studies that utilize 6AF must consider the possibility that the fucose analogue will modify a variety of acceptor substrates, perhaps at different levels of efficiency based on the particular fucosyltransferase involved, impacting the conclusions that can be drawn from such an experiment. Nevertheless, the potential uses of all sugar analogues and 6AF in particular, are vast and will likely continue to play a critical role in glycobiology research for the foreseeable future.

2.8. Materials and methods

Production of mouse Notch 1 EGF 1-5, human TSR 1-3, and human Lunatic Fringe constructs

The construct for mouse Notch 1 EGF repeats 1-5 was generated as described previously.[54] The construct for human TSR 1-3 was generated by modifying the procedure in[81] to include the cDNA sequence encoding TSP1-TSR1-3 instead of TSR3. Successful cloning was confirmed by direct DNA sequencing. The mouse Lunatic Fringe construct was generated as described previously.[111] All constructs were made in pSecTag2/Hygro C vector (Invitrogen (Life Technologies), Grand Island, NY), which is designed for high-level expression and protein secretion. The vector contains an Igκ chain leader sequence for protein secretion, as well as C-terminal myc epitope and 6-histidine residues, to aid in detection and purification.

Cell Culture, Metabolic Labeling, and Protein Expression

Pro5 and Lec1 Chinese Hamster Ovary (CHO) cells were grown in αMEM with 10% fetal calf serum (FCS), while Human Embryonic Kidney (HEK293T) cells were grown in DMEM with 10% FCS. Transfections were performed at 50% confluence with either pSecTag mNotch 1 EGF 1-5, pSecTag hTSP1 TSR1-3, or pSecTag Lunatic Fringe (Lfng). Briefly, 6 µg of plasmid was suspended in 600 µL 0.15M NaCl followed by addition of 36 µL PEI transfection reagent, vortexed, and incubated for 10 minutes at room temperature. The transfection mixture was added to cells on 100 mm plates, incubated at 37°C for 4 hours, media removed, and cells were washed once with 2.5 mL Tris-buffered saline (TBS). Control cells were incubated overnight in 10 mL of αMEM or DMEM, 10% FCS, followed by 72 hours in 5 mL of OptiMEM (GIBCO (Life Technologies), Grand Island, NY), while experimental cells were additionally supplemented with 200 µM peracetylated 6-alkynyl fucose (6AF) as previously described.[97]

Detection of Glycoproteins in Cell Lysate/Media

Media was collected and cells, washed twice in TBS, were lysed in 1 mL of chilled lysis buffer as described (1% Nonidet P-40, 150 mM NaCl, Protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN), 100 mM sodium phosphate, pH 7.5).[97] Cell lysates were incubated on ice and vortexed briefly every five minutes for 30 minutes. Lysates were centrifuged at 14,000 x g for 10 minutes at 4°C and the pellet discarded. Click reactions were performed as described[97] on both the labeled media and labeled cell lysate to tag glycoproteins containing 6AF with Biotin (0.1 mM Azido-Biotin (Click Chemistry Tools, Scottsdale, AZ), 0.1 mM tris-(benzyltriazolylmethyl)amine catalyst (AnaSpec, Fremont, CA), 1 mM copper sulfate, 2 mM sodium ascorbate, in PBS, at room temperature for one hour).

Tagged cell lysate and media were analyzed by immunoblot. Samples from media and lysates were separated by 10% SDS-PAGE transferred to nitrocellulose paper (BioRad, Hercules, CA) and probed with 9E10 α-Myc (1:1000 in PBST), followed by Alexa Fluor 680 Goat Anti-Mouse IgG (Invitrogen (Life Technologies), Grand Island, NY) or HRP-conjugated Goat Anti-Mouse IgG (Jackson ImmunoResearch, West Grove, PA) to detect the transfected proteins. To visualize biotinylated glycoproteins, tagged media and cell lysate samples were separated by 10% SDS-PAGE, transferred to nitrocellulose membrane, and probed with IRDye800 (Rockland Immunochemicals, Glibertsvile, PA) or Streptavidin-conjugated-peroxidase (1:20,000 in PBST) (Thermo Scientific, Rockford, IL). Western blots using fluorescent probes (Alexa Fluor 680 or IRDye800) were visualized using an Odyssey Imager (LI-COR, Lincoln, NE), while western blots with peroxidase conjugated probes were detected by enhanced chemiluminescence (ECL) blotting substrate (Thermo Scientific, Rockford, IL) and film development.

Specificity/Competition experiments

HEK293T cells were grown to 50% confluency in DMEM, 10% FCS, 1% penicillin/streptomycin. Cells were washed one time with 5 mL TBS and resuspended in optimum MEM media supplemented with (0, 8, 16, 40, 80, 160) µM **L**-fucose and 160 µM peracetylated 6AF or 200 µM galactose and 200 µM peracetylated 6AF. After 72 hours at 37 °C and 5 % CO_2 cells were lysed, biotinylated via click chemistry, and 1 µg per lane was resolved using SDS-PAGE and Western Blotting as described above. Protein concentration was quantified using the BCA protein assay with BSA as standard (Thermo Scientific, Rockford, IL).

Enzymatic cleavage of N-linked glyans by PNGase F

PNGase F digestion was performed as previously described.[106] Briefly, protein from media samples was precipitated in 8 volumes of acetone at -20° C overnight. The precipitated proteins were collected by centrifugation at 10,000 g at 4ºC for 15 minutes. The protein pellet was resuspended in 25 µL of 1% SDS, 1% β-mercaptoethanol by boiling for 5 minutes. After cooling, 225 µL of 50 mM Tris-HCl (pH 8.6), 0.7% Nonidet P-40, and Protease inhibitor cocktail (1 Protease inhibitor tablet per 10 mL of buffer) was added. Samples were digested with 5 U of PNGase F (produced in-house) for at least eight hours at 37°C.

Glycan analysis via nano-LC-MS/MS

Media from transiently transfected pSecTag Notch 1 EGF1-5 (and co-transfected with pSecTag Lfng, in a 2:1 ratio) Pro5 CHO cells or pSecTag hTSP1 TSR1-3 HEK293T cells supplemented with either 6AF or **L**-fucose (above) was collected after 72 hours, bound to Ni^{2+}-NTA, washed with 5 mL TBS containing 0.5M NaCl and 10 mM Imidazole, 5 mL cold RIPA buffer, and eluted with TBS containing 250 mM Imidazole. Protein concentration and purity were assessed by SDS/PAGE and Coomassie Brilliant Blue (Thermo Scientific, Rockford, IL) along BSA standards. Approximately 1 µg EGF1-5 or TSR1-3 was acetone precipitated with

four volumes acetone, vortexed, and frozen at -80°C overnight. Precipitated protein was reduced, alkylated, and typsinized, followed by analysis by nano-LC-MS/MS using a Zorbax 300SB-C18 nano-CHIP on an Agilent model 6340 ion trap mass spectrometer using collision-induced dissociation (CID) as previously described.[104]

Chapter 3: Towards Identifying the *O*-Fucome

3.1. Introduction

Following the identification and verification of 6-alkynylfucose (6AF) as a bioorthogonal analog for *O*-fucosylation of the targets of Protein O-Fucosyltransferase 1 and 2 (POFUT 1 and 2), namely Epidermal Growth Factor-like (EGF) repeats and Thrombospondin Type-1 repeats (TSRs) (Chapter 2), the next logical step was to use 6AF in a proteomics protocol, to elucidate all targets of *O*-fucosylation in a given system, such as proteins generated in a plate of HEK293T cells. Media of HEK293T cell plates would be conditioned with an appropriate concentration of peracetylated 6AF, which would then be taken across the cell membrane, converted to 6AF by nonspecific esterases, and incorporated onto target proteins using the same cellular machinery responsible for *O*-fucosylation. After an appropriate time optimized for maximum incorporation of the 6AF, the cells would then be lysed and subject to the copper catalyzed azide-alkyne cycloaddition ("click") reaction, by reacting the alkyne group in 6AF with the azide group of azido-biotin, effectively biotinylating all *O*-fucosylated proteins. The biotinylated proteins would then be selectively enriched by passage over streptavidin beads, taking advantage of the strong interaction between biotin and avidin (or streptavidin) to capture biotinylated proteins on the beads while washing away all other proteins. Next, the proteins were either captured by the streptavidin beads and subjected directly to protease digestion or the proteins would first be eluted from the streptavidin beads and then the eluent subjected to protease digestion. In either case, the peptides generated would then be further processed according to established protocols for mass spectrometry samples and subject to liquid chromatography / electrospray ionization mass spectrometry (LC/ESI-MS), and the ions generated compared to a database of all human proteins, in order to identify presumably *O*-fucosylated proteins. As the ensuing pages will demonstrate, in "proof-of-concept" experiments where HEK293T cells are transfected with and

overexpress proteins that are known targets for *O*-fucosylation, we are able to selectively purify

and identify the overexpressed protein in cells treated with 6AF but not in untreated control cells.

However, we have been unable to identify any endogenously expressing targets of *O*-

fucosylation, presumably because the portion of fucosylated protein in a given plate of

HEK293T cells is too low, and it would be technically as well as monetarily unfeasible to

generate a large enough supply of treated HEK293T cells to enable identification of *O*-

fucosylated proteins by this method.

3.2. Proof-of-concept experiments – HEK293T cells transfected with mN1 EGF1-5 and hTSP TSR1-3

One plate of HEK293T cells transfected with mN1 EGF1-5 (Figure 2.1) with no added sugar analog, as well as an equivalent plate of HEK293T cells transfected with mN1 EGF1-5 but with 200 µM 6AF, were generated. Media sample, rich in the overexpressed EGF1-5, was collected from each plate and subjected to the "on-bead digestion" proteomics method #1 with trypsin. The generated peptides were then subject to ESI-MS, and the predominant peaks in the analysis were compared to an *in silico* trypsin digest of the mouse genome, using the GPM database.

As expected, the GPM database search of the media sample generated from cells not conditioned with 6AF did not generate any matches for the overexpressed mouse Notch 1 protein (Figure 3.1A). This is because the proteins treated with 6AF are not expected to be biotinylated, and thus are not captured by the streptavidin beads. The results for that sample did indicate a large number of hits for a variety of "keratins", fibrous structural proteins that are the major component of human skin and hair, indicating that sample handling issues may have generated contamination of the sample. Other proteins detected were either the result of non-specific association of proteins with the streptavidin beads (beta-casein) or of the protease that was added for the digestion of the protein (trypsin).

The GPM database search of the media sample generated from cells conditioned with 6AF did generate the expected match for the overexpressed mouse Notch 1 protein (Figure 3.1B), which is highlighted with a red arrow. While only 3% of the entire mouse Notch 1 protein structure was identified in the tryptic digest in the sample, it is important to note that the GPM search considered the entire structure of the mouse Notch 1 protein, containing 2531 amino

acids, whereas only the portion containing EGF 1-5 was overexpressed in the sample. The EGF

1-5 portion consists of only 200 amino acids (amino acids 20-217), of which three peptides

containing 42 amino acids were detected (Figure 3.1C). Thus, 42/200, or 21% of the

overexpressed protein was detected. The search also revealed hits for keratins and trypsin

similar to the non-6AF treated sample, as well as a number of presumably non-specifically

bound proteins (transferrin and serum albumin) which are not predicted to be fucosylated

proteins.

EGF 1-5

A

rank	log(e) ▲	log(I)	%/%	#	total	Mr	Accession
1	-192.2	8.34	32/53	19	21	66.0	ENSP00000252244 gpmDB | psyt | snap [18780/37359] homo (1/43) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
2	-115.5	7.84	28/49	11½	12½	58.8	ENSP0000269576 gpmDB | psyt | snap [14419/28221] homo (1/4) protein KRT10:p, keratin 10 [Source: HGNC 6413] IPR016044 F IPR002957 (x5) Keratin I IPR009053 Prefoldin
3	-105.9	7.72	25/39	11	12	62.0	ENSP00000246662 gpmDB | psyt | snap [13330/26451] homo (1/1) protein KRT9:p, keratin 9 [Source: HGNC 6447] IPR016044 F IPR002957 (x5) Keratin I
4	-60.6	7.72	17/21	6	6	65.4	(H) ENSP00000310861 gpmDB | psyt | snap [14931/29363] homo (2/43) protein KRT2:p, keratin 2 [Source: HGNC 6439] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
5	-41.1	7.58	4.4/7	2	2	60.0	(H) ENSP00000369317 gpmDB | psyt | snap [9531/18503] homo (2/14) protein KRT6A:p, keratin 6A [Source: HGNC 6443] IPR016044 F IPR003054 (x6) Keratin II IPR009053 (x2) Prefoldin
6	-31.9	7.36	21/26	4	4	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [38518/74952] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
7	-8.0	6.51	8.0/20	1	2	25.1	sp|CASB_BOVIN| gpmDB | psyt | snap [2844/5607] protein Beta-casein; Contains: Casoparan; Contains: Antioxidant peptide; Contains: Casohypotensin; Flags: Precursor;
			μ½ =	#2	#2		

B

rank	log(e) ▲	log(I)	%/%	#	total	Mr	Accession
1	-107.3	8.07	20/32	11	11	77.0	ENSP00000385834 gpmDB | psyt | snap [3233/6146] homo (2/6) protein TF:p, transferrin [Source: HGNC 11740] IPR001156 (x14) Peptidase S60 IPR016357 Transferrin
2	-75.1	8.10	16/19	8	9	69.2	sp|ALBU_BOVIN| gpmDB | psyt | snap [20438/40789] homo (0/9) protein Serum albumin; BSA; Bos d 6; Flags: Precursor;
3	-73.4	7.92	14/24	8½	8½	66.0	ENSP00000252244 gpmDB | psyt | snap [18780/37359] homo (1/13) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
4	-28.4	7.59	8.0/14	4	4	58.8	ENSP0000269576 gpmDB | psyt | snap [14419/28221] homo (1/5) protein KRT10:p, keratin 10 [Source: HGNC 6413] IPR016044 F IPR002957 (x5) Keratin I IPR009053 Prefoldin
5	-27.0	7.62	20/26	4	4	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [38518/74952] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
6	-26.1	7.46	13/21	4	4	62.0	ENSP00000246662 gpmDB | psyt | snap [13330/26451] homo (1/1) protein KRT9:p, keratin 9 [Source: HGNC 6447] IPR016044 F IPR002957 (x5) Keratin I
7	-24.4	7.97	1.5/3	3	3	270.7	ENSMUSP00000028288 gpmDB | psyt | snap [269/479] homo (3/3) protein Notch1:p, Notch gene homolog 1 (Drosophila) [Source:MGI Symbol;Acc:MGI:97363] IPR002110 (x13) Ankyrin rpt IPR020683 (x2) Ankyrin rpt-contain dom IPR024600 DUF3454 notch IPR000742 (x36) EG-like dom IPR006210 (x36) EGF-like IPR001881 (x39) EGF-like Ca-bd IPR006209 (x27) EGF-like dom IPR013111 (x9) EGF extracell IPR008297 Notch IPR022362 (x5) Notch 1 IPR022355 (x3) Notch 4 IPR011656 Notch NODP dom IPR010660 Notch NOD dom IPR000800 (x12) Notch dom

50

C

```
  1 mprlltpllcltllpalaarglrcsqpsgtclnggrcevangteacvcsgafvgqrcqds 60
    MPRLLTPLLCLTLLPALAARGLRCSQPSGTCLNGGRCEVANGTEACVCSGAFVGQRCQDS
 61 npclstpcknagtchvvdhggtvdyacscplgfsgplcltpldnaclanpcrnggtcdll 120
    NPCLSTPCKNAGTCHVVDHGGTVDYACSCPLGFSGPLCLTPLDNACLANPCRNGGTCDLL
121 tlteykcrcppgwsgkscqqadpcasnpcanggqclpfessyicrcppgfhgptcrqdvn 180
    TLTEYKCRCPPGWSGKSCQQADPCASNPCANGGQCLPFESSYICRCPPGFHGPTCRQDVN
181 ecsqnpglcrhggtchneigsyrcacrathtgphcelpyvpcspspcqnggtcrptgdtt 240
    ECSQNPGLCRHGGTCHNEIGSYRCACRATHTGPHCELPYVPCSPSPCQNGGTCRPTGDTT
```

Figure 3.1: GPM database search results for HEK293T media samples overexpressing mN1 EGF 1-5. A. Results for sample not incubated in 6AF indicate no matches to the overexpressed Notch protein fragment. A large number of keratins, as well as trypsin and a non-specifically bound protein are also identified. **B.** Results for sample incubated in 200 µM 6AF. The search result indicates a match for Notch 1 in the tested sample (indicated by red arrows). A large number of keratin proteins as well as presumably non-specifically bound proteins (transferrin and serum albumin) are also identified. **C.** The amino acid sequence for EGF 1-5 (amino acids 20-217) with three peptides (highlighted in red) that were detected in the GPM database search.

The same experiment was attempted in HEK293T cells overexpressing hTSP1 TSR 1-3, with or without 200 µM 6AF. The experiment yielded similar results, in that the GPM database search on the sample from cells without 6AF did not indicate any matches to the overexpressed hTSP1 protein, whereas the GPM search on the sample from cells incubated in 6AF resulted in a match to hTSP1 protein (Figure 3.2A and 3.2B). As with the analysis for mN1 EGF 1-5, the GPM database search's sequence coverage match percentage (5% in this case) underestimates the true sequence coverage, as only 169 amino acids (amino acids 379-547) out of the total 1170 amino acids of hTSP1 are expressed in the TSR1-3 plasmid. The true sequence coverage is of 54 out of169 amino acids, approximately 32% of the total, and these amino acids are highlighted in red in Figure 3.2C. Also similar to the mN1 EGF 1-5 experiment, a large number of keratins and nonspecifically bound proteins were detected in all samples.

TSR 1-3

A

rank	log(e) ▲	log(l)	%/%	#	total	Mr	Accession
1	-83.5	7.86	14/24	8	9	66.0	ENSP00000252244 gpmDB	psyt	snap [18829/37457] homo (1/6) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
2	-75.2	7.62	19/33	8½	10½	58.8	ENSP00000269576 gpmDB	psyt	snap [14450/28283] homo (1/5) protein KRT10:p, keratin 10 [Source: HGNC 6413] IPR016044 F IPR002957 (x5) Keratin I IPR009053 Prefoldin
3	-46.0	7.20	14/22	5	6	62.0	ENSP00000246662 gpmDB	psyt	snap [13369/26528] homo (1/1) protein KRT9:p, keratin 9 [Source: HGNC 6447] IPR016044 F IPR002957 (x5) Keratin I
4	-31.0	7.37	5.2/6	3	3	65.4	(H) ENSP00000310861 gpmDB	psyt	snap [14981/29463] homo (2/6) protein KRT2:p, keratin 2 [Source: HGNC 6439] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
5	-3.1	6.26	0.8/1	1	1	174.7	ENSP00000355106 gpmDB	psyt	snap [1047/2057] homo (6/6) protein AGL:p, amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase [Source: HGNC 321] IPR010401 GDE C IPR006421 Glycogen debranch met
6	-1.9	5.83	2.6/3	1	1	74.8	ENSP00000305107 gpmDB	psyt	snap [72/131] protein GIMAP8:p, immune associated nucleotide 6; immune associated nucleotide. [Source: RefSeq (NM_175571)] Annotated domains: IPR006703 AIG1 family IPR001687 ATP/GTP-binding site motif A (P-loop)
7	-1.2	7.71	3.0/4	1	1	24.4	sp|TRYP_PIG| gpmDB	psyt	snap [38708/75332] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

B

rank	log(e) ▲	log(l)	%/%	#	total	Mr	Accession
1	-93.4	7.79	14/23	9	9	66.0	ENSP00000252244 gpmDB	psyt	snap [18829/37457] homo (1/20) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
2	-72.5	7.65	16/28	8	9	58.8	ENSP00000269576 gpmDB	psyt	snap [14450/28283] homo (1/9) protein KRT10:p, keratin 10 [Source: HGNC 6413] IPR016044 F IPR002957 (x5) Keratin I IPR009053 Prefoldin
3	-60.1	7.34	17/27	7½	7½	62.0	ENSP00000246662 gpmDB	psyt	snap [13369/26528] homo (1/1) protein KRT9:p, keratin 9 [Source: HGNC 6447] IPR016044 F IPR002957 (x5) Keratin I
4	-30.0	6.99	9.2/11	4	4	69.2	sp|ALBU_BOVIN| gpmDB	psyt	snap [20572/41058] homo (0/9) protein Serum albumin; BSA; Bos d 6; Flags: Precursor;
5	-23.4	7.60	4.3/5	3	4	129.3	ENSP00000260356 gpmDB	psyt	snap [4160/7833] protein THBS1:p, thrombospondin 1 [Source: HGNC 11785] IPR008985 (x2) ConA-like lec gl IPR000742 (x2) EG-like dom IPR006210 (x3) EGF-like IPR001881 (x3) EGF-like Ca-bd IPR001791 Laminin G IPR000884 (x12) Thrombospondin 1 rpt IPR003367 (x11) Thrombospondin 3-like rpt IPR008859 Thrombospondin C IPR001007 (x3) VWF C
6	-16.8	7.30	1.4/2	1	1	65.4	(H) ENSP00000310861 gpmDB	psyt	snap [14981/29463] homo (2/20) protein KRT2:p, keratin 2 [Source: HGNC 6439] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
7	-16.8	6.77	4.6/7	2	3	77.0	ENSP00000385834 gpmDB	psyt	snap [3256/6192] homo (2/3) protein TF:p, transferrin [Source: HGNC 11740] IPR001156 (x14) Peptidase S60 IPR016357 Transferrin
8	-1.4	7.61	3.0/4	1	1	24.4	sp|TRYP_PIG| gpmDB	psyt	snap [38708/75332] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

C

```
361 atvpdgeccprcwpsdsaddgwspwsewtscstscgngiqqrgrscdslnnrcegssvqt 420
    ATVPDGECCPRCWPSDSADDGWSPWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT
421 rtchiqecdkrfkqdggwshwspwsscsvtcgdgvitrirlcnspspqmngkpcegeare 480
    RTCHIQECDKRFKQDGGWSHWSPWSSCSVTCGDGVITRIRLCNSPSPQMNGKPCEGEARE
481 tkackkdacpinggwgpwspwdicsvtcgggvqkrsrlcnnptpqfggkdcvgdvtenqi 540
    TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNNPTPQFGGKDCVGDVTENQI
541 cnkqdcpidgclsnpcfagvkctsypdgswkcgacppgysgngiqctdvdeckevpdacf 600
    CNKQDCPIDGCLSNPCFAGVKCTSYPDGSWKCGACPPGYSGNGIQCTDVDECKEVPDACF
```

Figure 3.2: GPM database search results for HEK293T media samples overexpressing hTSP1 TSR 1-3. A. Results for sample not incubated in 6AF does not indicate any matches to the overexpressed TSP1 protein fragment. A large number of keratins, as well as trypsin and a non-specifically bound protein are also identified. **B.** Results for sample incubated in 200 µM 6AF. The search result indicates a match for thrombospondin 1 in the tested sample (indicated by red arrows). A large number of keratin proteins as well as presumably non-specifically bound proteins (transferrin and serum albumin) are also identified. **C.** The amino acid sequence for TSR 1-3 (amino acids 379-547) with two peptides (highlighted in red) that were detected in the GPM database search.

3.3. Proof-of-concept experiments – HEK293T cells transfected with ADAMTSL1 TSR1-4

While the results above did confirm that mN1 EGF1-5 and hTSP1 TSR 1-3 can be selectively detected only in samples from cells treated with 6AF, the proteomics protocol had much room for improvement, particularly in eliminating non-specifically bound proteins during the wash steps, as well as emphasizing greater care in sample handling to reduce or eliminate keratin hits. As a result, changes in the protocol to institute more extensive washing, specifically by increasing the number of washes from five to ten, increasing the wash volume from 1 mL to 10 mL, and using more stringent wash buffers (detailed in "methods" section) were attempted. Moreover, altering the choice of overexpressed protein from mN1 EGF1-5 or hTSP1 TSR1-3 to a larger fragment from a different protein, ADAMTSL1 TSR1-4, would be useful in increasing the "score" of the GPM hit generated, as well as show robustness of the proteomics protocol for an additional target protein. All of these changes were incorporated to the "on-bead digestion" proteomics method #2 (Section 3.9), which was used to analyze HEK293T samples transfected with ADAMTSL1 TSR1-4, -/+ 6AF.

As anticipated, the GPM database search on the non-6AF treated cells gave no hits for the overexpressed ADAMTSL1 fragment (Figure 3.3A). Gratifyingly, the only protein detected was trypsin, the protease that is exogenously added in order to carry out the protease digestion, and the presence of keratins and non-specifically bound proteins was completely eliminated.

The GPM database search on the 6AF-treated cells gave a convincing hit for the overexpressed ADAMTSL1 TSR 1-4 (Figure 3.3B). The detection of 9% of the ADAMTSL1 protein's peptides in the sample is an underestimate of the total percentage of the overexpressed fragment detected, as the GPM search is performed with respect to the entire ADAMTSL1 protein, not the shorter ADATMSL1 TSR1-4 fragment used in the experiment. In fact, 106 out

of the total 551 amino acids (amino acids 33-584) of ADAMTSL1 expressed, or just over 19%, were detected (Figure 3.3C). The GPM database search also revealed a weak hit for one keratin is as well as the expected match for trypsin.

ADAMTSL1

A

rank	log(e)▲	log(I)	%/%	#	total	Mr	Accession
1	-6.5	7.59	8.2/10	1	1	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [45253/87650] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

			μ½ =	#0	#3		

B

rank	log(e)▲	log(I)	%/%	#	total	Mr	Accession
1	-67.6	8.02	5.6/9	7½	7½	193.3	ENSP00000369921 gpmDB | psyt | snap [8/219] homo (2/5) protein ADAMTSL1:p, ADAMTS-like 1 [Source: HGNC 14632] IPR007110 (x4) Ig-like dom IPR013098 (x4) Ig I-set IPR003599 (x4) Ig sub IPR003598 (x4) Ig sub2 IPR013151 (x2) Immunoglobulin IPR010909 (x2) PLAC IPR013273 (x3) Peptidase M12B ADAM-TS IPR000884 (x43) Thrombospondin 1 rpt
2	-6.2	6.22	1.7/3	1	1	66.0	ENSP00000252244 gpmDB | psyt | snap [21520/42769] homo (1/1) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 (x6) Keratin II IPR009053 Prefoldin
3	-4.2	6.96	8.2/10	1	1	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [45253/87650] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

C

```
  1 meccrratpgtllllflafllllssrtarseedrdglwdawgpwsecsrtcgggasyslrrc  60
    MECCRRATPGTLLLLFLAFLLLLSSRTARSEEDRDGLWDAWGPWSECSRTCGGGASYSLRRC
 61 lsskscegrniryrtcsnvdcppeagdfraqqcsahndvkhhgqfyewlpvsndpdnpcs   120
    LSSKSCEGRNIRYRTCSNVDCPPEAGDFRAQQCSAHNDVKHHGQFYEWLPVSNDPDNPCS
121 lkcqakgttlvvelapkvldgtrcytesldmcisglcqivgcdhqlgstvkedncgvcng  180
    LKCQAKGTTLVVELAPKVLDGTRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVCNG
181 dgstcrlvrgqyksqlsatksddtvvaipygsrhirlvlkgpdhlyletktlqgtkgens  240
    DGSTCRLVRGQYKSQLSATKSDDTVVAIPYGSRHIRLVLKGPDHLYLETKTLQGTKGENS
241 lsstgtflvdnssvdfqkfpdkeilrmagpltadfivkirnsgsadstvqfifyqpiihr  300
    LSSTGTFLVDNSSVDFQKFPDKEILRMAGPLTADFIVKIRNSGSADSTVQFIFYQPIIHR
301 wretdffpcsatcgggyqltsaecydlrsnrvvadqychyypenikpkpklqecnldpcp  360
    WRETDFFPCSATCGGGYQLTSAECYDLRSNRVVADQYCHYYPENIKPKPKLQECNLDPCP
361 asdgykqimpydlyhplprweatpwtacssscgggiqsravscveediqghvtsveewkc  420
    ASDGYKQIMPYDLYHPLPRWEATPWTACSSSCGGGIQSRAVSCVEEDIQGHVTSVEEWKC
421 mytpkmpiaqpcnifdcpkwlaqewspctvtcgqglryrvvlcidhrgmhtggcspktkp  480
    MYTPKMPIAQPCNIFDCPKWLAQEWSPCTVTCGQGLRYRVVLCIDHRGMHTGGCSPKTKP
481 hikeecivptpcykpkeklpveaklpwfkqaqeleegaavseepsfipeawsactvtcgv  540
    HIKEECIVPTPCYKPKEKLPVEAKLPWFKQAQELEEGAAVSEEPSFIPEAWSACTVTCGV
```

Figure 3.3: GPM database search results for HEK293T media samples overexpressing ADAMTSL1 TSR1-4. A. Results for sample not incubated in 6AF do not indicate any matches to the overexpressed ADAMTSL1 protein fragment. **B.** Results for sample incubated in 200 µM 6AF. The search result indicates a match for ADAMTS-like 1 in the tested sample as the main hit. **C.** The amino acid sequence for ADAMTSL1 TSR1-4 (amino acids 33-584) with six peptides (highlighted in red) that were detected in the GPM database search.
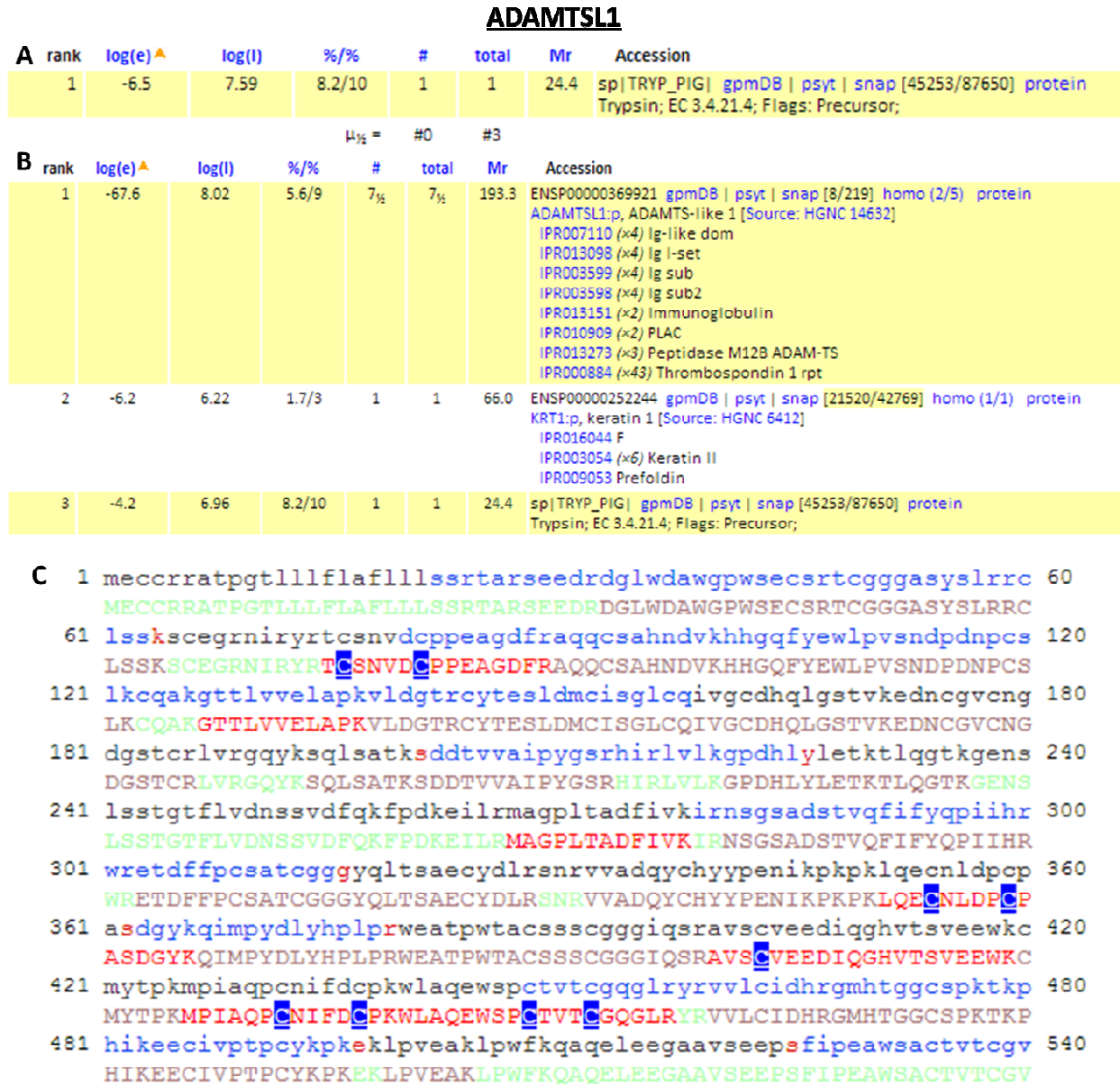
3.4. Detection of endogenously expressed *O*-fucosylated proteins

Encouraged by the proof-of-concept experiments, the next logical step was to analyze the cell lysates of HEK293T cells, which endogenously express many proteins which are labeled with 6AF (Figure 2.2A), to determine whether any protein targets of *O*-fucosylation could be selectively detected in these cells after 6AF labeling.

A representative database search result of samples generated from HEK293T cell lysates not incubated in 6AF using the "on-bead digestion" proteomics method #2 (Section 3.9) shows, as would be expected, that no predicted *O*-fucosylated proteins are detected (Figure 3.4A). Unfortunately, a representative database search result from an equivalent sample that is incubated in 200 µM 6AF is similarly devoid of any protein that is predicted to be *O*-fucosylated (Figure 3.4B). The experiment was carried out on numerous occasions with slight variations to the number or volume of washes in the proteomics protocol, the type and volume of streptavidin beads used, and other such optimizations, but all such experiments have generated results very similar to the representative figures used.

Endogenous HEK293T cell lysate

	rank	log(e)▲	log(I)	%/%	#	total	Mr	Accession
A	1	-2.2	5.05	4.0/5	1½	1	42.0	ENSP00000355645 gpmDB | psyt | snap [13805/26726] homo (26/26) protein ACTA1:p, actin, alpha 1, skeletal muscle [Source: HGNC 129] IPR004000 *(x8)* Actin-related
	2	-1.3	6.56	3.0/4	1	2½	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [43451/84218] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
				μ½ =	#1	#2		
B	rank	log(e)▲	log(I)	%/%	#	total	Mr	Accession
	1	-10.8	5.22	3.1/5	2½	2½	66.0	ENSP00000252244 gpmDB | psyt | snap [20481/40703] homo (1/5) protein KRT1:p, keratin 1 [Source: HGNC 6412] IPR016044 F IPR003054 *(x6)* Keratin II IPR009053 Prefoldin
	2	-1.7	5.95	3.0/4	1	1	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [43451/84218] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
				μ½ =	#1	#1		

Figure 3.4: Representative GPM database search results for HEK293T cell lysate samples. A. Search result for untransfected HEK293T cell lysate, not incubated with 6AF. **B.** Search result for untransfected HEK293T cell lysate incubated with 200 µM 6AF. No *O*-fucosylated proteins are detected.

3.5. "Off-bead" proteomics protocol

Undiscouraged by the inability to purify and detect endogenously expressed *O*-fucosylated proteins, I next focused my efforts on improving the yield of the proteomics protocol. Changes in the amount and type of streptavidin beads used, as well as the amount and volume of washes, had little or no impact on GPM database search results, so attention turned to the "on-bead" digestion step. Backed by informal discussions at conferences and lectures, we hypothesized that protease digestion of captured proteins on streptavidin beads may be inefficient and responsible for sample loss. Recognizing that the sample loss might be acceptable where cells were intentionally overexpressed with a protein target for *O*-fucosylation but perhaps catastrophic in attempting to identify the far lower amounts of endogenously expressed *O*-fucosylated protein, we set out on creating an "off-bead" proteomics protocol, using strategies where the captured proteins were first eluted or disassociated from the streptavidin beads, before being subjected to protease digestion. Two methods we attempted were the "thiol-cleavable linker" and "acid-cleavable linker" elution strategies.

The thiol-cleavable linker strategy involved using an azido-biotin with an S-S linker between the azide
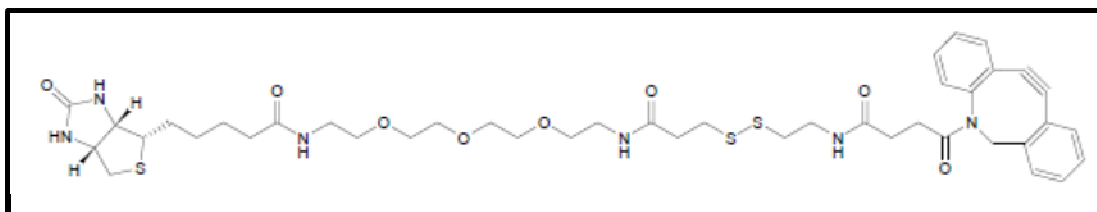
Figure 3.5: Structure of thiol-cleavable azido-biotin.

and biotin groups. After the click reaction targeting 6AF-containing proteins with the modified azido-biotin (Figure 3.5), streptavidin bead incubation, and extensive washes, the protein is then eluted from the streptavidin beads by cleaving the linker by reducing the S-S to -SH HS- with

TCEP. The eluted protein can then be subjected to protease digestion and mass spectrometry.

Unfortunately, as the representative database searches for the thiol-cleavable linker strategy

demonstrated, no relevant proteins were detected in either the -/+6AF HEK293T cell lysate

samples (Figure 3.6A and 3.6B).

Endogenous HEK293T cell lysate – "Thiol Cleavable" Proteomics Strategy

A.

rank	log(e) ▲	log(I)	%/%	#	total	Mr	Accession
1	-1.4	5.00	3.0/5	1	1	55.2	ENSP00000358880 gpmDB | psyt | snap [20/28] homo (1/1) protein AMIGO1:p, adhesion molecule with Ig-like domain 1 [Source: HGNC 20824] IPR000483 Cys-rich flank reg C IPR007110 Ig-like IPR013098 Ig I-set IPR013106 Ig V-set IPR003599 Ig sub IPR001611 (x3) Leu-rich rpt IPR003591 (x5) Leu-rich rpt typical-subtyp
2	-1.4	6.16	3.0/4	1½	1½	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [43451/84218] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
3	-1.1	5.24	4.5/10	1	1	40.1	ENSP00000304236 gpmDB | psyt | snap [1416/2615] homo (2/2) protein CD14 molecule [Source: HGNC 1628] IPR001611 Leu-rich rpt IPR016337 Monocyte diff Ag CD14
			μ½ =	#2	#2		

B.

rank	log(e) ▲	log(I)	%/%	#	total	Mr	Accession
1	-1.6	6.03	3.0/4	1½	1½	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [43451/84218] protein Trypsin; EC 3.4.21.4; Flags: Precursor;
			μ½ =	#1	#1		

Figure 3.6: Representative GPM database search results for HEK293T lysate sample using the "thiol-cleavable" strategy proteomics protocol. A. Search result for HEK293T cell lysate sample not incubated in 6AF. **B.** Search result for HEK293T cell lysate sample incubated in 200 μM 6AF. No *O*-fucosylated protein targets are detected.

Part of the difficulty in using a thiol-cleavable linker is that our proteomics method

requires the extensive reduction and alkylation of samples prior to protease digestion,

particularly as the known targets of fucosylation, EGF repeats and TSRs, are small but

extensively folded domains that are unlikely to be significantly digested without efficient

reduction and alkylation. We were careful in the protocol to avoid adding any reducing agent

prematurely, but nevertheless it was difficult to detect whether the thiol-cleavable probe was

incorporated at all, because Western blots under reducing conditions would cause cleavage of the

thiol-cleavable probe and Western blots under non-reducing conditions for cell lysate samples

were notoriously of poor quality and inconclusive. As such, an alternative method for eluting the

captured protein through an acid-cleavable probe (Figure 3.7) was pursued. This would have the

added benefit of being able to track capture of the *O*-fucosylated protein on streptavidin beads by Western blot, prior to protease digestion and mass spectrometry workup.
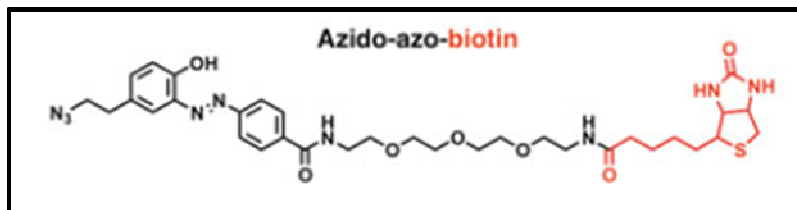
Figure 3.7: Azido-azo-biotin, the "acid-cleavable linker"

In pursuing the acid-cleavable linker elution strategy, we decided to first test the incorporation of the acid-cleavable linker by using media from HEK293T cells overexpressing hTSP TSR1-3 (Figure 3.8A). The left lane represents sample from cells not incubated with 6AF, while the right lane represents sample incubated in 50 µM 6AF. Both samples were "clicked" with the acid-cleavable linker azido-biotin. The myc tag expressed on the hTSP TSR1-3 construct allows identification of the overexpressed protein, which is present in both samples (red), while a fluorescent streptavidin antibody (green) detects biotin, which allows for direct detection of the acid-cleavable linker. As Figure 3.8A demonstrates, the acid-cleavable linker azido-biotin is successfully incorporated in the media sample obtained from cells treated with 6AF.

Next, we wanted to test our ability to elute biotinylated 6AF-modified proteins (hTSP TSR1-3) captured on streptavidin beads, using 50 mM sodium dithionite. As Figure 3.8B demonstrates, mock elution with water is not successful in eluting the hTSP TSR1-3, which remains almost exclusively bound to the streptavidin beads. By contrast, eluting with 50 mM sodium dithionite results in virtually complete cleavage of the linker (loss of green signal) and with the majority of the hTSP TSR 1-3 protein eluted from the streptavidin beads into solution.
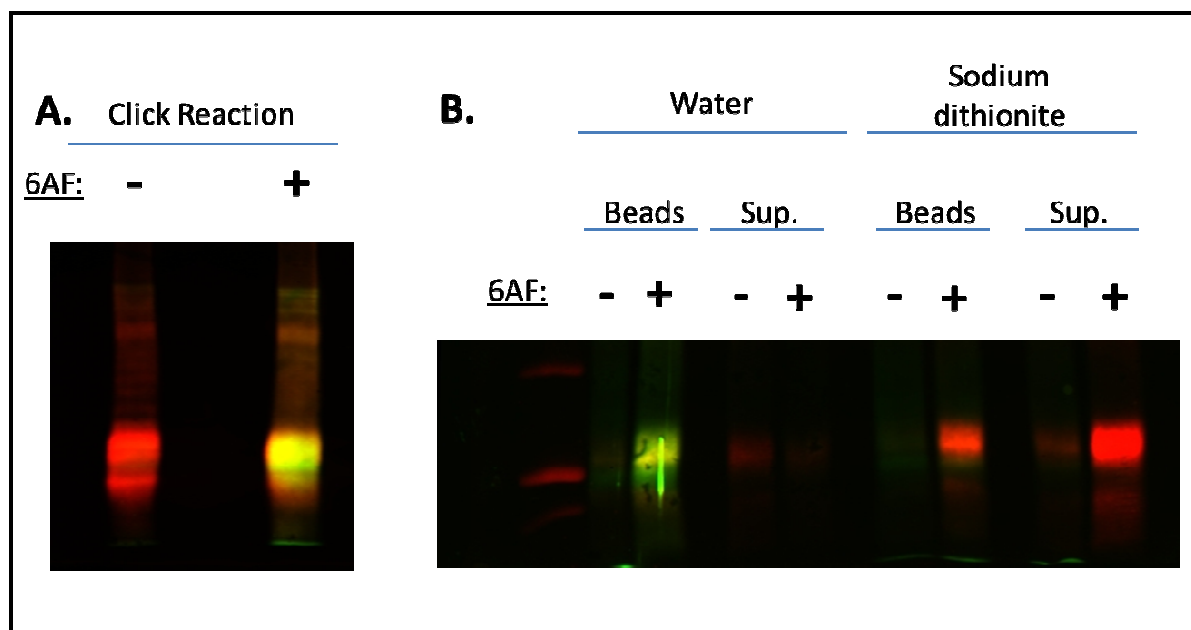
Figure 3.8: Incorporation and elution of acid-cleavable azido-biotin in media samples of HEK293T cells overexpressing hTSP1 TSR1-3, -/+6AF. **A.** Acid-cleavable azido-biotin is successfully and exclusively incorporated onto hTSP TSR1-3 from cells treated with 50 μM 6AF. **B.** Mock cleavage of acid-cleavable azido-biotin with water does not cleave the linker, resulting in the vast majority of hTSP1 TSR1-3 being retained on the streptavidin beads. Cleavage with 50 mM of sodium dithionite, however, results in complete cleavage of the linker, as evidenced by the absence of green signal. Moreover, the majority of the hTSP TSR1-3 protein is eluted from the streptavidin beads into the supernatant. Red: anti-myc antibody used to detect hTSP TSR1-3; Green: IRDye800 Streptavidin probe used to detect biotin; Yellow: Overlay of red and yellow signals.

Having demonstrated the incorporation of the acid-cleavable azido-biotin and its cleavage with sodium dithionite resulting in 6AF-modified proteins being eluted from streptavidin beads into the supernatant, we tested this strategy to do a "proof-of-concept" experiment, attempting to detect by GPM search an overexpressed protein (i.e. hTSP1 TSR1-3); a feat we were able to accomplish with the "on-bead" digestion method. Unfortunately, the GPM search on HEK293T media transfected with hTSP1 TSR1-3 and incubated with 6AF failed to detect any of the overexpressed protein following the proteomics protocol (Figure 3.9). Absolutely no hits were detected for the control sample that was not incubated with 6AF. As the proof-of-concept

experiment failed, an attempt to use this strategy for finding endogenous proteins was abandoned.

rank	log(e) ▲	log(I)	%/%	#	total	Mr	Accession
1	-1.8	5.52	1.1/2	1	1	181.5	ENSP00000300896 gpmDB | psyt | snap [915/971] homo (3/3) protein USP32:p, ubiquitin specific peptidase 32 [Source: HGNC 19143] IPR018249 EF HAND 2 IPR002048 EF hand Ca-bd IPR006615 Pept C19 DUSP IPR001394 Peptidase C19 IPR001125 Recoverin
2	-1.6	5.73	0.7/1	1	1	154.7	ENSP00000303928 gpmDB | psyt | snap [98/122] homo (1/1) protein KIAA0232:p, No description Annotated domains: IPR001687 ATP/GTP-binding site motif A (P-loop)
3	-1.6	7.02	3.0/4	1½	1½	24.4	sp|TRYP_PIG| gpmDB | psyt | snap [98803/101484] protein Trypsin; EC 3.4.21.4; Flags: Precursor;

Figure 3.9: Representative GPM database search result for HEK293T media sample overexpressing hTSP1 TSR1-3 incubated in 6AF using the "acid-cleavable linker" proteomics strategy. The search result does not show any hits for the overexpressed protein. The control sample from HEK293T media sample overexpressing hTSP1 TSR1-3 but not incubated in 6AF showed no hits whatsoever.

3.6. Collaboration with Bertozzi lab – "IsoTag" proteomics method

During the visit of Dr. Caroline Bertozzi, a pioneer of "click" chemistry and whose lab has published many papers on developing chemical and computational tools for "mining" the glycoproteome, on May 1st, 2014 to Stony Brook University, I was able to speak with Dr. Bertozzi briefly on the status of the proteomics project. This quickly led to a collaborative project with Christina Woo, a graduate student in her laboratory, in order to utilize the recently developed "IsoTag" to probe the *O*-fucose proteome. The IsoTag (Figure 3.10A) consists of four functional parts: An "affinity enrichment" component utilizing biotin; a "cleavable functionality" that allows cleavage of the probe when reacted with the appropriate reagents; an "isotope pattern" incorporating two bromine atoms; and a "bioorthogonal azide reactive group", allowing the probe to engage in the "click" reaction with the alkyne group that is in the 6AF used to metabolically treat cells. The innovation in the IsoTag comes from the "isotope pattern" component, as bromine naturally exists in two nearly equal stable isotopes as Br-79 and Br-81. The presence of two bromine atoms thus naturally confers a unique charge distribution pattern (Figure 3.10B) in a mass spectrometric analysis which would only be detected in proteins that incorporate the IsoTag. A computer program compiles all instances of the unique bromine-derived pattern in a mass spectrometric analysis and a subsequent run of the same sample can exclusively select those ions for fragmentation and subsequent identification by comparing them to the GPM database (or equivalent). This would allow for an increase in the sensitivity of mass spectrometric analysis, as typically only abundant ions in a mass spectrometric run are chosen for further fragmentation, but by being able to track and then exclusively fragment ions with the unique bromine-derived pattern, even low abundance proteins could, in theory, be fragmented and subsequently identified.

Figure 3.10: IsoTag and resulting unique bromine-derived mass spectrometric pattern. A. Structure of the IsoTag, consisting of four distinct components, including the "isotope pattern" component. **B.** An example of IsoTag labeling of a known glycopeptide. The predicted bromine-derived mass spectrometric pattern (bottom panel – "Isotope Simulation") is closely matched by the experimentally produced spectra.

NOTE: This data was generated by Christina Woo of the Bertozzi lab

Similar to the approach I adopted in my previously described proteomics methods, we elected to begin with "proof-of-concept" experiments testing the utility of the IsoTag-based Bertozzi lab proteomics protocol on cell lysates from HEK293T cells overexpressing hTSP1 TSR1-3. I therefore generated and shipped HEK293T cell pellets, from cells transfected with hTSP1 TSR1-3 and incubated in 100 µM 6AF, as well as untransfected controls and controls where no 6AF was added. Ms. Woo lysed one plate's worth of cells, performed the click reaction with the IsoTag, and performed anti-biotin Western blots to check incorporation of the IsoTag.

Unfortunately, the Western blot revealed no additional signal in the +6AF treated cell lysate samples as compared to the -6AF treated cell lysates (Figure 3.11A). There was no indication that the IsoTag had failed, since an anti-biotin Western blot on cells generated in the Bertozzi lab labeled with an alkyne-tagged sialic acid precursor showed robust labeling after click reaction with the IsoTag (Figure 3.11B). There was also no indication of any problems with the HEK293T cells, as a Western blot I performed on 5% of the HEK293T cell media (250 µL) showed robust labeling of the anti-myc antibody in HEK293T cells transfected with hTSP TSR1-3 and robust labeling with streptavidin in cells incubated in 100 µM 6AF (Figure 3.11C). Despite these setbacks, Ms. Woo performed the Bertozzi lab proteomics protocol using the IsoTag on the HEK293T cell lysates transfected with hTSP1 TSR1-3 and incubated in 6AF, but no spectra were generated that correlated with the overexpressed hTSP1 TSR1-3 protein.
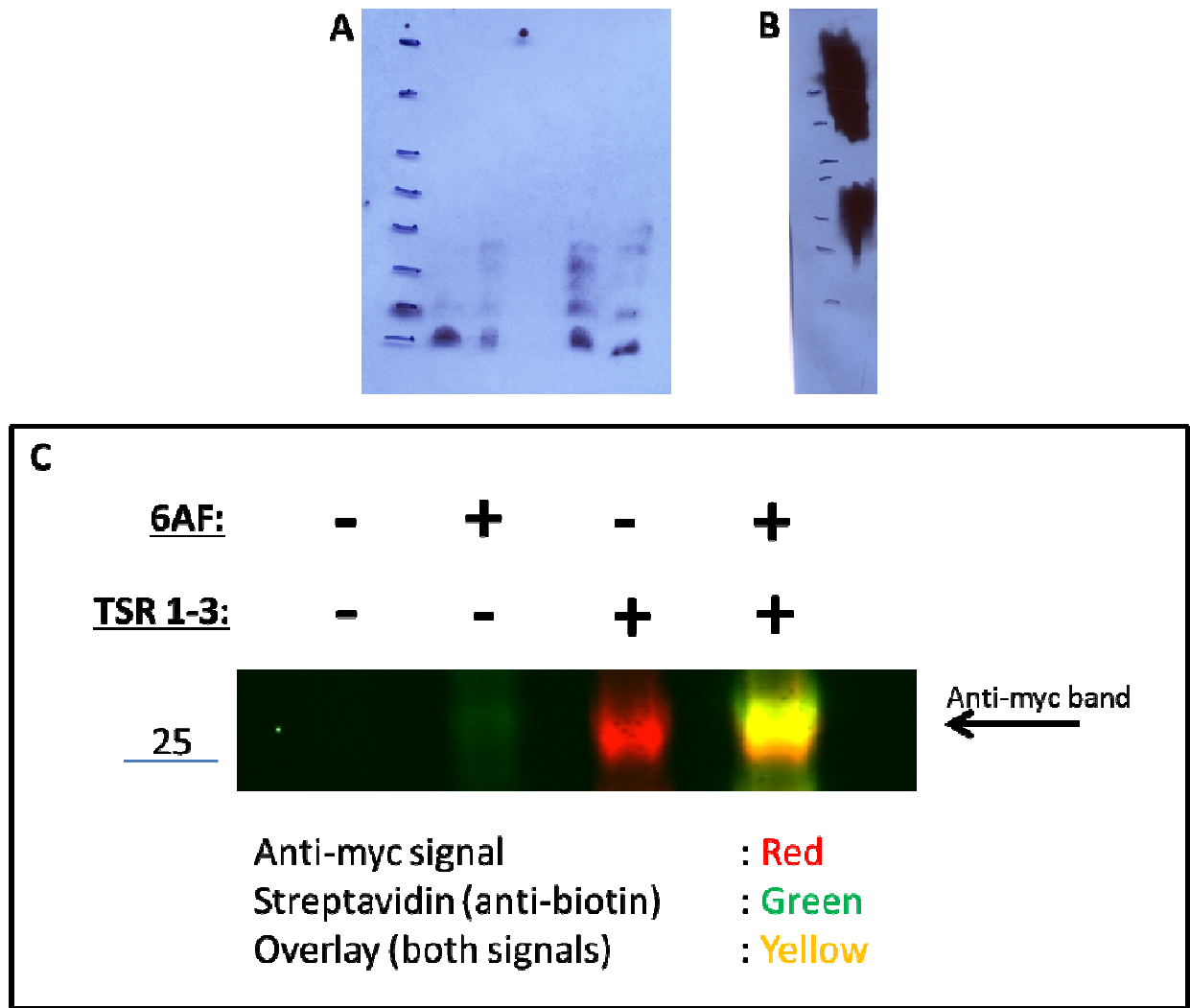
Figure 3.11. Western blots to check IsoTag labeling. A. One minute exposure of four samples using anti-biotin antibody, from left to right: HEK293T cell lysates transfected with hTSP1 TSR1-3 and incubated in 6AF; HEK293T cell lysates transfected with hTSP1 TSR1-3 and not incubated in 6AF; untransfected HEK293T cell lysates incubated in 6AF; untransfected HEK293T cell lysates not incubated in 6AF. **B.** Fifteen second exposure using anti-biotin antibody of Bertozzi lab generated cell lysate incubated in ManN-alkyne, a sugar analog of a sialic acid precursor. **C.** Western blot on HEK293T cell media.

NOTE: Figure 3.11A and 3.11B was generated by Christina Woo

3.7. Conclusions

The overall strategy of labeling cells with 6AF, performing click chemistry to biotinylate fucosylated proteins, pulling down those proteins with streptavidin beads, followed by protease digestion and mass spectrometric analysis, yielded mixed results. The "proof-of-concept" experiments on overexpressed mN1 EGF1-5, hTSP1 TSR1-3, and ADAMTSL1 TSR1-4 demonstrated that the chemistry underlying the "on-bead" digestion protocol is sound, and in fact these overexpressed proteins were successfully pulled down and identified by mass spectrometry, exclusively in samples incubated with 6AF and never in the non-6AF treated controls. Using "off-bead" proteomics strategies involving the dissociation of biotinylated proteins from streptavidin beads were unsuccessful. Additionally, all attempts to pull down and identify endogenous targets of O-fucosylation in HEK293T cells failed.

3.8. Future Directions

Using fucose analogs to identify the *O*-fucose proteome presents enormous challenges. The success of the "on-bead" digestion proteomics protocols in pulling down overexpressed proteins confirms that the underlying approach is sound from a chemical basis (i.e. it is possible to incorporate fucose analogs into cells, biotinylate them with the click reaction, and then purify them using streptavidin beads), however the major challenge is in addressing the low abundance of *O*-fucosylated proteins in the cells, which poses the overriding difficulty in purifying and identifying this small population of proteins in the total cell lysate.

Much time has been spent in this project attempting to optimize a variety of conditions in the proteomics protocol, so it is unlikely that further minor optimizations will yield success. Major breakthroughs, including new reagents being developed by Dr. Peng Wu's laboratory that enable "supersensitive" click chemistry[112] or extremely sensitive mass spectrometers capable of resolving the many thousands of cell lysate proteins coupled with advanced algorithms that enable analysis of the massive amounts of spectra generated to pinpoint *O*-fucose modified proteins, might be needed to make significant new progress.

Additionally, the recent availability of new enrichment tools, such as azide-coupled magnetic beads (available on clickchemistrytools.com) allows the prospect of using a biotin/streptavidin-free method of purifying and capturing proteins that incorporate 6AF, potentially allowing for efficient enrichment of 6AF-incorporating proteins without requiring extensive stringent washes as in the current protocol, which may improve the chances of retaining sufficient sample to allow for mass spectral identification. Certainly, it would be a logical first step in the continuation of this project.

3.9. Materials and methods

"On-bead" digestion proteomics method #1

Stock Solutions (sources for unusual reagents are indicated):

- 25 mM Azido-biotin (ClickChemistryTools.com, Item #AZ104-25)

- 25 mM TBTA (ClickChemistryTools.com, Item #1061-100)

- 250 mM copper sulfate (Sigma-Aldrich, Catalog #451657)

- 10% SDS in water

- 100% β-mercaptoethanol

- 400 mM tris-HCl pH = 8.0

- 100% NP-40

- 100% Triton X-100 (substitute with NP-40 if unavailable)

- 100% formic acid

- 10X PBS

- 100 mM EDTA

- 1 M NaCl

- 100 mM $CaCl_2$

- 0.5 M TCEP (Thermo Scientific, Product #77720)

Freshly prepared solutions:

- 500 mM sodium ascorbate (MW = 198.11 g/mol)

- 6M Urea (MW = 60.06 g/mol) in 1X PBS

- 2M Urea in 1X PBS

- 6M Urea in 1X PBS + 10 mM TCEP

- Buffer A: 1% SDS + 1% β-mercaptoethanol in water

- Buffer B: 0.7% NP-40 in 50 mM tris-HCl (pH = 8.0) + 1 pic tab / 10 mL

- Lysis Buffer: 5 mM EDTA; 150 mM NaCl; 1% Triton X-100 in 50 mM tris-HCl (pH = 8.0) + 1 protease inhibitor cocktail (pic) tab / 10 mL of lysis buffer

- RIPA Buffer: 25mM Tris-HCl (pH 7.6), 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS

- Trypsin Solution: 1 mM $CaCl_2$ + 10 µg/mL trypsin in 2M Urea in 1X PBS

Procedure:

1. Perform Click Reaction on 250 uL of media/cell lysate sample. (For larger samples, increase proportionally):

    a. Add 1 µL of 25 mM Azido-biotin. Vortex.

    b. Add 1 µL of 25 mM TBTA. Vortex. (*Note:* Solution will be slightly cloudy due to low solubility of TBTA).

    c. Add 1 µL of 500 mM sodium ascorbate. Vortex.

    d. Add 1 µL of 250 mM copper sulfate. Vortex.

    e. Allow reaction to proceed at room temperature for 1 hour.

2. Acetone precipitation:

    a. Add 1000 µL of -20 °C acetone to each sample. Incubate in -20 °C freezer for 2 hours to overnight.

    b. After incubation, spin down sample for 10 minutes at maximum speed in 4 °C centrifuge.

    c. Aspirate acetone, taking care not to disturb protein pellet.

3. Applying sample to streptavidin beads:

a. Wash 50 µL packed volume (100 µL 50% slurry) of streptavidin beads per sample with lysis buffer. (Example: For 8 samples, wash 400 µL of streptavidin beads.)

b. Resuspend protein pellet in 50 µL of 1% SDS with aggressive vortexing.

c. Sonicate for 10 min and then boil for 10 min.

d. Cool down, centrifuge sample for 1 min at maximum speed, then dilute with 450 µL of Lysis buffer for a total volume of 500 µL.

e. Apply sample solution to 50 µL (packed volume) of streptavidin beads.

f. Apply samples to end-over-end rotator in cold room. Incubate for 2 hours to overnight.

4. Wash the beads (all washes at 1 mL):

a. Wash 3 times with Lysis Buffer. After each wash, centrifuge the beads for 2.5 minutes at 2500 rpm.

b. Wash 1 time with 6M Urea in 1X PBS.

c. Wash 1 time with 1X PBS.

d. Aspirate the beads with a 1 mL syringe to withdraw all solution from beads.

5. On-bead reduction and alkylation:

a. Add 200 µL of "6M Urea in 1X PBS + 10 mM TCEP" to the beads.

b. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 15 minutes.

c. Add 50 µL of 200 mM Iodoacetamide (Final concentration: 20 mM Iodoacetamide) to the beads.

d. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 30 minutes. Turn off the lights and cover sample tray in aluminum foil.

6. On-bead trypsin digestion:

    a. Wash the beads with 2M Urea in 1X PBS.

    b. Aspirate the beads with 1 mL syringe.

    c. Add 150 µL of trypsin solution.

    d. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 8 hours.

7. Sample collection:

    a. Centrifuge beads for 2.5 minutes at 2500 rpm and collect supernatant in a "Lo-bind" tube.

    b. Wash the beads with 100 µL of HPLC water. Centrifuge the beads and add supernatant to the same "Lo-bind" tube.

    c. Add 13 µL of 100% formic acid (Final concentration: 5% formic acid) to the solution.

8. Perform Zip-Tip clean-up.

    a. Follow Zip-Tip protocol. Briefly, the Zip-Tip is activated with 2 washes of Solution A (95% acetonitrile, 0.1% acetate), followed by equilibration with 3 washes of Solution B (0.1% acetate). Sample is then pipetted to fresh tube. Sample is then pipetted from fresh tube back to original tube. Zip-Tip is then washed 6 times with Solution B. Sample is eluted from Zip-Tip by pipetting 10 times into 3 µL of Solution C (50% acetonitrile, 0.1% acetate.) Eluted sample is then readjusted to 15% HPLC Buffer B by adding 7 µL of Zip-Tip Solution B (0.1% acetate), yielding 10 µL of final sample for MS analysis.

9. Submit samples for MS analysis.

Figure 3.12. **Summary of steps required for "on-bead" trypsin digestion proteomics methods.**

"On-bead" digestion proteomics protocol #2

All steps are the same as proteomics protocol #1, except:

-Pay special attention to conducting all steps of sample separation in a well-ventilated clean

hood.

72

-Replace #4 (washing the beads) in the procedure with:

4. Wash the beads (all washes at 10 mL):

   a. Wash 4 times with RIPA Buffer. After each wash, centrifuge the beads for 2.5 minutes at 2500 rpm.

   b. Wash 3 times with 6M Urea in 1X PBS.

   c. Wash 3 times with 1X PBS.

   d. Carefully remove the top 9-9.5 mL of 1X PBS with a pipette-aid.

   e. Aspirate the beads with a 1 mL syringe to withdraw all solution from beads.

"Off-bead" thiol-cleavable linker proteomics protocol

Stock Solutions:

- 25 mM Azido-biotin

- 25 mM TBTA

- 250 mM copper sulfate

- 10% SDS in water

- 100% 2-mercaptoethanol

- 400 mM tris-HCl pH = 8.0

- 100% NP-40

- 100% Triton X-100 (substitute with NP-40 if unavailable)

- 100% formic acid

- 10X PBS

- 100 mM EDTA

- 1 M NaCl

- 100 mM CaCl$_2$

- 0.5 M TCEP

Freshly prepared solutions:

- 500 mM sodium ascorbate (MW = 198.11 g/mol)

- 6M Urea (MW = 60.06 g/mol) in 1X PBS

- 2M Urea in 1X PBS

- 6M Urea in 1X PBS + 10 mM TCEP

- Buffer A: 1% SDS + 1% 2-mercaptoethanol in water

- Buffer B: 0.7% NP-40 in 50 mM tris-HCl (pH = 8.0) + 1 pic tab / 10 mL

- Lysis Buffer: 5 mM EDTA; 150 mM NaCl; 1% Triton X-100 in 50 mM tris-HCl (pH = 8.0) + 1 pic tab

- RIPA Buffer: 25mM Tris-HCl (pH 7.6), 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS

- Trypsin Solution: 1 mM CaCl$_2$ + 10 µg/mL trypsin in 2M Urea in 1X PBS

Procedure:

1. Perform Click Reaction on 250 uL of media/cell lysate sample. (For larger samples, increase proportionally):

    a. Add 1 µL of 25 mM thiol-cleavable azido-biotin. Vortex.

    b. Add 1 µL of 25 mM TBTA. Vortex. (*Note:* Solution will be slightly cloudy due to low solubility of TBTA).

  c. Add 1 µL of 500 mM sodium ascorbate. Vortex.

  d. Add 1 µL of 250 mM copper sulfate. Vortex.

  e. Allow reaction to proceed at room temperature for 1 hour.

2. Acetone precipitation:

  a. Add 1000 µL of -20 °C acetone to each sample. Incubate in -20 °C freezer for 2 hours to overnight.

  b. After incubation, spin down sample for 10 minutes at maximum speed in 4 °C centrifuge.

  c. Aspirate acetone, taking care not to disturb protein pellet.

3. Applying sample to streptavidin beads:

  a. Wash 50 µL packed volume (100 µL 50% slurry) of streptavidin beads per sample with lysis buffer. (Example: For 8 samples, wash 400 µL of streptavidin beads.)

  b. Resuspend protein pellet in 50 µL of 1% SDS with aggressive vortexing.

  c. Sonicate for 10 min and then boil for 10 min.

  d. Cool down, centrifuge sample for 1 min at maximum speed, then dilute with 450 µL of Lysis buffer for a total volume of 500 µL.

  e. Apply sample solution to 50 µL (packed volume) of streptavidin beads.

  f. Apply samples to end-over-end rotator in cold room. Incubate for 2 hours to overnight.

4. Wash the beads (all washes at 10 mL):

  a. Wash 4 times with RIPA Buffer. After each wash, centrifuge the beads for 2.5 minutes at 2500 rpm.

b. Wash 3 times with 6M Urea in 1X PBS.

c. Wash 3 times with 1X PBS.

d. Carefully remove the top 9-9.5 mL of 1X PBS with a pipette-aid.

e. Transfer the beads to "Lo-bind" tubes.

f. Aspirate the beads with a 1 mL syringe to withdraw all solution from beads.

4. Elution by thiol-cleavage (reduction):

a. Add 200 µL of "6M Urea in 1X PBS + 10 mM TCEP" to the beads.

b. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 15 minutes.

c. Collect supernatant in "Lo-bind" tube. Then, add an additional 200 µL of "6M

Urea in 1X PBS + 10 mM TCEP" to the beads.

d. For both the supernatant and the beads: Add 50 µL of 200 mM Iodoacetamide

(Final concentration: 20 mM Iodoacetamide) to the beads.

e. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 30 minutes.

Turn off the lights and cover sample tray in aluminum foil.

f. Combine both samples.

5. Trypsin digestion:

a. Add 150 µL of trypsin solution.

b. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 8 hours.

6. Sample collection:

a. Centrifuge tubes and add 25 µL of 100% formic acid (Final concentration: 5%

formic acid) to the solution.

7. Perform Zip-Tip clean-up.

a. Follow Zip-Tip protocol. Briefly, the Zip-Tip is activated with 2 washes of Solution A (95% acetonitrile, 0.1% acetate), followed by equilibration with 3 washes of Solution B (0.1% acetate). Sample is then pipetted to fresh tube. Sample is then pipetted from fresh tube back to original tube. Zip-Tip is then washed 6 times with Solution B. Sample is eluted from Zip-Tip by pipetting 10 times into 3 µL of Solution C (50% acetonitrile, 0.1% acetate.) Eluted sample is then readjusted to 15% HPLC Buffer B by adding 7 µL of Zip-Tip Solution B (0.1% acetate), yielding 10 µL of final sample for MS analysis.

8. Submit samples for MS analysis.
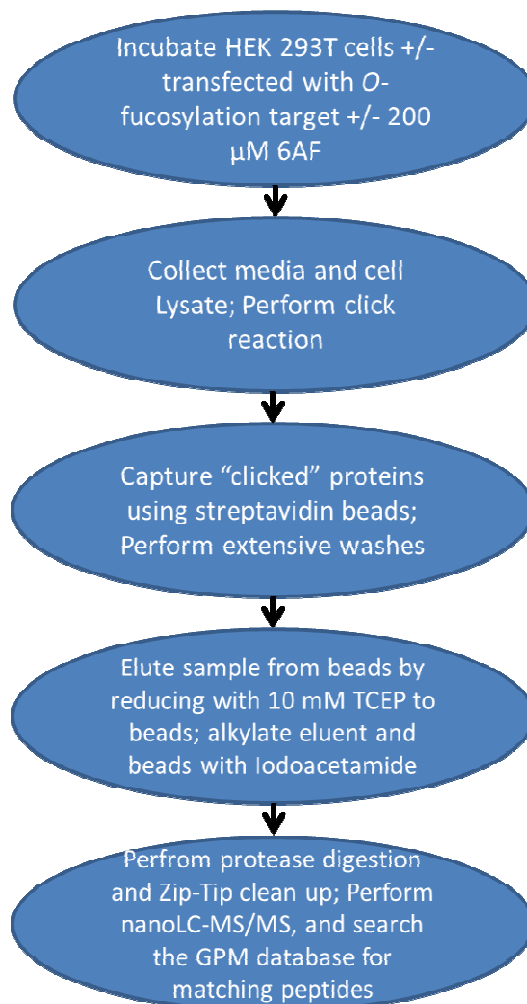
Figure 3.13. Summary of steps required for "off-bead" thiol-cleavable linker proteomics strategy.

"Off-bead" acid-cleavable linker proteomics protocol

Stock Solutions:

- 25 mM Azido-biotin

- 25 mM TBTA

- 250 mM copper sulfate

- 10% SDS in water

- 100% 2-mercaptoethanol

- 400 mM tris-HCl pH = 8.0

- 100% NP-40

- 100% Triton X-100 (substitute with NP-40 if unavailable)

- 100% formic acid

- 10X PBS

- 100 mM EDTA

- 1 M NaCl

- 100 mM CaCl$_2$

- 0.5 M TCEP

Freshly prepared solutions:

- 500 mM sodium ascorbate (MW = 198.11 g/mol)

- 6M Urea (MW = 60.06 g/mol) in 1X PBS

- 2M Urea in 1X PBS

- 6M Urea in 1X PBS + 10 mM TCEP

- Buffer A: 1% SDS + 1% 2-mercaptoethanol in water

- Buffer B: 0.7% NP-40 in 50 mM tris-HCl (pH = 8.0) + 1 pic tab / 10 mL

- Lysis Buffer: 5 mM EDTA; 150 mM NaCl; 1% Triton X-100 in 50 mM tris-HCl (pH = 8.0) + 1 pic tab

- RIPA Buffer: 25mM Tris-HCl (pH 7.6), 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS

- Trypsin Solution: 1 mM $CaCl_2$ + 10 µg/mL trypsin in 2M Urea in 1X PBS

Procedure:

1. Perform Click Reaction on 250 uL of media/cell lysate sample. (For larger samples, increase proportionally):

    a. Add 1 µL of 25 mM acid-cleavable azido-biotin. Vortex.

    b. Add 1 µL of 25 mM TBTA. Vortex. (*Note:* Solution will be slightly cloudy due to low solubility of TBTA).

    c. Add 1 µL of 500 mM sodium ascorbate. Vortex.

    d. Add 1 µL of 250 mM copper sulfate. Vortex.

    e. Allow reaction to proceed at room temperature for 1 hour.

2. Acetone precipitation:

    a. Add 1000 µL of -20 °C acetone to each sample. Incubate in -20 °C freezer for 2 hours to overnight.

    b. After incubation, spin down sample for 10 minutes at maximum speed in 4 °C centrifuge.

    c. Aspirate acetone, taking care not to disturb protein pellet.

3. Applying sample to streptavidin beads:

a. Wash 50 µL packed volume (100 µL 50% slurry) of streptavidin beads per sample with lysis buffer. (Example: For 8 samples, wash 400 µL of streptavidin beads.)

b. Resuspend protein pellet in 50 µL of 1% SDS with aggressive vortexing.

c. Sonicate for 10 min and then boil for 10 min.

d. Cool down, centrifuge sample for 1 min at maximum speed, then dilute with 450 µL of Lysis buffer for a total volume of 500 µL.

e. Apply sample solution to 50 µL (packed volume) of streptavidin beads.

f. Apply samples to end-over-end rotator in cold room. Incubate for 2 hours to overnight.

5. Wash the beads (all washes at 10 mL):

a. Wash 4 times with RIPA Buffer. After each wash, centrifuge the beads for 2.5 minutes at 2500 rpm.

b. Wash 3 times with 6M Urea in 1X PBS.

c. Wash 3 times with 1X PBS.

d. Carefully remove the top 9-9.5 mL of 1X PBS with a pipette-aid.

e. Aspirate the beads with a 1 mL syringe to withdraw all solution from beads.

4. On-bead reduction and alkylation:

a. Add 200 µL of "6M Urea in 1X PBS + 10 mM TCEP" to the beads.

b. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 15 minutes.

c. Add 50 µL of 200 mM Iodoacetamide (Final concentration: 20 mM Iodoacetamide) to the beads.

d. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 30 minutes.

Turn off the lights and cover sample tray in aluminum foil.

5. Elution:

a. Wash the beads with 2M Urea in 1X PBS.

b. Transfer the beads to "Lo-Bind" tubes.

c. Aspirate the beads with 1 mL syringe.

d. Add 100 µL of 50 mM sodium dithionite and incubate at room temperature for 30 minutes.

e. Collect the supernatant into a fresh "Lo-Bind" tube.

6. Trypsin digestion:

a. Add 50 µL of trypsin solution.

b. Incubate with gentle agitation at 37 °C (in Hot Room shaker) for 8 hours.

7. Sample collection:

a. Centrifuge tubes and then add 8 µL of 100% formic acid (Final concentration: 5% formic acid) to the solution.

8. Perform Zip-Tip clean-up.

a. Follow Zip-Tip protocol. Briefly, the Zip-Tip is activated with 2 washes of Solution A (95% acetonitrile, 0.1% acetate), followed by equilibration with 3 washes of Solution B (0.1% acetate). Sample is then pipetted to fresh tube. Sample is then pipetted from fresh tube back to original tube. Zip-Tip is then washed 6 times with Solution B. Sample is eluted from Zip-Tip by pipetting 10 times into 3 µL of Solution C (50% acetonitrile, 0.1% acetate.) Eluted sample is

then readjusted to 15% HPLC Buffer B by adding 7 µL of Zip-Tip Solution B

(0.1% acetate), yielding 10 µL of final sample for MS analysis.

9. Submit samples for MS analysis.

```
┌─────────────────────────────────┐
│  Incubate HEK 293T cells +/-    │
│  transfected with O-            │
│  fucosylation target +/- 200    │
│  µM 6AF                         │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Collect media and cell          │
│  Lysate; Perform click           │
│  reaction                        │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Capture "clicked" proteins      │
│  using streptavidin beads;       │
│  Perform extensive washes        │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Perform on-bead reduction       │
│  and alkylation; elute sample by │
│  adding sodium dithionite        │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Perfrom protease digestion      │
│  and Zip-Tip clean up; Perform   │
│  nanoLC-MS/MS, and search        │
│  the GPM database for            │
│  matching peptides               │
└─────────────────────────────────┘
```

**Figure 3.14. Summary of steps required for "off-bead" acid-cleavable linker proteomics strategy.**

Chapter 4: Fucose Analogs – Substrates or Inhibitors for FUT8, POFUT1, and POFUT2?

4.1. Introduction

Following the observation that 6-alknylfucose modification of the *N*-linked glycan of Lunatic Fringe may be inefficient (Figure 2.3B), the tantalizing possibility emerged of utilizing fucose analogs as inhibitors of fucosylation, particularly the core fucosylation of the *N*-linked glycans on the single site of *N*-linked glycosylation on the Fc portion of human Immunoglobulin G (IgG) that is carried out by fucosyltransferase 8 (FUT8). One role of IgG antibodies is to bind antigen and promote antibody dependent cellular cytotoxicity (ADCC) carried out by effector cells of the immune system.[113] ADCC can be induced more potently in therapeutic antibodies that do not undergo core fucosylation on their *N*-linked asparagine.[114,115] Thus, developing inhibitors of FUT8 could be useful for generating potent anticancer therapeutic antibodies devoid of core fucosylation. This chapter demonstrates the impact of 6-alkynylfucose (6AF), as well as additional fucose analogs 5-thiofucose (5TF) and 2-fluorofucose (2FF), on the core fucosylation of *N*-glycan heavy chain of human Immunogloblin G (IgG) carried out by FUT8. I chose to investigate 5-thiofucose (5TF) because it has been demonstrated to be an inhibitor of sialyl-Lewis[x] biosynthesis,[103] while 2-fluorofucose (2FF) was selected for study because it was shown to inhibit core fucosylation by FUT8 of IgG in Chinese Hamster Ovary (CHO) cells.[53] I also performed experiments to study the impact of the 5TF and 2FF on the *O*-fucosylation of mouse Notch1 EGF 1-5 by POFUT1 and *O*-fucosylation of human Thrombospondin1 TSR 1-3 by POFUT2, to test whether these fucose analogs acted as substrates or inhibitors for the fucosyltransferases responsible for *O*-fucosylation.

4.2. Glycan standard – Man₉GlcNAc₂

The first step in the glycan analysis was to use a commercially available glycan as a standard. MAN-9 glycan (Sigma-Aldrich) was used. This oligosaccharide contains two GlcNAc and nine mannose residues ($Man_9GlcNAc_2$, molecular weight = 1883.67) and was used to test the mass spectrometer's glycan chip. Figure 4.1A illustrates the MS2 spectra of an ion matching the expected mass of the glycan standard in the 2+ charge state, with the individual peaks corresponding to the breakdown products of the glycan after undergoing collision induced dissociation. Figure 4.1B shows the Extracted Ion Chromatogram (EIC) for the $Man_9GlcNAc_2$ standard. EICs provide a semi-quantitative estimate of the relative amounts of each ion detected in the sample. In this case, we see two distinct peaks for the same mass, referring to the α-anomer and β-anomer of the glycan at the reducing end carbon, which elute separately on the glycan chip.
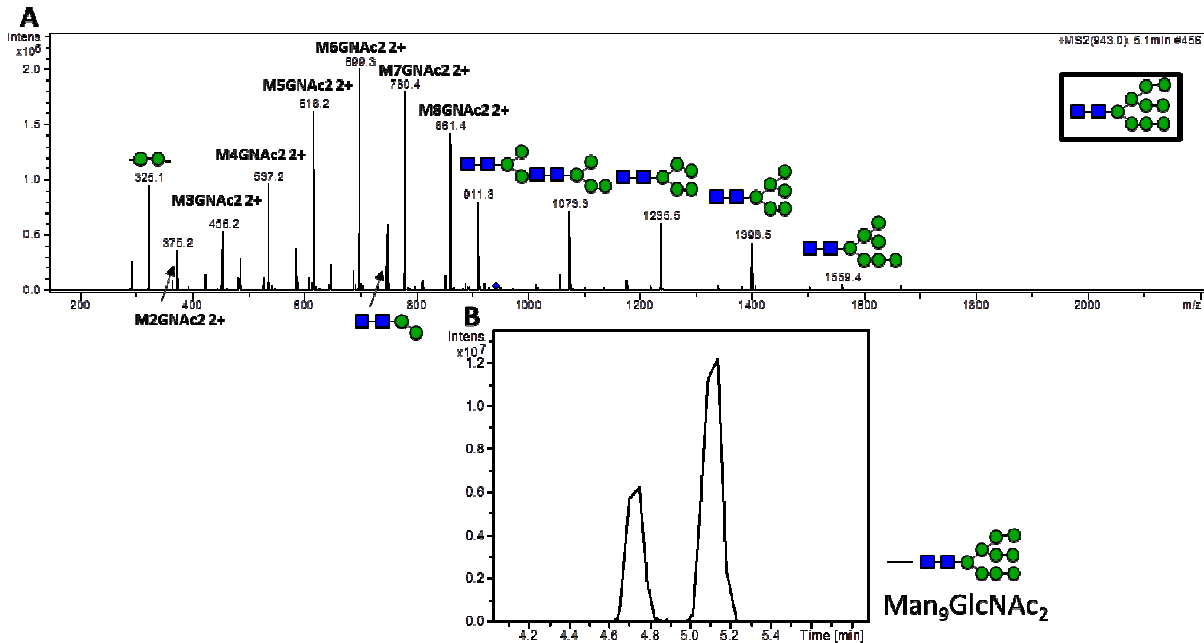
Figure 4.1: MS2 spectrum and EIC of glycan standard Man₉GlcNAc₂. A. The parent ion of 943.0 closely matches that of the glycan standard, Man9GlcNAc2 in the 2+ charge state. Numerous labeled breakdown products indicated in the MS2 spectrum further confirm

85

identification. **B.** The EIC shows two peaks for the glycan standard, likely referring to the α- and β-anomers which elute separately on the glycan chip. Blue square, GlcNAc; green circle, mannose.

4.3. The effect of fucose analogs on core fucosylation of *N*-linked glycans by FUT8 on human IgG

HEK293T cells transfected with IgG were incubated with conditioned media containing 50 µM of the fucose analogs 6AF, 5TF, 2FF, or with no added fucose analog. The samples were then processed by the "*N*-glycan purification and identification protocol" (see methods).

4.3a. Control samples

Figure 4.2A shows the MS2 spectra for the *N*-linked glycans released from IgG for the sample in which no fucose analog was added. There are three *N*-linked glycan species detected – the first has a mass-to-charge ratio (m/z) of 732.7 in the 2+ charge state, corresponding to the known G0F oligosaccharide structure, the second has an m/z of 813.9 corresponding to the G1F oligosaccharide structure, and the third has an m/z of 894.9, corresponding to the G2F structure. G0F, G1F, and G2F, all of which are core fucosylated, are the most common *N*-linked glycan structures modifying IgG.[114] Identifying these structures confirms that the *N*-glycan purification protocol method is effective, in addition, it provides the control from which the effect of the fucose analogs on *N*-glycosylation of IgG can be elucidated. An exhaustive search for the non-core fucosylated (afucosylated) structures (G0, G1, and G2) was negative, indicating that in the absence of fucose analogs, the *N*-linked glycans modifying IgG are ~100% core fucosylated, or at least that any afucosylated structures are so rare in this sample as to be below the detection threshold.

Figure 4.2B shows the EIC obtained after searching for the masses of the afucosylated structures (G0, G1, G2), which in this case were not detected, while figure 4.3B shows the EIC for the fucosylated structures (G0F, G1F, G2F). For this sample, the G1 subset of structures is

the most abundant. In figures 4.3C, 4.3D, and 4.3E, I calculate the area under the curve (AUC)

for each detected structure, and provide a graph comparing the relative abundance of

afucosylated and fucosylated structures in the G0, G1, and G2 subset of structures. Since there

were no detectable afucosylated structures, the relative abundance of the fucosylated structures

with respect to the afucosylated structures (i.e. G0F compared to G0, G1F compared to G1, and

G2F compared to G2) is 100%:0% in each instance.

The EIC spectra are presented primarily to show the quality of the data, as they are the

raw data used for the analysis. Multiple peaks exist primarily due to the separation of the

anomeric forms of the released *N*-glycan structures, but additional artifact peaks can also be

detected, particularly when the signal-to-noise ratio is low. The EIC spectra can also be useful

for comparing the relative abundance of detected G0, G1, and G2 subset of structures. The most

important analysis, however, is of the graphs that are generated after calculating the AUC of

each peak, as the graphs provide a breakdown of the relative abundance of afucosylated,

fucosylated, and in later figures, fucose analog conjugated structures. By comparing the graphs

generated in samples where fucose analogs are added with the graphs generated in this control

sample, the inhibitory effects of the fucose analogs on core fucosylation can be detected.

IgG (no fucose analog added)
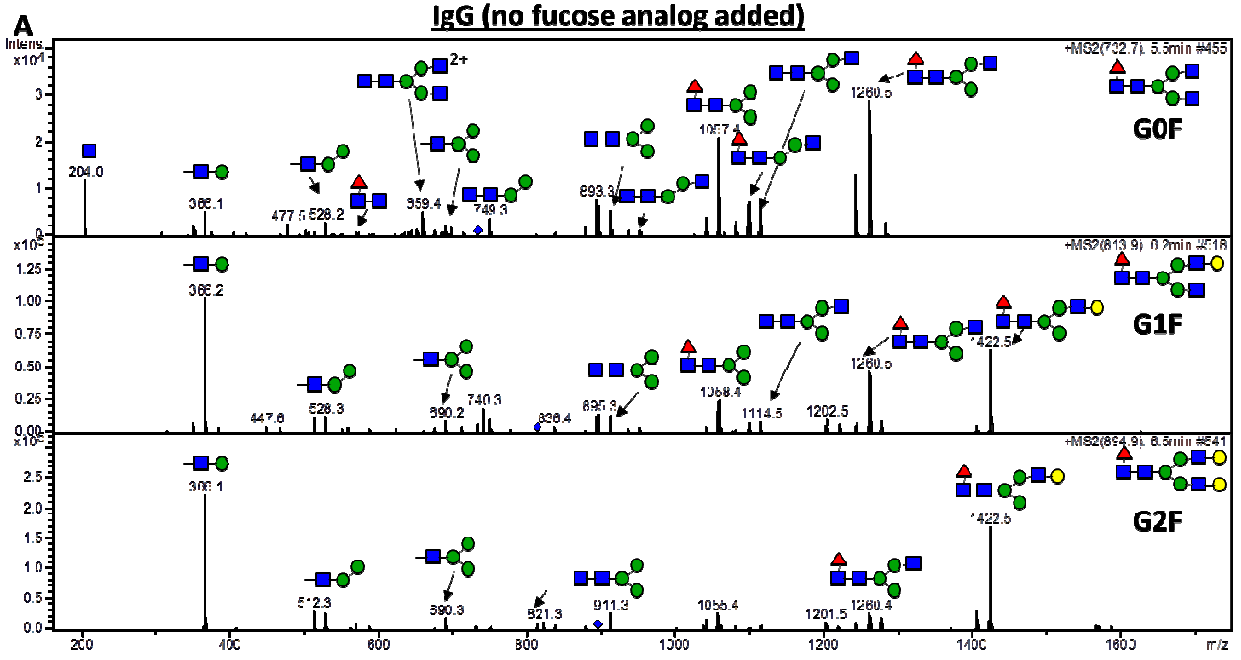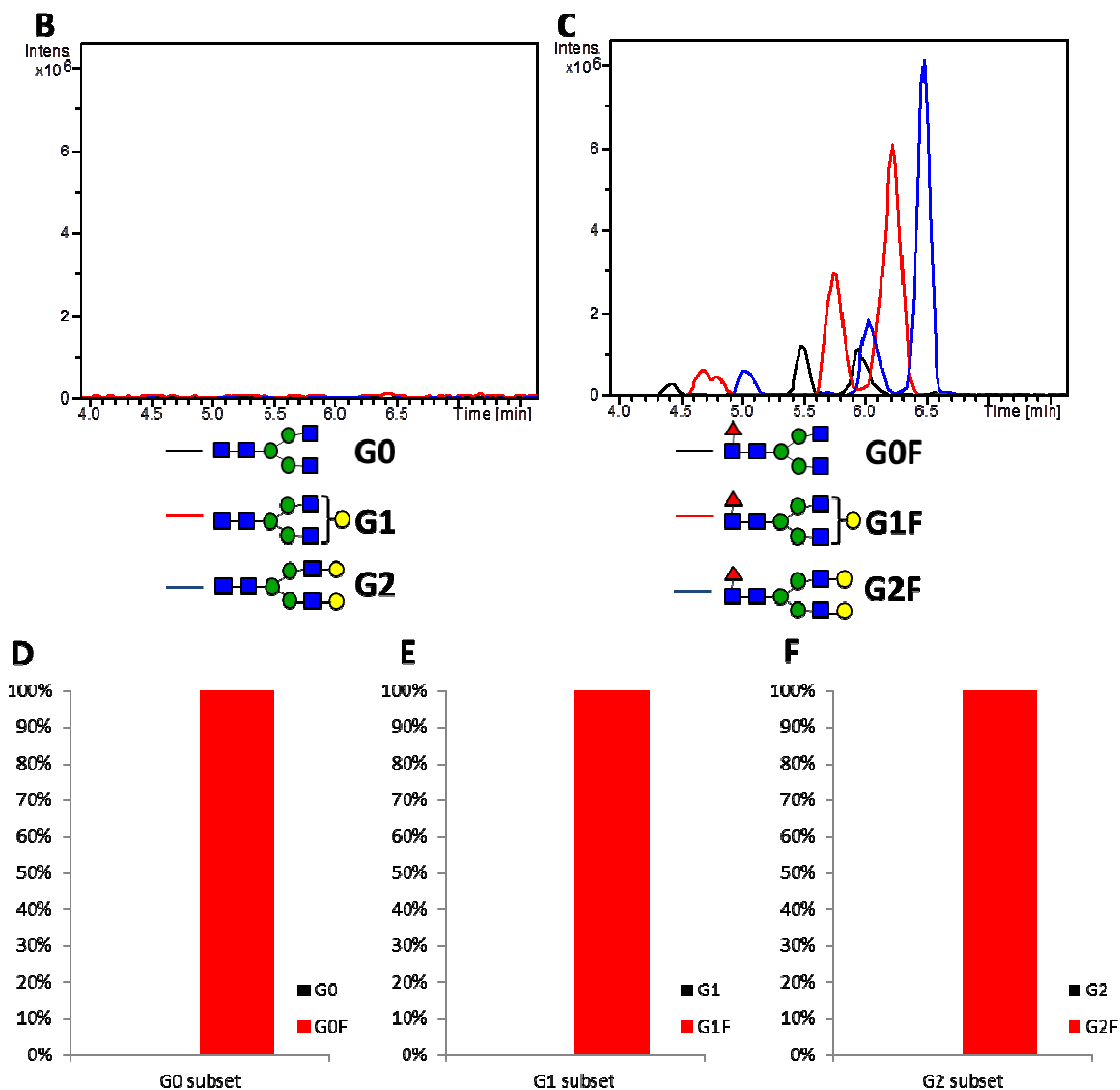
Figure 4.2: MS2 spectra, EICs, and AUC graphs for IgG (no fucose analog added) sample.
A. MS2 spectra corresponding to the core-fucosylated G0F, G1F, and G2F structures are indicated. Many breakdown products produced by collision induced dissociation are highlighted. MS2 spectra for the afucosylated G0, G1, and G2 structures were not detected. **B.** Extracted Ion Chromatograms (EICs) for the afucosylated structures are indicated. **C.** EICs for the fucosylated structures are indicated. It appears that the G1 subset of structures is most abundant for this sample. **D.** Graph displaying the area under the curve (AUC) for the G0/G0F structures. **E.** Graph displaying the area under the curve (AUC) for the G1/G1F structures. **F.** Graph displaying the area under the curve (AUC) for the G2/G2F structures. For all subsets, only the fucosylated structures were detected. Red triangle, fucose; blue square, GlcNAc; green circle, mannose; yellow circle, galactose.

90

4.3b. Effect of 6AF on FUT8

Figure 4.3 shows the MS2 spectra for the *N*-linked glycans released from IgG for the sample treated with 6AF. Spectra corresponding to the G0, G1, G2, G0F, G1F, and G2F structures were positively identified (Figure 4.3A and 4.3B). Additionally, a spectrum corresponding to the G2+6AF structure was also identified (Figure 4.3C), indicating that some 6AF was incorporated by FUT8 onto the *N*-linked glycan of IgG. EICs for the G0/G0F/G0+6AF (G0 subset), G1/G1F/G1+6AF (G1 subset), and G2/G2F/G2+6AF (G2 subset) structures (Figure 4.3D, 4.3E, 4.3F) show that the G0 subset was most abundant in this sample. Calculating the area under the curve (AUC) for each structure's EIC, followed by calculating the percentage of the abundance of each structure in each class provides a convenient way of comparing the relative abundance of afucosylated, fucosylated, and fucose analog-modified *N*-linked glycan structures and a graph displaying this data is provided for each subset (Figure 4.3G, 4.3H, 4.3I.)

In the case of 6AF, the vast majority of the *N*-linked glycan structures are in the afucosylated state, contrary to what was observed in the control sample to which no fucose analog was added, indicating that FUT8 is potently inhibited by 6AF. Small amounts of core-fucosylated and core-6AF modified structures were detected, indicating that the inhibition of FUT8 is not absolute, but the striking observation is that 6AF inhibits core fucosylation as compared to the control sample.
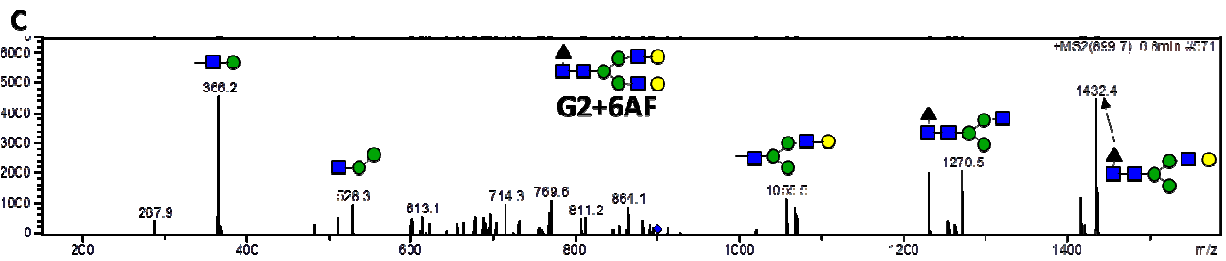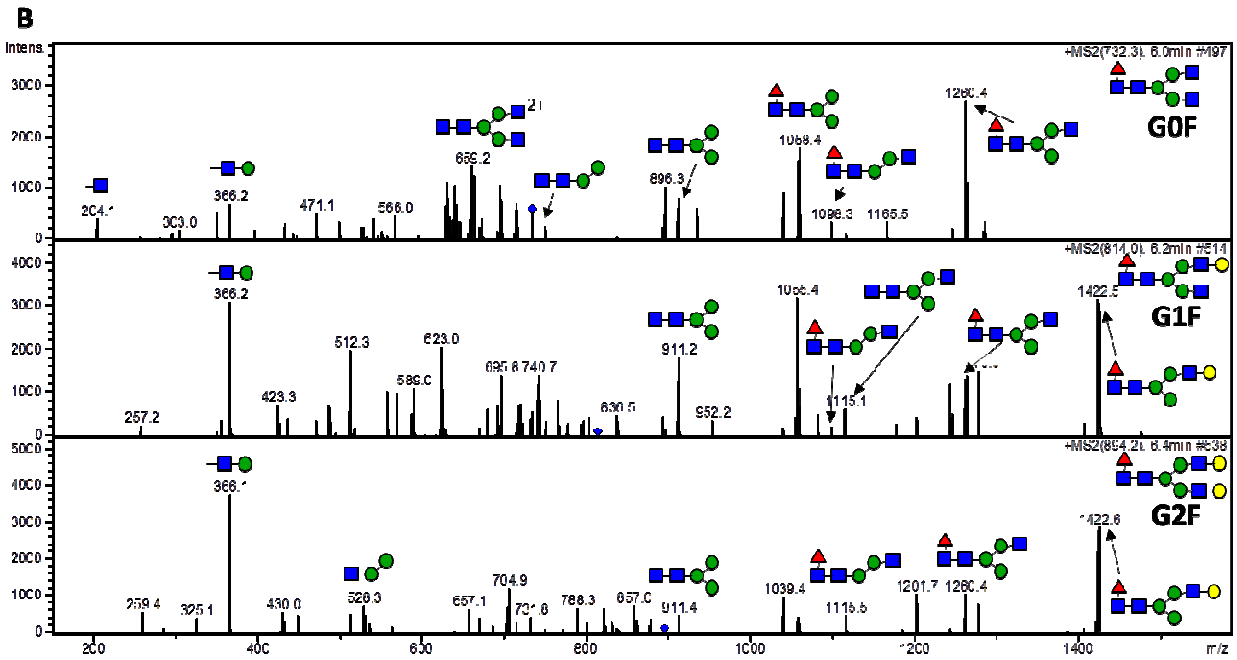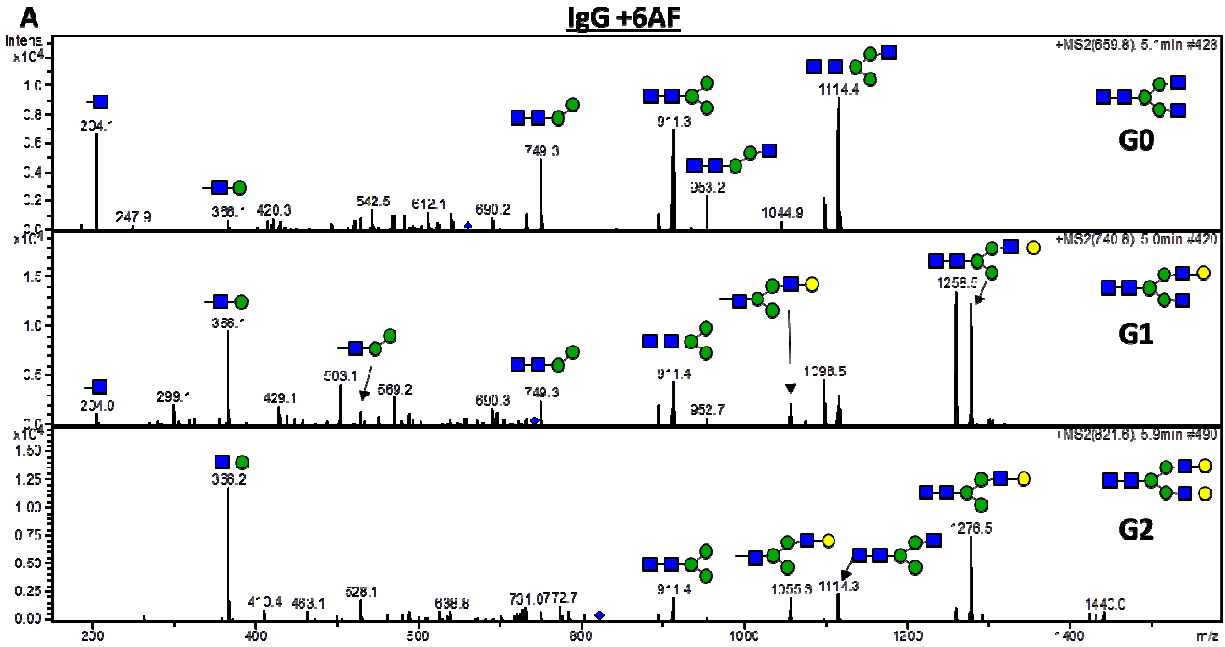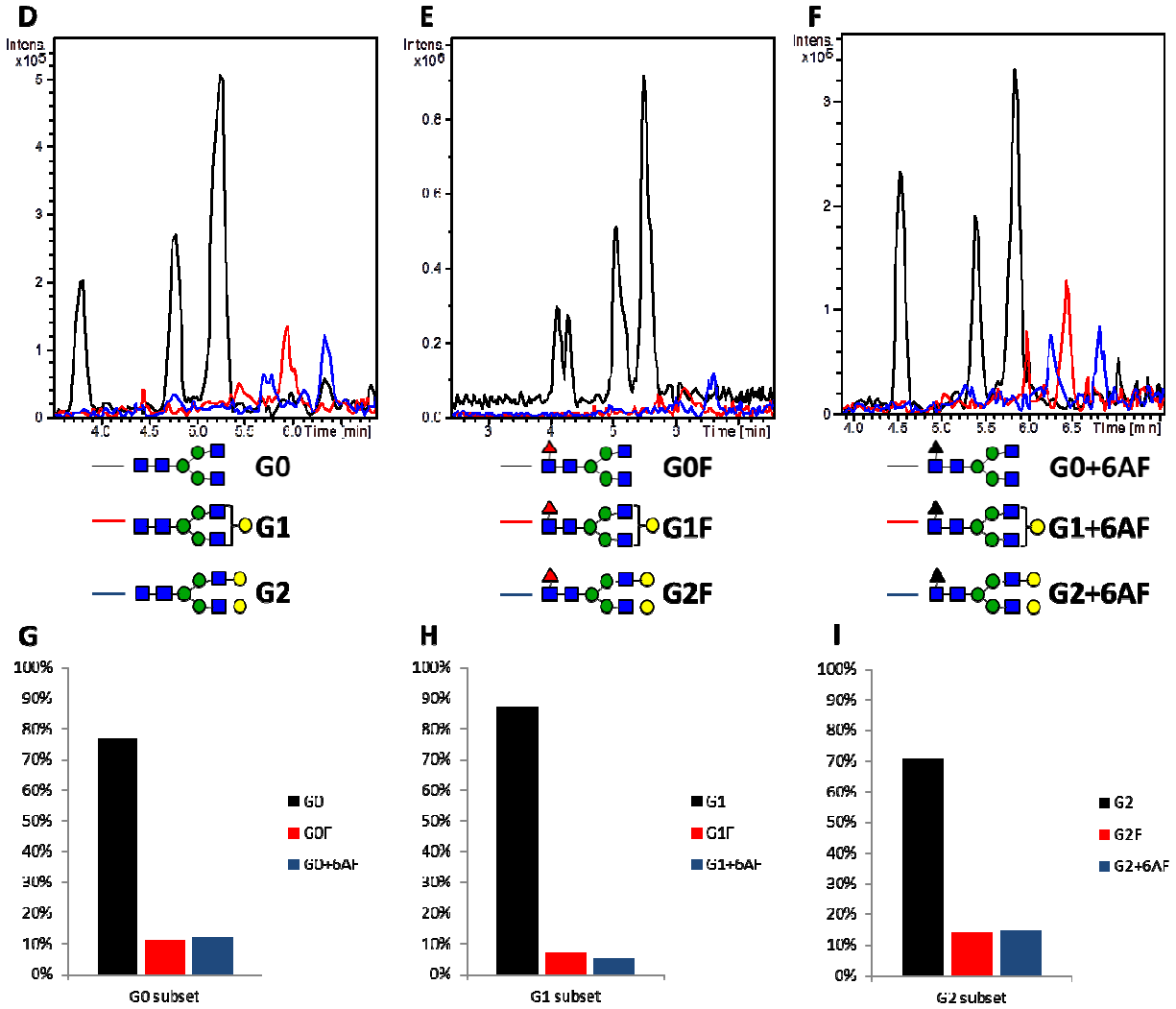
Figure 4.3: MS2 spectra, EICs, and AUC graphs for the IgG + 6AF sample. **A**. MS2 spectra for the afucosylated G0, G1, and G2 structures. **B**. MS2 spectra for the core-fucosylated G0F, G1F, and G2F structures. **C**. MS2 spectrum for the G2+6AF structure. G0+6AF and G1+6AF were not selected for fragmentation. **D**. EIC for the G0 subset structures. The G0 subset structures were most abundant for this sample. **E**. EIC for the G1 subset structures. **F**. EIC for the G2 subset structures. **G.** AUC graph for the G0 subset structures. **H.** AUC graph for the G1 subset structures. **I.** AUC graph for the G2 subset structures. The afucosylated structures were most abundant in this sample, due to 6AF inhibition of core fucosylation. Red triangle, fucose; black triangle, 6AF; blue square, GlcNAc; green circle, mannose; yellow circle, galactose.

4.3c. Effect of 5TF on FUT8

Figure 4.4 shows the MS2 spectra for the *N*-linked glycans released from IgG for the sample treated with 5TF. Spectra corresponding to the G0, G1, G2, G0F, G1F, and G2F structures are positively identified (Figure 4.4A and 4.4B). An exhaustive search of structures modified by 5TF (G0+5TF, G1+5TF, and G2+5TF) was unsuccessful. This does not mean that no 5TF-modified structure was present, but rather that any 5TF-modified structures were not abundant enough to be selected for fragmentation. It is important to note that due to the mass of the 5TF fucose analog, the mass of the G0+5TF and G1 structures as well as the G1+5TF and G2 structures is identical, but analysis of the breakdown products in both cases strongly suggested that the G1 and G2 structures were detected, not the 5TF-modified alternatives. Estimating the relative abundance of each subset by looking at each subset's EIC (Figure 4.4C, 4.4D, 4.4E) revealed that the G1 subset of structures was most abundant in this sample.

The AUC for each structure was calculated and grouped for each subset (figure 4.4F for G0 subset, figure 4.4G for G1 subset, figure 4.4H for G2 subset). Compared to the control sample, only a mild increase in the non-core fucosylated structure was detected, as well as a trace amount of the 5TF-modified structures. The vast majority of the *N*-linked glycan remained core fucosylated as in the case of the control sample. Thus, 5TF only mildly inhibited core fucosylation of the *N*-linked glycan of IgG by FUT8.
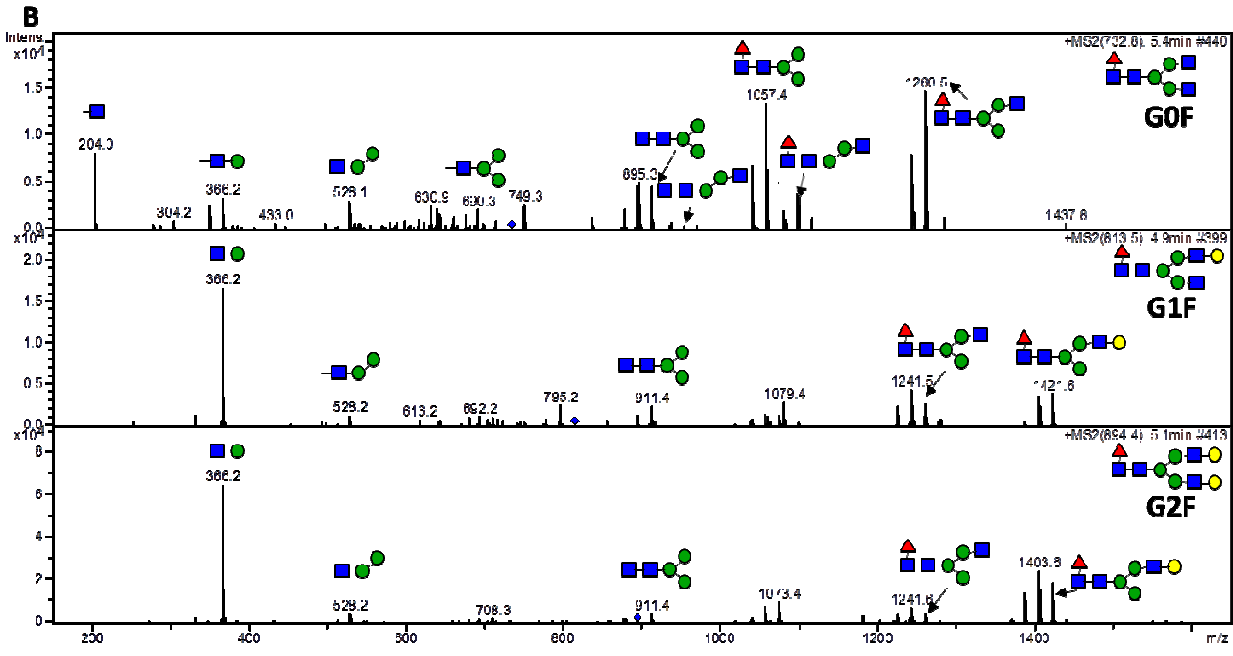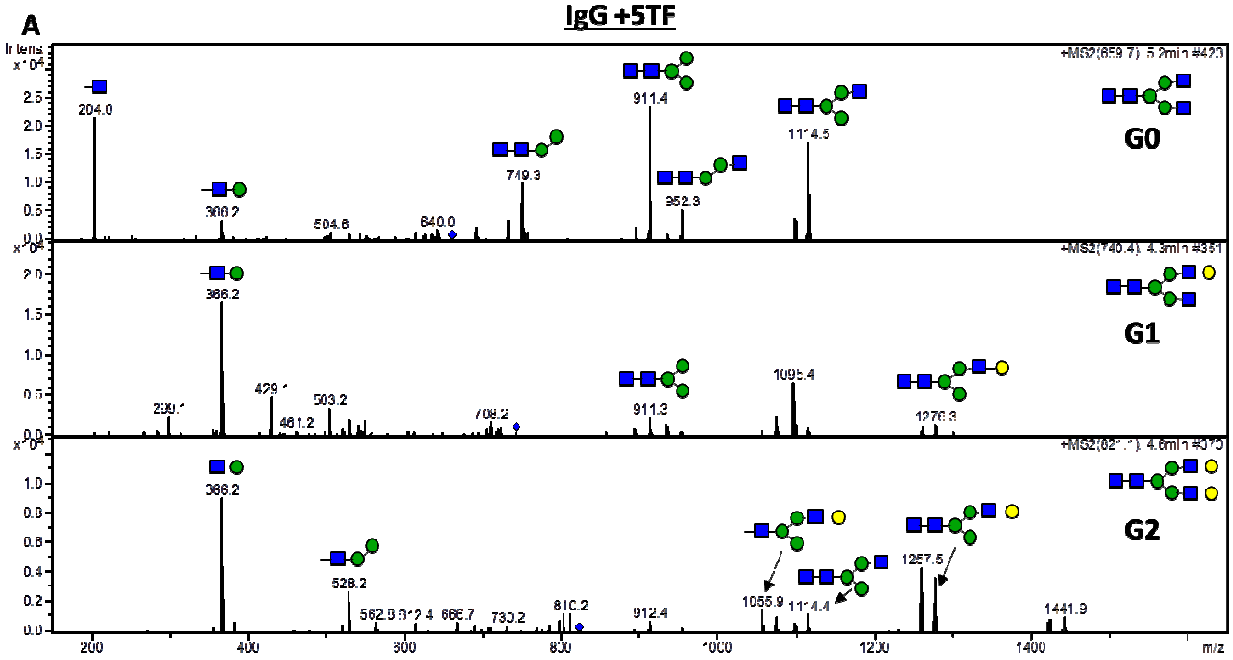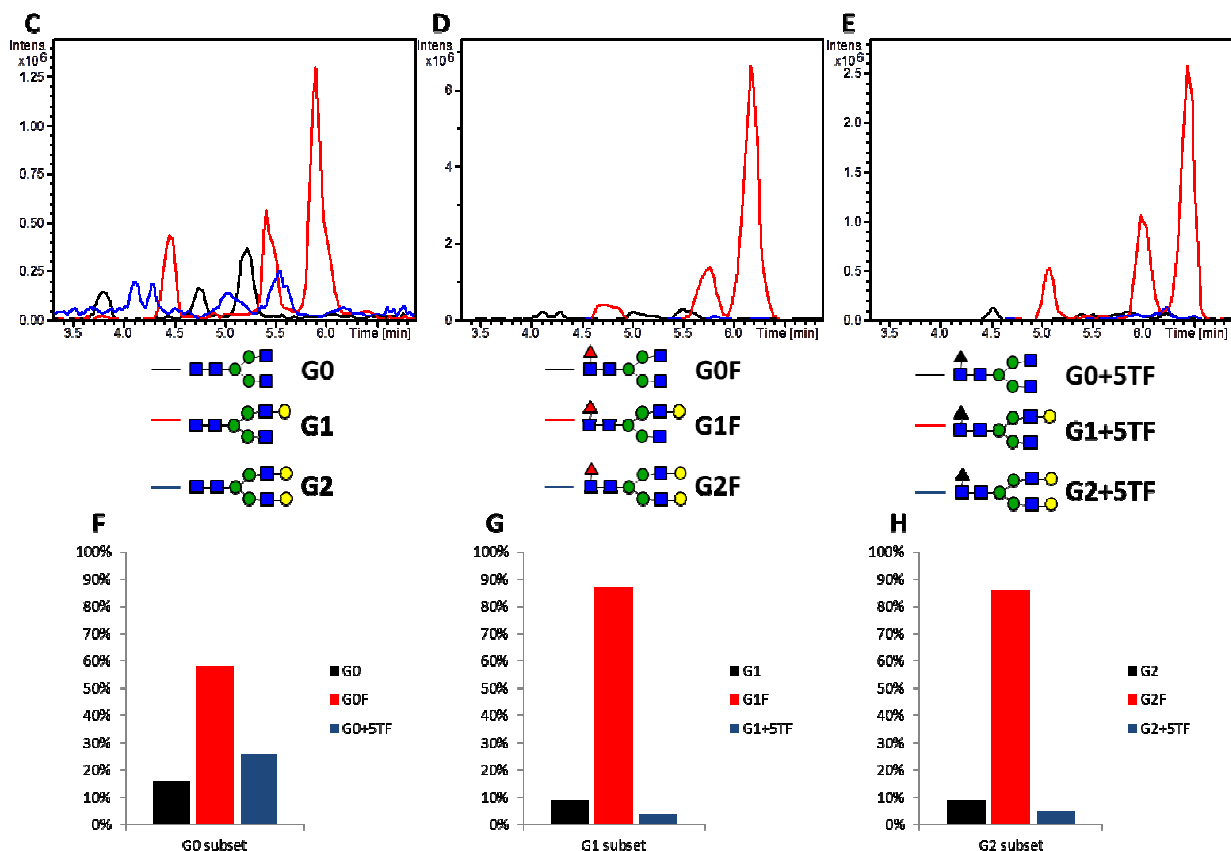
IgG +5TF

Figure 4.4: MS2 spectra, EICs, and AUC graphs for the IgG + 5TF sample. **A**. MS2 spectra for the afucosylated G0, G1, and G2 structures. **B**. MS2 spectra for the core-fucosylated G0F, G1F, and G2F structures. G0+5TF, G1+5TF, and G2+5TF structures were not selected for fragmentation. **C**. EIC for the G0 subset structures. **D**. EIC for the G1 subset structures. The G1 subset structures were most abundant in this sample. **E**. EIC for the G2 subset structures. **F.** AUC graph for the G0 subset structures. **G.** AUC graph for the G1 subset structures. **H.** AUC graph for the G2 subset structures. The fucosylated structures were most abundant in this sample, indicating that 5TF did not potently inhibit core fucosylation. Red triangle, fucose; black triangle, 5TF; blue square, GlcNAc; green circle, mannose; yellow circle, galactose.

4.3d. Effect of 2FF on FUT8

Figure 4.5 shows the MS2 spectra for the *N*-linked glycans released from IgG for the sample treated with 2FF. Similar to the sample treated with 5TF, only spectra corresponding to the G0, G1, G2, G0F, G1F, and G2F structures are positively identified (Figure 4.5A and 4.5B). An exhaustive search of structures modified by 2FF (G0+2FF, G1+2FF, and G2+2FF) was unsuccessful. This does not necessarily indicate that 2FF-modified structure was absent, but rather that any 2FF-modified structures were not abundant enough to be selected for fragmentation. It is important to note that the mass of the 2FF fucose analog is only 2 Da greater than the naturally occurring L-fucose, thus the m/z of the 2FF modified structures would only differ by 1 as compared to the fucose-modified structures. Thus, in generating the EIC for each structure, I modified the usual method of programming the EIC to calculate the desired m/z +/- 0.5 for each structure, to instead select the m/z +/- 0.3 for each structure, in order to better differentiate 2FF-modified structures from fucose-modified structures. While this runs the risk of introducing some error in the quantitation of the EICs, because spectra with an m/z that is greater than 0.3 but less than 0.5 away from the target m/z would not be included in the quantitation, the added benefit of being able to better resolve the 2FF-modified from fucose-modified structures made this difference in the quantitation method reasonable for the 2FF-treated sample. Estimating the relative abundance of each subset by looking at each subset's EIC (Figure 4.5C, 4.5D, 4.5E) revealed that the G2 subset of structures was most abundant in this sample.

The AUC for each structure was calculated and grouped for each subset (figure 4.5F for G0 subset, figure 4.5G for G1 subset, figure 4.4H for G2 subset). Compared with the control sample, only a mild increase in the afucosylated structure was detected, as well as a trace amount

of the 2FF-modified structures. An exception existed for the G2/G2F/G2+2FF structures (G2 subset), where it appeared that 2FF had a more profound inhibitory effect on core fucosylation, similar to the effect of 6AF. Thus, with the exception of the G2-class of structures, the vast majority of the *N*-linked glycan remained core fucosylated as in the case of the control sample. Overall, 2FF only inhibited to varying amounts the core fucosylation of the *N*-linked glycan of IgG by FUT8.
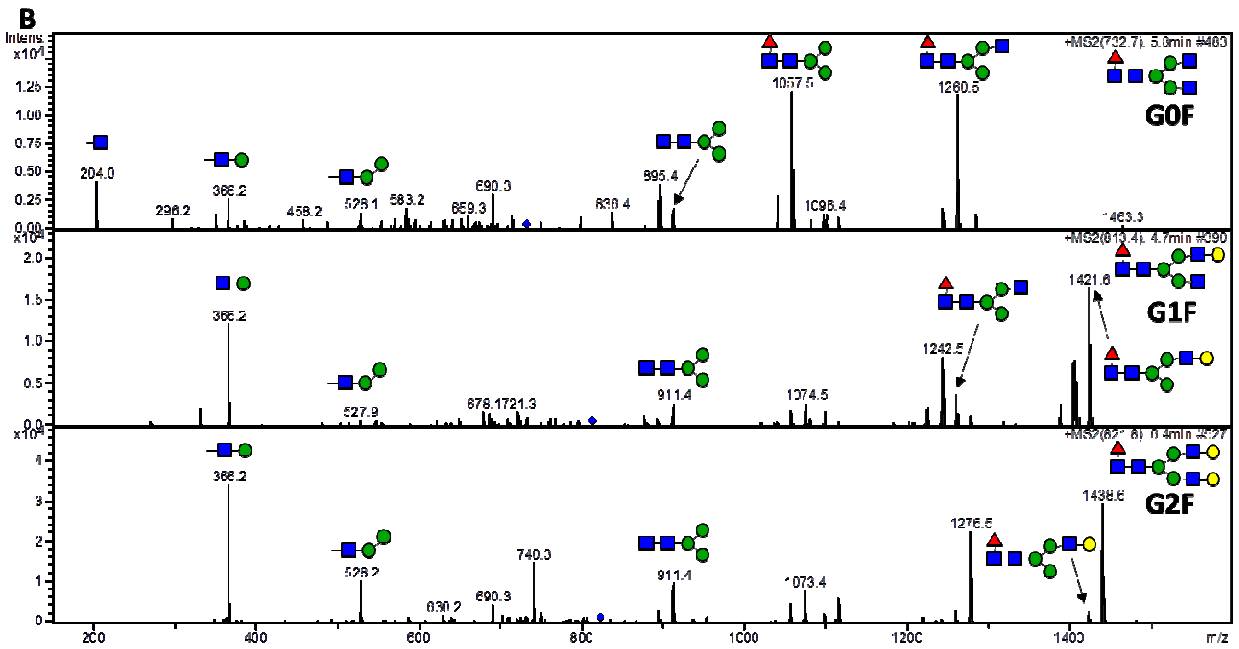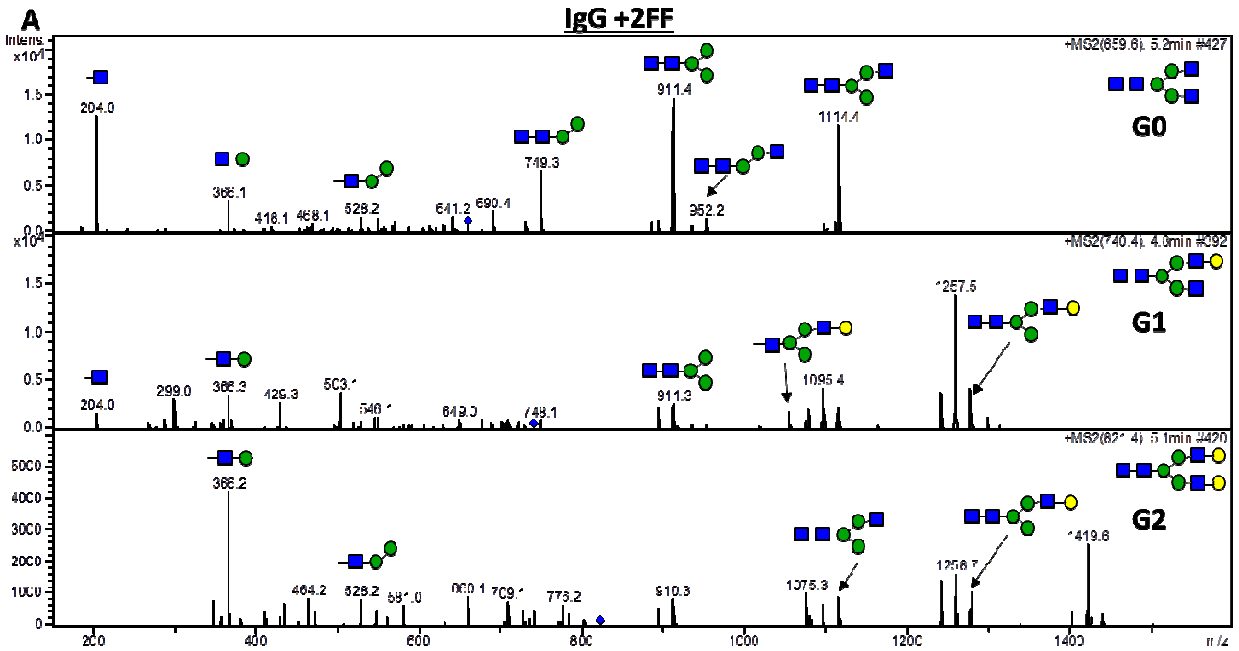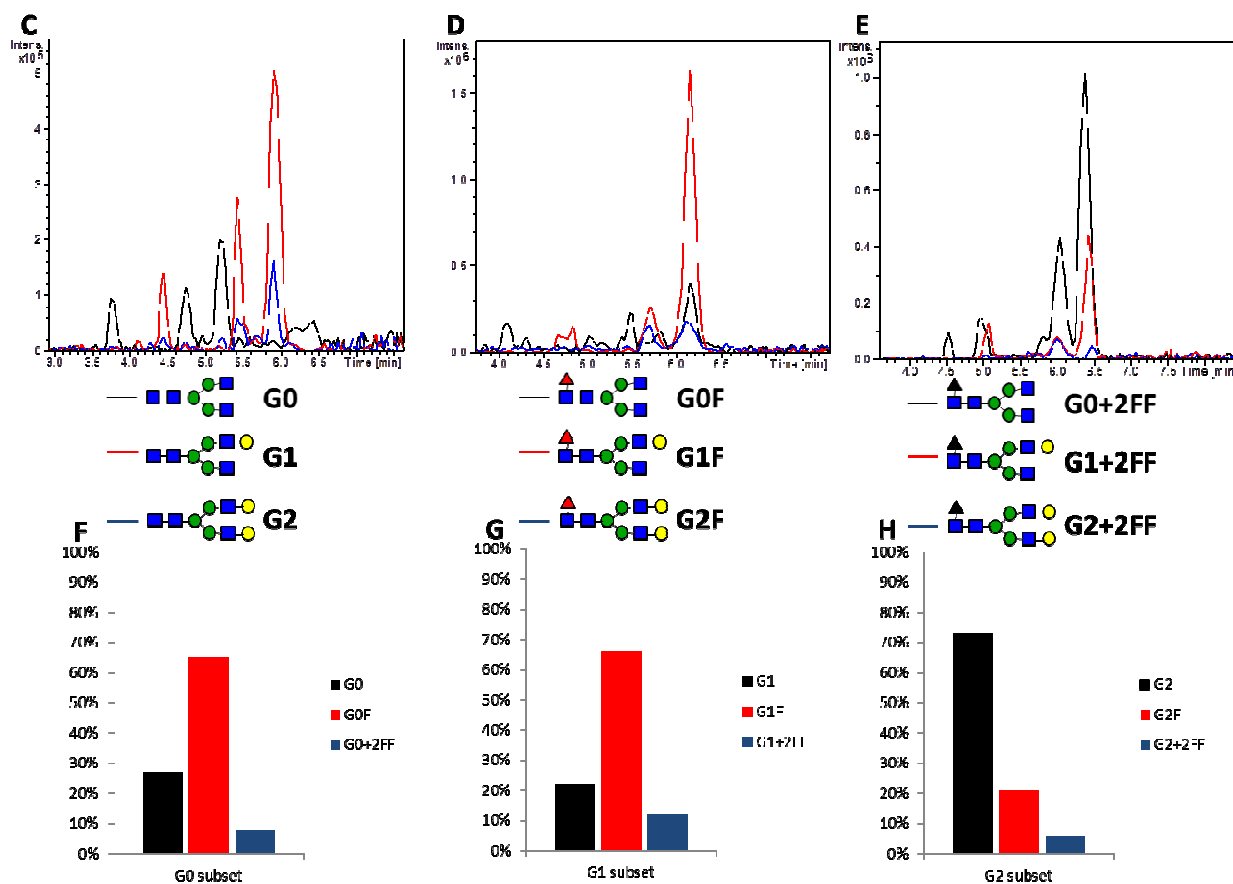
IgG +2FF

Figure 4.5: MS2 spectra, EICs, and AUC graphs for the IgG + 2FF sample. A. MS2 spectra for the afucosylated G0, G1, and G2 structures. **B**. MS2 spectra for the core-fucosylated G0F, G1F, and G2F structures. G0+2FF, G1+2FF, and G2+2FF structures were not selected for fragmentation. **C**. EIC for the G0 subset structures. **D**. EIC for the G1 subset structures. **E**. EIC for the G2 subset structures. The G2 subset structures were most abundant in this sample. **F.** AUC graph for the G0 subset structures. **G.** AUC graph for the G1 subset structures. **H.** AUC graph for the G2 subset structures. The fucosylated structures were most abundant for the G0 and G1 subsets, while the afucosylated structure was most abundant for the G2 subset, indicating that 2FF variably inhibited core fucosylation. Red triangle, fucose; black triangle, 2FF; blue square, GlcNAc; green circle, mannose; yellow circle, galactose.

4.4. The effect of fucose analogs on *O*-fucosylation by POFUT2

In Chapter 2, we concluded that labeling HEK293T cells with 200 µM 6AF was effective and sufficient for *O*-fucosylating TSRs by POFUT2. That study did not address the impact, if any, on inhibiting *O*-fucosylation, instead focusing only on the relative amount of *O*-fucosylated versus *O*-6AF-modified POFUT2 targets. In order to understand the impact of fucose analogs on *O*-fucosylation of POFUT2 targets, I studied the impact of 6AF as well as 5TF and 2FF on the *O*-fucosylation of hTSP1 TSR 1-3. The analysis was performed using the same methods as described in Chapter 2, but with 50 µM of each fucose analog instead of 200 µM (see "*O*-linked glycan identification protocol" in methods).

4.4a. Untreated samples

Figure 4.6 shows the MS and MS2 spectra of the tryptic peptide containing TSR2 on the control sample in which no fucose analog was added. The MS spectra show the so-called "naked" TSR2 peptide (Figure 4.6A), which is not *O*-fucosylated, as well as the TSR2 peptide modified with the fucose-glucose disaccharide (Figure 4.6B). The associated MS2 spectra show the "b and y" ions and neutral loss of the sugars, as expected after collision induced dissociation (CID) of the parent ions, confirming the identification. Next, Figure 4.6C indicates the EIC which shows the relative abundance of the naked TSR2 peptide and the TSR2 peptide modified with the disaccharide, followed by a graph showing the relative abundance of each species by calculating the AUC for the EIC of each structure. The results indicate that approximately three-quarters of the TSR2 peptides exist as naked peptides while approximately one-quarter are modified with the disaccharide for the sample to which no fucose analog is added.
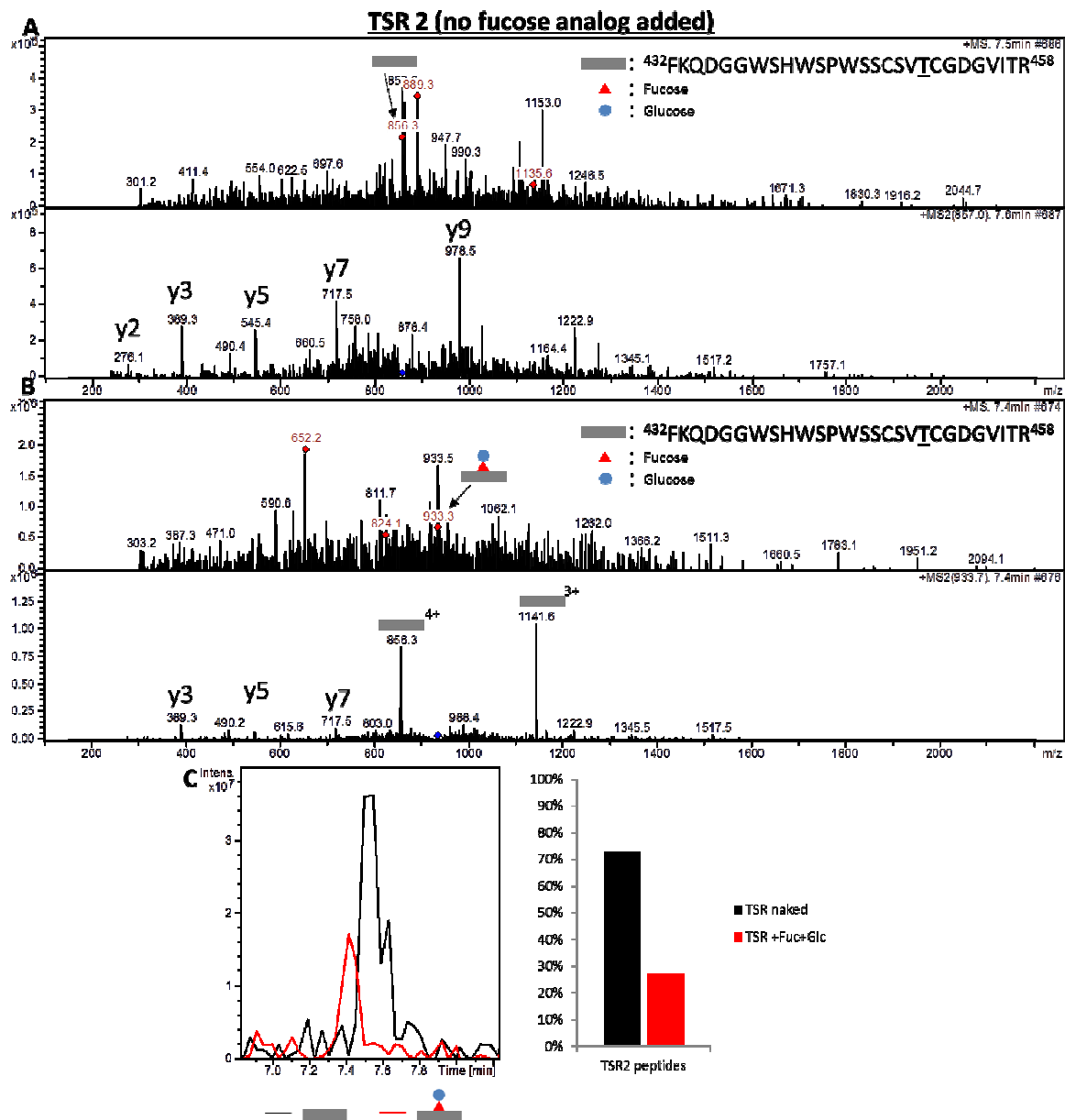
Figure 4.6: MS and MS2 spectra and EIC for the TSR1-3 (no fucose analog added) sample. **A.** MS and MS2 spectra of the TSR2 naked peptide. Several "b and y" ions are indicated in the MS2 spectra. **B.** MS and MS2 spectra of the TSR2 fucose-glucose disaccharide glycopeptide. The MS2 spectrum shows neutral loss of the disaccharide. **C.** EIC of the TSR2 peptides and glycopeptides with associated quantitation of the AUC. Gray rectangle, TSR2 peptide; red triangle, fucose; blue circle, glucose.

4.4b. Effect of 6AF on POFUT2

     Figure 4.7 shows the MS and MS2 spectra of the tryptic peptide containing TSR2 for the sample incubated with 6AF. The MS spectrum contains the naked TSR2 peptide as well as the TSR2 glycopeptide modified with fucose-6AF (Figure 4.7A) and the confirmatory MS2 spectra for each species are indicated (Figure 4.7B1 for the TSR2 naked peptide and figure 4.7B2 for the TSR2 6AF-glucose disaccharide). The TSR2 modified with fucose-glucose disaccharide was not selected for fragmentation. Calculating the AUC for the EIC of each structure (Figure 4.7C) indicated that for the 6AF-treated sample, a greater proportion of the TSR2 peptide exists as the naked peptide as compared to the control sample, suggesting that 6AF may mildly inhibit POFUT2. Examination of only the glycosylated fraction of the peptide, however, showed that the vast majority was 6AF-modified instead of fucosylated (Figure 4.7D). These two observations suggest that while *O*-fucosylation by POFUT2 may be mildly inhibited by 6AF, 6AF can still serve as a good substrate at the concentration of 50 μM used in labeling targets of *O*-fucosylation with this analog. Comparing this data (the examination of the EIC for the glycosylated fraction of the peptide) with the similar analysis performed in Chapter 2 (Figure 2.5C), suggests that using a lower concentration of 50 μM 6AF as compared to 200 μM resulted in a slight decrease in the incorporation of the 6AF fucose analog, as the percentage of the glycosylated fraction of TSR2 in the 200 μM appeared to be somewhat higher than that detected in this analysis at 50 μM 6AF.
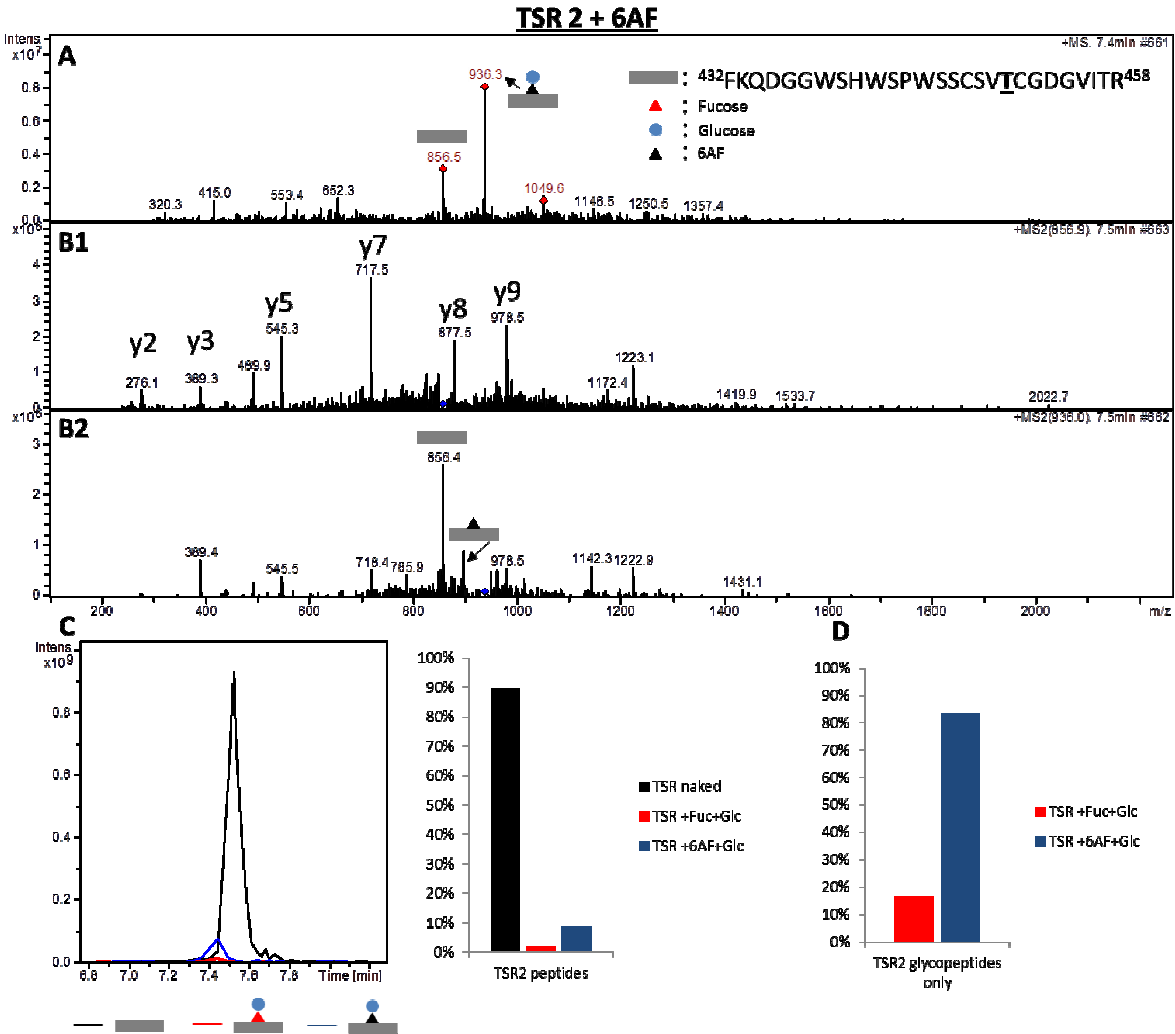
Figure 4.7: MS and MS2 spectra and EIC for the TSR1-3 + 6AF sample. **A.** MS spectrum indicating the TSR2 naked peptide and TSR2 6AF-glucose glycopeptide. TSR2 fucose-glucose glycopeptide was not detected. **B.** *1.* MS2 spectrum of the TSR2 peptide revealing several "b and y" ions confirming identification. *2.* MS2 spectrum of the TSR2 6AF-glucose glycopeptide showing neutral loss of the disaccharide, confirming identification. **C.** EIC of the TSR2 peptides detected, followed by quantitation of the AUC. **D.** AUC quantitation exclusively for the TSR2 glycopeptides (excluding the TSR2 naked peptide contribution). Gray rectangle, TSR2 peptide; red triangle, fucose; blue circle, glucose; black triangle, 6AF.
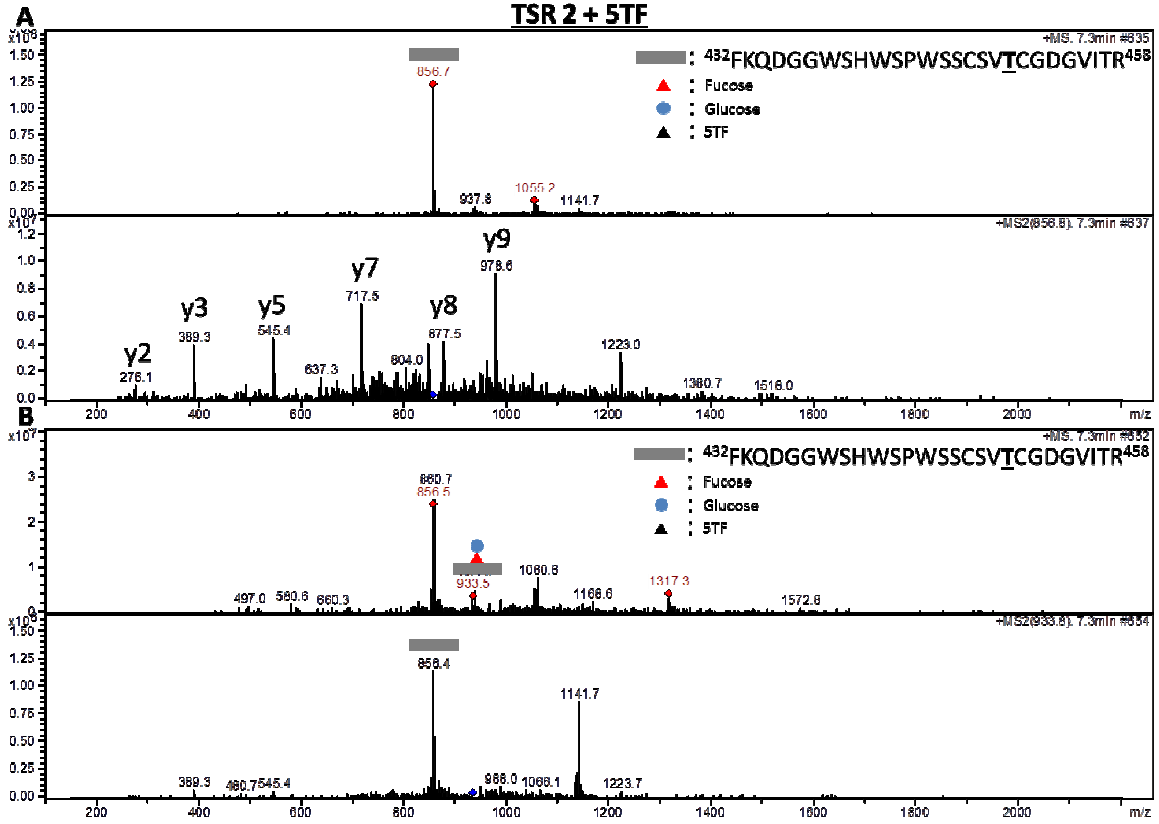
4.4c. Effect of 5TF on POFUT2

Figure 4.8 shows the MS and MS2 spectra of the tryptic peptide containing TSR2 for the sample incubated with 5TF. The MS and confirmatory MS2 spectra for the naked TSR2 peptide (Figure 4.8A), disaccharide modified TSR2 peptide (Figure 4.8B), as well as the TSR2 peptide modified with 5TF-glucose disaccharide are detected (Figure 4.8C). Calculating the AUC for the EIC of each structure (Figure 4.8D) indicated that for the 5TF-treated sample (similar to the 6AF-treated sample), a greater proportion of the TSR2 peptide exists as the naked peptide as compared to the control sample, suggesting that 5TF may mildly inhibit POFUT2. Examination of only the glycosylated fraction of the peptide (Figure 4.8E) showed that the majority was 5TF-modified instead of fucosylated, although incorporation of the 5TF fucose analog was clearly less than that for 6AF. As in the case of 6AF, *O*-fucosylation by POFUT2 may be mildly inhibited by 5TF. 5TF may still serve as an adequate substrate for labeling targets of *O*-fucosylation, though 6AF labeling is more complete at the same concentration added and 6AF would be the superior fucose analog for labeling POFUT2 targets.

TSR 2 + 5TF

A

^{432}FKQDGGWSHWSPWSSCSV**T**CGDGVITR^{458}
▲ : Fucose
● : Glucose
▲ : 5TF

+MS, 7.3min #835

856.7
937.8
1055.2
1141.7

+MS2(856.8), 7.3min #837

y2
276.1
y3
389.3
y5
545.4
637.3
y7
717.5
804.0
y8
877.5
y9
978.6
1223.0
1380.7
1516.0

B

^{432}FKQDGGWSHWSPWSSCSV**T**CGDGVITR^{458}
▲ : Fucose
● : Glucose
▲ : 5TF

+MS, 7.3min #852

860.7
856.5
497.0 580.6 660.3
933.5
1080.8
1166.6
1317.3
1572.8

+MS2(933.8), 7.3min #854

858.4
389.3 460.7 545.4
966.0 1068.1
1141.7
1223.7

200   400   600   800   1000   1200   1400   1600   1800   2000   m/z

Figure 4.8: MS and MS2 spectra and EIC for the TSR1-3 + 5TF sample. A. MS and MS2 spectra for TSR2 naked peptide. **B.** MS and MS2 spectra for TSR2 fucose-glucose glycopeptide. **C.** MS and MS2 spectra for TSR2 5TF-glucose glycopeptide. **D.** EIC of the TSR2 peptides detected, followed by quantitation of the AUC. **E.** AUC quantitation exclusively for the TSR2 glycopeptides (excluding the TSR2 naked peptide contribution). Gray rectangle, TSR2 peptide; red triangle, fucose; blue circle, glucose; black triangle, 5TF.

Figure 4.9 shows the MS and MS2 spectra of the tryptic peptide containing TSR2 for the

sample incubated with 2FF. The MS and confirmatory MS2 spectra show the naked TSR2

peptide (Figure 4.9A and 4.9B1), however, additional spectra cannot unambiguously

differentiate between the disaccharide modified TSR2 peptide and the TSR2 peptide modified

with 2FF-glucose disaccharide (Figure 4.9A and 4.9B2). Differentiation between the 2FF-

modified disaccharide and the fucose-glucose disaccharide is difficult because the 2FF fucose

analog is only 2 Da greater in mass than fucose; as the tryptic peptide is detected in the 4+

charge state, the difference in the m/z ratio between 2FF and fucose is only 0.5, which is within

the error threshold established for the mass spectrometer. Calculating the AUC for the EIC of

each structure indicated that for the 2FF-treated sample, a greater proportion of the TSR2 peptide

exists as the naked peptide as compared to the control sample, suggesting that 2FF may mildly

inhibit POFUT2. Examination of only the glycosylated fraction of the peptide cannot

unambiguously differentiate between the 2FF-glucose disaccharide and the fucose-glucose

disaccharide as explained above. In any case, as with the other fucose analogs tested, *O*-

fucosylation by POFUT2 may be mildly inhibited by 2FF. 2FF is unlikely to serve as an

adequate substrate for labeling targets of *O*-fucosylation, especially as compared to 6AF, because

6AF labeling is more complete and because the m/z ratio difference between 6AF and fucose is

greater than that between 2FF and fucose, avoiding the problem of ambiguous labeling with 2FF.
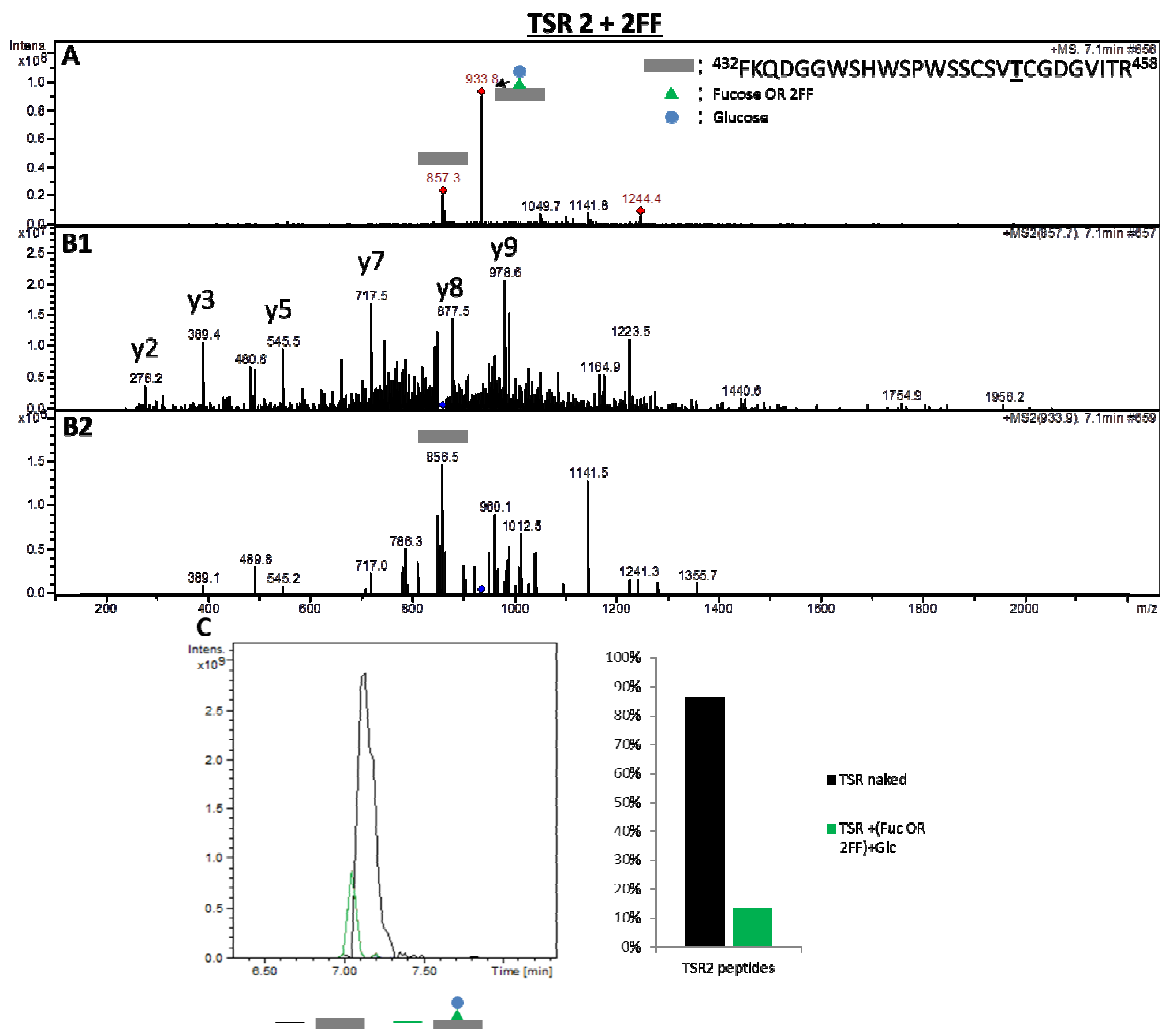
Figure 4.9: MS and MS2 spectra and EIC for the TSR1-3 + 2FF sample. A. MS spectrum indicating the TSR2 naked peptide and TSR2 disaccharide glycopeptide. Due to the small mass of 2FF, the TSR2 disaccharide glycopeptide could be either fucose-glucose or 2FF-glucose . **B. 1.** MS2 spectrum of the TSR2 peptide revealing several "b and y" ions confirming identification. *2.* MS2 spectrum of the TSR2 disaccharide glycopeptide showing neutral loss of the disaccharide. **C.** EIC of the TSR2 peptides detected, followed by quantitation of the AUC. Gray rectangle, TSR2 peptide; green triangle, fucose or 2FF; blue circle, glucose.

4.5. The effect of fucose analogs on O-fucosylation by POFUT1

In Chapter 2, we concluded that labeling HEK293T cells with 200 µM 6AF was effective

and sufficient for *O*-fucosylating EGF repeats by POFUT1. That study did not address the

impact, if any, on inhibiting *O*-fucosylation, instead focusing only on the relative amount of *O*-

fucosylated versus *O*-6AF-modified POFUT1 targets. In order to more generally understand the

impact of fucose analogs on *O*-fucosylation of POFUT1 targets, I studied the impact of 6AF as

well as 5TF and 2FF on the *O*-fucosylation of mN1 EGF 1-5. The analysis was performed using

the same methods as described in Chapter 2, but with 50 µM of each fucose analog (see "*O*-

linked glycan identification protocol" in methods).

4.5a. Control samples

Figure 4.10 shows the MS and MS2 spectra of the tryptic peptide containing EGF3 for

the control sample in which no fucose analog was added. Only the EGF3-fucose species is

detected; no MS spectra for the "naked" EGF3 peptide (which is not *O*-fucosylated) can be

found, but the MS spectrum for EGF3-fucose is clearly identified. The MS2 spectra show the "b

and y" ions expected after collision induced dissociation (CID) of the parent ions, confirming the

identification (Figure 4.10A). Next, the figure indicates the EIC which shows the relative

abundance of the naked EGF peptide and the EGF peptide modified with the monosaccharide,

followed by a graph showing the relative abundance of each species by calculating the AUC for

the EIC of each structure (Figure 4.10B). The results indicate that approximately 100% of the

EGF3 peptide exists as the fucose-modified peptide, as no clear peak was detected for the naked

peptide, for the sample in which no fucose analog is added.
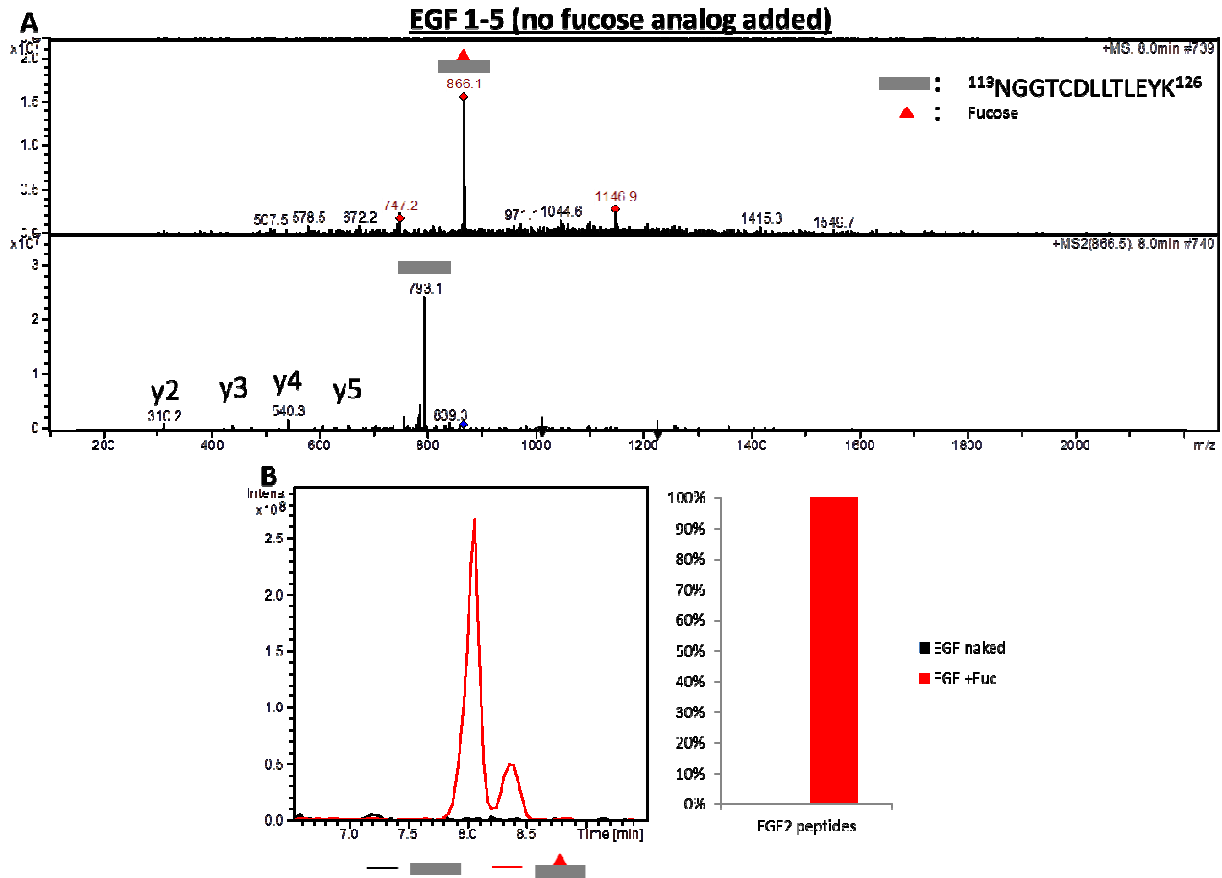
Figure 4.10: MS and MS2 spectra and EIC for EGF 1-5 (no fucose analog added) sample.
A. MS and confirmatory MS2 spectra for EGF3 peptide modified with fucose. The EGF3 naked peptide was not identified. **B.** EIC followed by AUC quantitation of the EGF3 peptides. Gray rectangle, EGF3 peptide; red triangle, fucose.

4.5b. Effect of 6AF on POFUT1

An analysis of EGF3 modified with 6AF was not carried out due to the exhaustion of 6AF stocks. 6AF labeling at the higher concentration of 200 µM was carried out previously (Figure 2.4C), which indicated that in the glycosylated fraction of EGF3, virtually all of the EGF3 peptide was labeled with 6AF, with no detectable levels of EGF3-fucose. This earlier observation does not directly answer the question of whether or not 6AF mildly inhibits POFUT1, nor does it directly address whether the superior efficiency with which 6AF labels EGF3 as compared to 5TF or 2FF labeling is accounted for by the greater concentration of 6AF used.

4.5c. Effect of 5TF on POFUT1

     Figure 4.11 shows the MS and MS2 spectra of the tryptic peptide containing EGF3 for the sample in which 5TF was added. MS spectra for the EGF3 naked peptide (Figure 4.11A), EGF3-fucose peptide (Figure 4.11A), and EGF3+5TF (Figure 4.11C) peptides are clearly detected. Confirmatory MS2 spectra show the expected fragmentation ions of the EGF3 peptides, and the neutral loss of a fucose in the EGF3+fucose spectrum is confirmed (Figure 4.11B1 and 4.11B2). For the EGF3+5TF peptide, however, the neutral loss of the 5TF would be expected to generate an ion with a m/z ratio of 793.1, whereas the detected ion has a mass of 801.4. The spectra were generated for ions in the 2+ charge state, meaning that an additional 16 Da on the neutral loss peak is unaccounted for. This is particularly odd as it would almost suggest a selective loss of the fucose instead of 5TF for the EGF3+5TF, which is difficult to explain. Aside from this aberration, however, the remainder of the MS2 spectrum indicates the expected fragmentation ion for the EGF3 tryptic peptide (Figure 4.11C). Next, the figure indicates the EIC which shows the relative abundance of the naked EGF3 peptide, EGF3+fucose peptide, and the EGF3+5TF peptide, followed by a graph showing the relative abundance of each species by calculating the AUC for the EIC of each structure (Figure 4.11D). A graph showing the relative abundance of the glycopeptides only (excluding the EGF3 naked peptide) is also displayed (Figure 4.11E). The results indicate that while the majority of the EGF3 remains fucosylated, as in the control sample, small but detectable levels of EGF3 naked peptide and EGF3+5TF are detected. This result suggests that, as in the case of POFUT2, 5TF is a mild inhibitor of POFUT1, however the ability of 5TF to replace fucose labeling of EGF3 is quite limited at the concentration of 50 µM, as the vast majority of the EGF3 peptide exists as EGF3+fucose.
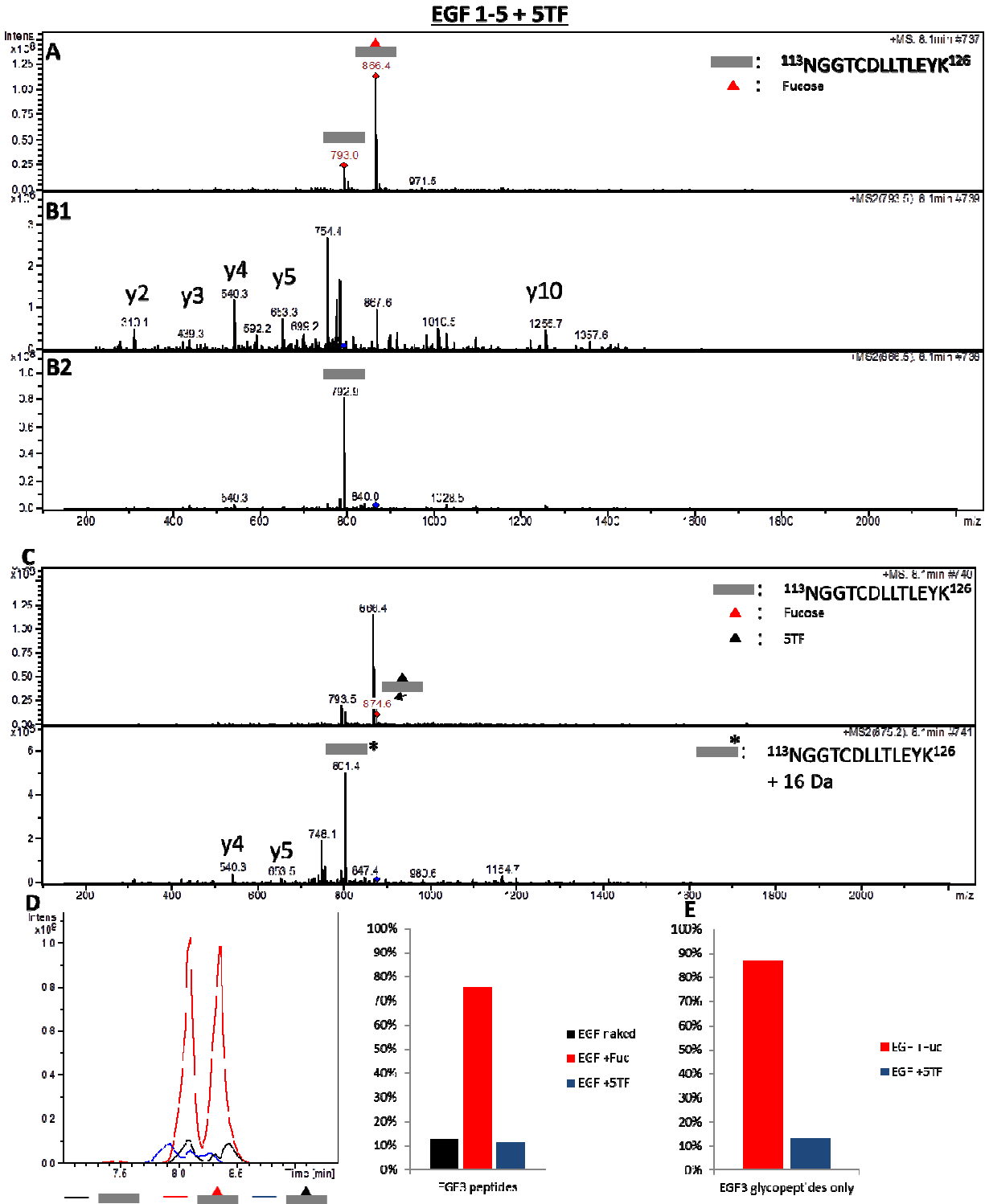
Figure 4.11: MS and MS2 spectra and EIC for the EGF1-5 + 5TF sample. A. MS spectrum indicating the EGF3 naked peptide and EGF3+fucose glycopeptide. **B.** *1.* MS2 spectrum of the EGF3 naked peptide revealing several "b and y" ions confirming identification. *2.* MS2 spectrum of the EGF3+fucose glycopeptide showing neutral loss of the monosaccharide. **C.** MS and MS2 spectra of the EGF3+5TF glycopeptide. The MS2 spectrum shows an aberration in that the

114

neutral loss of the 5TF is expected to yield the EGF3 naked peptide, but a structure with an additional 16 Da is detected. **D.** EIC of the EGF3 peptides detected, followed by quantitation of the AUC. **E.** AUC quantitation for the glycopeptides only. Gray rectangle, EGF3 peptide; red triangle, fucose; black triangle, 5TF.

Figure 4.12 shows the MS and MS2 spectra of the tryptic peptide containing EGF3 for the sample in which 2FF was added. MS spectra for the EGF3 naked peptide (Figure 4.12A), EGF3+fucose peptide (Figure 4.12B), and EGF3+2FF peptides (Figure 4.12C) are clearly detected. Confirmatory MS2 spectra show the expected fragmentation ions of the EGF3 peptides, including the neutral loss of a fucose in the EGF3+fucose spectrum and the neutral loss of 2FF in the EGF+2FF spectrum. Next, the figure indicates the EIC which shows the relative abundance of the naked EGF3 peptide, EGF3+fucose peptide, and the EGF3+2FF peptide, followed by a graph showing the relative abundance of each species by calculating the AUC for the EIC of each structure (Figure 4.11D) and an additional graph showing the relative abundance of the glycopeptides only (Figure 4.11E). The results indicate that while the majority of the EGF3 remains fucosylated, as in the control sample, small but detectable levels of EGF3 naked peptide and EGF3+2FF are observed. This result closely mirrors that observed for 5TF (Figure 11), indicating that 2FF is a mild inhibitor of POFUT1, however the ability of 2FF to replace fucose labeling of EGF3 is quite limited at the concentration of 50 µM, as the vast majority of the EGF3 peptide exists as EGF3+fucose.
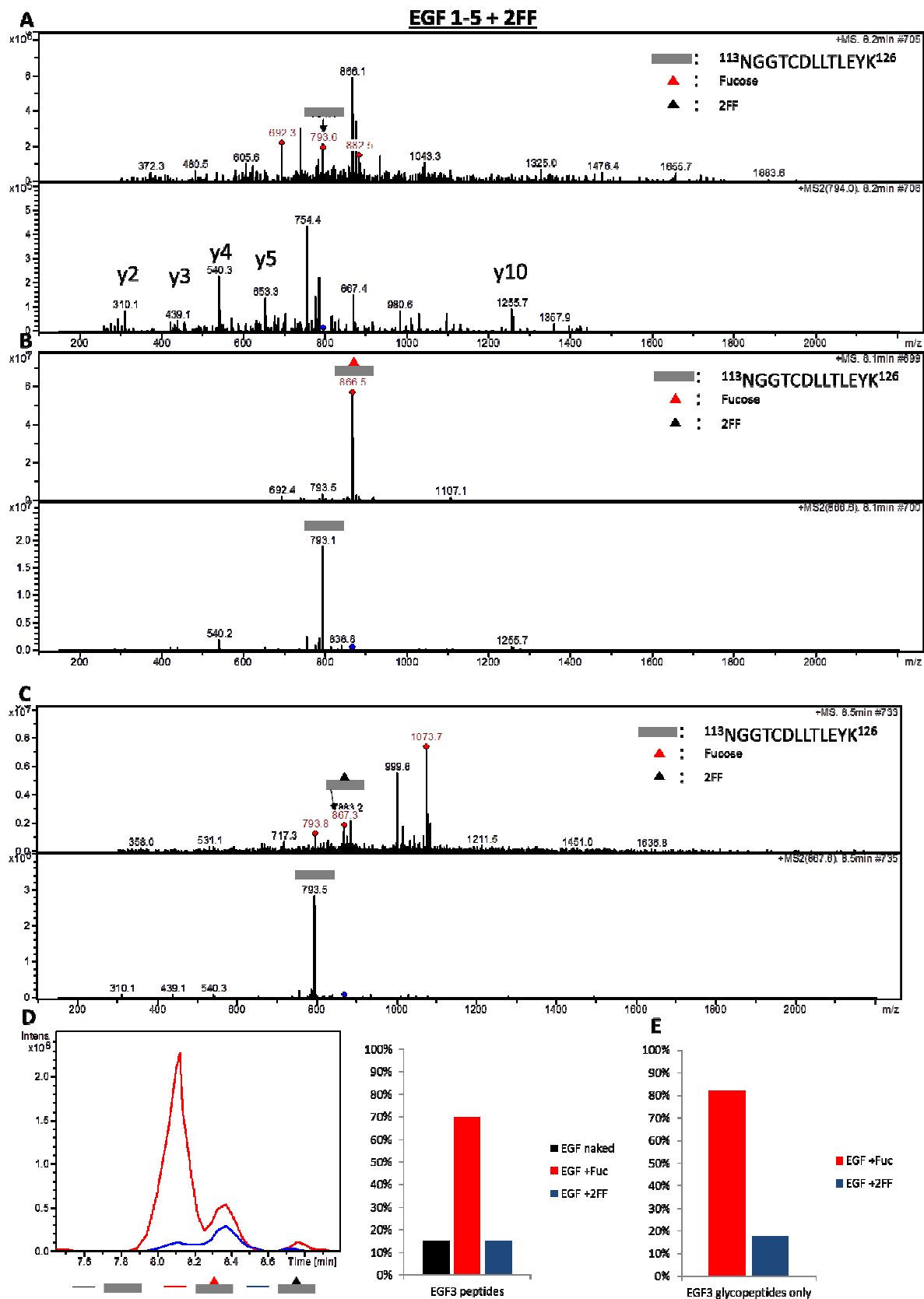
Figure 4.12: MS and MS2 spectra and EIC for the EGF1-5 + 2FF sample. A. MS and MS2

spectra indicating the EGF3 naked peptide. **B.** MS and MS2 spectra indicating the EGF3 peptide modified with fucose. **C.** MS and MS2 spectra indicating the EGF3 peptide modified with 2FF. **D.** EIC of the EGF3 peptides detected, followed by quantitation of the AUC. **E.** AUC quantitation for the glycopeptides only. Gray rectangle, EGF3 peptide; red triangle, fucose; black triangle, 2FF.

4.6. Discussion

The use of fucose analogs as potential inhibitors of core fucosylation of *N*-linked glycans was previously attempted by Okeley et al. in 2013.[53] In that paper, the authors examined the use of 50 µM 6AF (called 5-alkynylfucose in their paper) and 2FF as potential inhibitors of core fucosylation of the *N*-linked glycan of an IgG$_1$ monoclonal antibody directed against CD70 carried out by FUT8 samples grown in Chinese hamster ovary (CHO) cells. The work in this chapter expanded the Okeley analysis to additionally considering 5TF as well as examined the impact of the fucose analogs as substrates or inhibitors of POFUT1 and POFUT2 in addition to FUT8. The work was carried out in HEK293T cells and the mass spectrometric analysis carried out by different instruments and methods. We also focused on all three common subclasses of IgG (G0, G1, and G2) whereas the Okeley analysis focused only on the G0 subclass.

Where our analyses overlap, my data largely agrees with that published by Okeley et al., in that we both observe inhibition of core fucosylation of a FUT8 target by 6AF, but whereas Okeley et al. observed an even greater inhibition of FUT8 with 2FF, my results suggest that 2FF is only a mild inhibitor of FUT8. The discrepancy is likely explained by the fact that my analysis was of samples grown in HEK293T cells whereas Okeley used CHO cells. It is known that glycosylation patterns of recombinant proteins produced in human cells differs from those grown in CHO cells[116], possibly explaining why Okeley only observed and commented on the G0 subclass of glycosylations and also explaining the differences observed in samples generated in CHO cells as compared to HEK293T cells in terms of the inhibitory potential of 2FF on FUT8.

4.7. Conclusions

6AF is a potent inhibitor of core fucosylation of the *N*-linked glycan of the Fc heavy chain of IgG carried out by FUT8, whereas 5TF and 2FF only mildly inhibited core fucosylation of IgG. All fucose analogs partially inhibited overall *O*-fucosylation by POFUT1 and POFUT2 while simultaneously being used as substrates by these fucosyltransferases and incorporated onto mN1 EGF 1-5 and hTSP TSR 1-3. As substrates for POFUT1 and POFUT2, incubation of HEK293T cells with 200 µM of 6AF was the most effective method of labeling *O*-fucose targets.

4.8. Materials and Methods

N-glycan purification and identification protocol

Media from transiently transfected HEK293T cells transfected with human IgG and supplemented with either 50 µM peracetylated 6AF (Life Technologies, catalog #C10264), 5TF (obtained from Vocadlo lab[103]), 2FF (CarboSynth, product code MD06089), or not supplemented with fucose analog, was collected after 72 hours. Protein concentration and purity were assessed by Coomassie Brilliant Blue (Thermo Scientific, Rockford, IL) along BSA standards. Approximately 1 µg IgG was acetone precipitated with four volumes acetone, resuspended with 2x sample buffer, and subjected to SDS-PAGE, followed by transfer for 50 min at 0.40 A to PVDF membrane (Millipore). Staining of the PVDF membrane with Direct Blue 71 (Sigma-Aldrich, catalog #212407-50G) allowed visualization of the IgG bands, which were cut with a clean razor and transferred to separate wells in a flat-bottom polypropylene 96-well plate (Corning). The remainder of the protocol directly followed the previously described method[117]. Briefly, 100 µL of polyvinylpyrrolidone (PVP40; Sigma-Aldrich, catalog #PVP40-50G) was applied to each well, to block the membranes and wells in order to prevent nonspecific PNGaseF enzyme binding. After shaking the plate and washing three times with water, the water is removed and 5 µL of PNGaseF enzyme (produced in-house; concentration ~5U/µL) was added to each well, followed by 10 µL of water, and the plate placed in a 37 °C incubator overnight. The next day, the plate is sonicated, the samples in each well are collected, and the wells are washed twice and the washes are pooled with the sample. 10 µL of 100 mM ammonium acetate (pH 5) is added and incubated with each sample for one hour at room temperature, followed by vacuum concentration of samples in a SpeedVac concentrator (without heating) until dry. The samples are then resuspended in 3% B mass spectrometer buffer (A: 0.1% formic acid in water;

B: 0.1% formic acid, 95% acetonitrile, 4.9% water) and analyzed by nano-LC-MS/MS using the "glycan chip" (Agilent, G4240-62003) on an Agilent model 6340 ion trap mass spectrometer using collision-induced dissociation (CID) as described previously.[118]

O-linked glycan identification protocol

Glycopeptide analysis was done as described in Chapter 2. Briefly, media from transiently transfected pSecTag Notch 1 EGF1-5 pSecTag hTSP1 TSR1-3 HEK293T cells supplemented with either 50 µM 6AF, 5TF, 2FF, or not supplemented with fucose analog, was collected after 72 hours, bound to Ni^{2+}-NTA, washed with 5 mL TBS containing 0.5M NaCl and 10 mM Imidazole, 5 mL cold RIPA buffer, and eluted with TBS containing 250 mM Imidazole. Protein concentration and purity were assessed by Coomassie Brilliant Blue (Thermo Scientific, Rockford, IL) along BSA standards. Approximately 1 µg EGF1-5 or TSR1-3 was acetone precipitated with four volumes acetone, vortexed, and frozen at -80°C overnight. Precipitated protein was reduced, alkylated, and typsinized, followed by analysis by nano-LC-MS/MS using a Zorbax 300SB-C18 nano-CHIP on an Agilent model 6340 ion trap mass spectrometer using collision-induced dissociation (CID) as previously described.[104]

Chapter 5: Conclusion

5.1. Final Conclusions

In this dissertation, I have outlined my work with fucose analogs. I was able to demonstrate that POFUT1 and POFUT2 efficiently incorporate 6AF onto predicted *O*-fucosylated proteins mN1 EGF 1-5 and hTSP1 TSR 1-3, respectively, and that incorporation of the fucose analog did not perturb the normal elongation of the monosaccharide in those proteins.

I also demonstrated that 6AF-labeling of HEK293T cells could be used as part of a proteomics protocol to metabolically label and identify *O*-fucosylated proteins, specifically mN1 EGF 1-5, hTSP1 TSR 1-3, and ADAMTSL1 TSR1-4 (both POFUT1 and POFUT2 targets), but that the method was only successful in the "proof-of-concept" stage where *O*-fucosylated proteins were overexpressed. The ultimate goal of being able to identify the far less abundant amounts of endogenously expressed *O*-fucosylated proteins was completely unsuccessful. Nevertheless, improvements in mass spectral resolution, creative new methods using novel tools (like further pursuing the Bertozzi lab "IsoTag" or the Wu lab "supersensitive" click reagents), or moving away from the biotin-streptavidin purification may be helpful as part of new approaches to tackling the proteomics puzzle and "cracking" the *O*-fucose proteome. In particular, exploring the newly available azido-magnetic beads and performing "on-bead" click chemistry as a way of retaining, without biotinylation, 6AF-modified proteins is a particularly tantalizing new approach which I would recommend as the first experiment for an ambitious future graduate student to pursue if he/she wishes to further explore this project.

I have also confirmed that 6AF is a potent inhibitor of FUT8 and can be used to drastically reduce the amount of core-fucosylated IgG, potentially making 6AF a useful adjunct to anticancer therapeutics as non-core fucosylated therapeutic antibodies display far greater

potency than fucosylated antibodies in triggering antibody-dependent cellular cytotoxicity, enhancing an important mechanism of anticancer therapeutic antibodies. The other two fucose analogs I tested, 5TF and 2FF, were poor substrates for FUT8, but did not have the inhibitory qualities of 6AF and only marginally reduced the levels of core fucosylation of IgG. 5TF and 2FF were also likely less efficiently utilized by POFUT1 and POFUT2 compared to 6AF, making 6AF the optimal fucose analog of the three tested both in labeling *O*-fucose targets by POFUT1 and POFUT2 and in inhibiting core fucosylation carried out by FUT8. The methods I have worked out in testing the fucose analogs for their potential as inhibitors are particularly convenient, requiring a small amount of sample (1/20th of a single 100-mm plate of cells) and reliably generating data from metabolic labeling of cells to semi-quantitative %-inhibition results in as little as a week. These methods can be easily adapted to screening dozens of additional FUT8 inhibitors in the future.

Several important future directions remain. The elucidation of the *O*-fucose proteome is certainly a worthy endeavor, both for confirming that many of the predicted targets of *O*-fucosylation are indeed modified, but also for the potential of discovering new targets of *O*-fucosylation. The knowledge that POFUT1 and POFUT2 knockout mice result in embryonic lethality underscores the importance of *O*-fucosylation, and understanding the mechanism(s) by which disruptions in *O*-fucosylation result in such a severe phenotype will require generating a list of *O*-fucosylated proteins as a necessary first step. Additionally, while 6AF is a potent inhibitor of core fucosylation by FUT8, the data generated suggest that 6AF does not completely abrogate core fucosylation. Continuing work on potent FUT8 inhibitors until such a compound is discovered can be of enormous importance for future anticancer therapeutics, as many emerging anticancer drugs are monoclonal antibodies designed to trigger ADCC. Finding an

inhibitor that can completely eliminate core fucosylation could be part of an effective strategy to generate core fucose depleted antibodies that are orders of magnitude more potent than their core fucosylated counterparts in triggering ADCC and achieving desired clinical outcomes in cancer patients.

References

1.   Becker DJ & Lowe JB (2003) Fucose: biosynthesis and biological function in mammals. *Glycobiology* 13(7):41R-53R.

2.   Werz DB*, et al.* (2007) Exploring the structural diversity of mammalian carbohydrates ("glycospace") by statistical databank analysis. *ACS chemical biology* 2(10):685-691.

3.   Ma B, Simala-Grant JL, & Taylor DE (2006) Fucosylation in prokaryotes and eukaryotes. *Glycobiology* 16(12):158R-184R.

4.   Pang P-C*, et al.* (2011) Human Sperm Binding Is Mediated by the Sialyl-Lewisx Oligosaccharide on the Zona Pellucida. *Science* 333(6050):1761-1764.

5.   Kim Y & Varki A (1997) Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate Journal* 14(5):569-576.

6.   Wang X, Gu J, Miyoshi E, Honke K, & Taniguchi N (2006) Phenotype changes of Fut8 knockout mouse: core fucosylation is crucial for the function of growth factor receptor(s). *Methods in enzymology* 417:11-22.

7.   Seth A, Machingo QJ, Fritz A, & Shur BD (2010) Core fucosylation is required for midline patterning during zebrafish development. *Developmental dynamics : an official publication of the American Association of Anatomists* 239(12):3380-3390.

8.   Listinsky JJ, Siegal GP, & Listinsky CM (2011) The emerging importance of alpha-L-fucose in human breast cancer: a review. *American journal of translational research* 3(4):292-322.

9.   Miyoshi E*, et al.* (2012) Fucosylation is a promising target for cancer diagnosis and therapy. *Biomolecules* 2(1):34-45.

10.   Miyoshi E (2008) Fucosylation and Cancer. *Experimental Glycoscience*, eds Taniguchi N, Suzuki A, Ito Y, Narimatsu H, Kawasaki T, & Hase S (Springer Japan), pp 235-237.

11.   Sawke NG & Sawke GK (2010) Serum fucose level in malignant diseases. *Indian journal of cancer* 47(4):452-457.

12.   Mehta A & Block TM (2008) Fucosylated glycoproteins as markers of liver disease. *Disease markers* 25(4-5):259-265.

13.   Hiraoka A*, et al.* (2012) Huge pancreatic acinar cell carcinoma with high levels of AFP and fucosylated AFP (AFP-L3). *Internal medicine (Tokyo, Japan)* 51(11):1341-1349.

14.	Villar-Portela S, Muinelo-Romay L, Cuevas E, Gil-Martin E, & Fernandez-Briera A (2013) FX enzyme and GDP-L-Fuc transporter expression in colorectal cancer. *Histopathology* 63(2):174-186.

15.	Listinsky JJ, Siegal GP, & Listinsky CM (2013) Glycoengineering in cancer therapeutics: a review with fucose-depleted trastuzumab as the model. *Anti-cancer drugs* 24(3):219-227.

16.	Merino P*, et al.* (2012) Recent progress on fucosyltransferase inhibitors. *Mini reviews in medicinal chemistry* 12(14):1455-1464.

17.	Yamane-Ohnuki N & Satoh M (2009) Production of therapeutic antibodies with controlled fucosylation. *mAbs* 1(3):230-236.

18.	Yamakawa T, Ayukawa T, & Matsuno K (2012) Metabolism and transportation pathways of GDP-fucose that are required for the O-fucosylation of Notch. *Advances in experimental medicine and biology* 727:37-46.

19.	Ginsburg V (1960) Formation of Guanosine Diphosphate l-Fucose from Guanosine Diphosphate d-Mannose. *Journal of Biological Chemistry* 235(8):2196-2201.

20.	Foster DW & Ginsburg V (1961) Biosynthesis of L-fucose by mammalian tissue. *Biochimica et biophysica acta* 54:376-378.

21.	Sullivan FX*, et al.* (1998) Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis in vitro. *The Journal of biological chemistry* 273(14):8193-8202.

22.	Chang S, Duerr B, & Serif G (1988) An epimerase-reductase in L-fucose synthesis. *Journal of Biological Chemistry* 263(4):1693-1697.

23.	Yurchenco PD & Atkinson PH (1975) Fucosyl-glycoprotein and precursor pools in HeLa cells. *Biochemistry* 14(14):3107-3114.

24.	Ishihara H, Massaro DJ, & Heath EC (1968) The Metabolism of l-Fucose : III. THE ENZYMATIC SYNTHESIS OF β-l-FUCOSE 1-PHOSPHATE. *Journal of Biological Chemistry* 243(6):1103-1109.

25.	Haltiwanger RS (2009) Fucose is on the TRAIL of colon cancer. *Gastroenterology* 137(1):36-39.

26.	Luhn K, Wild MK, Eckhardt M, Gerardy-Schahn R, & Vestweber D (2001) The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. *Nature genetics* 28(1):69-72.

27.   Hirschberg CB (2001) Golgi nucleotide sugar transport and leukocyte adhesion deficiency II. *The Journal of clinical investigation* 108(1):3-6.

28.   Song Z (2013) Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Molecular aspects of medicine* 34(2-3):590-600.

29.   Yakubenia S*, et al.* (2008) *Leukocyte trafficking in a mouse model for leukocyte adhesion deficiency II/congenital disorder of glycosylation IIc* pp 1472-1481.

30.   Ishikawa HO*, et al.* (2005) Notch deficiency implicated in the pathogenesis of congenital disorder of glycosylation IIc. *Proceedings of the National Academy of Sciences of the United States of America* 102(51):18532-18537.

31.   Hellbusch CC*, et al.* (2007) Golgi GDP-fucose transporter-deficient mice mimic congenital disorder of glycosylation IIc/leukocyte adhesion deficiency II. *The Journal of biological chemistry* 282(14):10762-10772.

32.   Ishikawa HO*, et al.* (2010) Two pathways for importing GDP-fucose into the endoplasmic reticulum lumen function redundantly in the O-fucosylation of Notch in Drosophila. *The Journal of biological chemistry* 285(6):4122-4129.

33.   Lu L, Hou X, Shi S, Korner C, & Stanley P (2010) Slc35c2 promotes Notch1 fucosylation and is required for optimal Notch signaling in mammalian cells. *The Journal of biological chemistry* 285(46):36245-36254.

34.   Wang X*, et al.* (2005) Dysregulation of TGF-beta1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 102(44):15791-15796.

35.   Shi S & Stanley P (2003) Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* 100(9):5234-5239.

36.   Okamura Y & Saga Y (2008) Pofut1 is required for the proper localization of the Notch receptor during mouse development. *Mechanisms of Development* 125(8):663-673.

37.   Du J*, et al.* (2010) O-fucosylation of thrombospondin type 1 repeats restricts epithelial to mesenchymal transition (EMT) and maintains epiblast pluripotency during mouse gastrulation. *Developmental biology* 346(1):25-38.

38.   Mollicone R*, et al.* (2009) Activity, splice variants, conserved peptide motifs, and phylogeny of two new alpha1,3-fucosyltransferase families (FUT10 and FUT11). *The Journal of biological chemistry* 284(7):4723-4738.

39.   Kumar A*, et al.* (2013) The Lewis X-related alpha1,3-fucosyltransferase, Fut10, is required for the maintenance of stem cell populations. *The Journal of biological chemistry* 288(40):28859-28868.

40.   Chen CY*, et al.* (2013) Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proceedings of the National Academy of Sciences of the United States of America* 110(2):630-635.

41.   Ihara H*, et al.* (2007) Crystal structure of mammalian α1,6-fucosyltransferase, FUT8. *Glycobiology* 17(5):455-466.

42.   Wang X*, et al.* (2009) Requirement of Fut8 for the expression of vascular endothelial growth factor receptor-2: a new mechanism for the emphysema-like changes observed in Fut8-deficient mice. *Journal of biochemistry* 145(5):643-651.

43.   Osumi D*, et al.* (2009) Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer science* 100(5):888-895.

44.   Zhao Y*, et al.* (2006) Deletion of core fucosylation on alpha3beta1 integrin down-regulates its functions. *The Journal of biological chemistry* 281(50):38343-38350.

45.   Kötzler MP, Blank S, Bantleon FI, Spillner E, & Meyer B (2012) Donor substrate binding and enzymatic mechanism of human core α1,6-fucosyltransferase (FUT8). *Biochimica et Biophysica Acta (BBA) - General Subjects* 1820(12):1915-1925.

46.   Varki A CR, Esko JD, et al. (2009) *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)) 2nd edition Ed.

47.   Flores A & Marrero JA (2014) Emerging trends in hepatocellular carcinoma: focus on diagnosis and therapeutics. *Clinical Medicine Insights. Oncology* 8:71-76.

48.   Mathew S*, et al.* (2014) Biomarkers for virus-induced hepatocellular carcinoma (HCC). *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 26C:327-339.

49.   Takahashi M, Kuroki Y, Ohtsubo K, & Taniguchi N (2009) Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydrate Research* 344(12):1387-1390.

50.   Shields RL*, et al.* (2002) Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *The Journal of biological chemistry* 277(30):26733-26740.

51.   Shinkawa T*, et al.* (2003) The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the

critical role of enhancing antibody-dependent cellular cytotoxicity. *The Journal of biological chemistry* 278(5):3466-3473.

52.   Yamane-Ohnuki N*, et al.* (2004) Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnology and bioengineering* 87(5):614-622.

53.   Okeley NM*, et al.* (2013) Development of orally active inhibitors of protein and cellular fucosylation. *Proceedings of the National Academy of Sciences of the United States of America*.

54.   Shao L, Moloney DJ, & Haltiwanger R (2003) Fringe Modifies O-Fucose on Mouse Notch1 at Epidermal Growth Factor-like Repeats within the Ligand-binding Site and the Abruptex Region. *Journal of Biological Chemistry* 278(10):7775-7782.

55.   Rampal R, Luther KB, & Haltiwanger RS (2007) Notch signaling in normal and disease States: possible therapies related to glycosylation. *Current molecular medicine* 7(4):427-445.

56.   Moloney DJ*, et al.* (2000) Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *The Journal of biological chemistry* 275(13):9604-9611.

57.   Rana NA & Haltiwanger RS (2011) Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Current opinion in structural biology* 21(5):583-589.

58.   Okajima T & Irvine KD (2002) Regulation of notch signaling by o-linked fucose. *Cell* 111(6):893-904.

59.   Sasamura T*, et al.* (2003) neurotic, a novel maternal neurogenic gene, encodes an O-fucosyltransferase that is essential for Notch-Delta interactions. *Development (Cambridge, England)* 130(20):4785-4795.

60.   Shi S & Stanley P (2003) Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proceedings of the National Academy of Sciences* 100(9):5234-5239.

61.   Rampal R, Arboleda-Velasquez JF, Nita-Lazar A, Kosik KS, & Haltiwanger RS (2005) Highly conserved O-fucose sites have distinct effects on Notch1 function. *The Journal of biological chemistry* 280(37):32133-32140.

62.   Yao D*, et al.* (2011) Protein O-fucosyltransferase 1 (Pofut1) regulates lymphoid and myeloid homeostasis through modulation of Notch receptor ligand interactions. *Blood* 117(21):5652-5662.

63.   Li M, *et al.* (2013) Mutations in POFUT1, encoding protein O-fucosyltransferase 1, cause generalized Dowling-Degos disease. *American journal of human genetics* 92(6):895-903.

64.   Schuster-Gossler K, Harris B, Johnson KR, Serth J, & Gossler A (2009) Notch signalling in the paraxial mesoderm is most sensitive to reduced Pofut1 levels during early mouse development. *BMC developmental biology* 9:6.

65.   Kim ML, *et al.* (2008) O-fucosylation of muscle agrin determines its ability to cluster acetylcholine receptors. *Molecular and cellular neurosciences* 39(3):452-464.

66.   Kielty CM, *et al.* (2002) Fibrillin: from microfibril assembly to biomechanical function. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 357(1418):207-217.

67.   Timpl R, Sasaki T, Kostka G, & Chu M-L (2003) Fibulins: a versatile family of extracellular matrix proteins. *Nat Rev Mol Cell Biol* 4(6):479-489.

68.   Moloney DJ, *et al.* (2000) Fringe is a glycosyltransferase that modifies Notch. *Nature* 406(6794):369-375.

69.   Bruckner K, Perez L, Clausen H, & Cohen S (2000) Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* 406(6794):411-415.

70.   Panin VM, *et al.* (2002) Notch ligands are substrates for protein O-fucosyltransferase-1 and Fringe. *The Journal of biological chemistry* 277(33):29945-29952.

71.   Haltiwanger RS & Stanley P (2002) Modulation of receptor signaling by glycosylation: fringe is an O-fucose-beta1,3-N-acetylglucosaminyltransferase. *Biochimica et biophysica acta* 1573(3):328-335.

72.   Xu A, *et al.* (2007) In vitro reconstitution of the modulation of Drosophila Notch-ligand binding by Fringe. *The Journal of biological chemistry* 282(48):35153-35162.

73.   Luther KB & Haltiwanger RS (2009) Role of unusual O-glycans in intercellular signaling. *The international journal of biochemistry & cell biology* 41(5):1011-1024.

74.   Takeuchi H & Haltiwanger RS (2014) Significance of glycosylation in Notch signaling. *Biochemical and biophysical research communications*.

75.   Lira-Navarrete E, *et al.* (2011) Structural insights into the mechanism of protein O-fucosylation. *PloS one* 6(9):e25365.

76.   Breton C, Snajdrova L, Jeanneau C, Koca J, & Imberty A (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology* 16(2):29R-37R.

77.   Okajima T, Xu A, Lei L, & Irvine KD (2005) Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding. *Science (New York, N.Y.)* 307(5715):1599-1603.

78.   Wang Y & Spellman MW (1998) Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *The Journal of biological chemistry* 273(14):8112-8118.

79.   Milde-Langosch K*, et al.* (2014) Prognostic relevance of glycosylation-associated genes in breast cancer. *Breast cancer research and treatment* 145(2):295-305.

80.   Yokota S*, et al.* (2013) Protein O-fucosyltransferase 1: a potential diagnostic marker and therapeutic target for human oral cancer. *International journal of oncology* 43(6):1864-1870.

81.   Luo Y, Nita-Lazar A, & Haltiwanger RS (2006) Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *The Journal of biological chemistry* 281(14):9385-9392.

82.   Luo Y, Koles K, Vorndam W, Haltiwanger RS, & Panin VM (2006) Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *The Journal of biological chemistry* 281(14):9393-9399.

83.   Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, & Majerus EM (2007) O-fucosylation is required for ADAMTS13 secretion. *The Journal of biological chemistry* 282(23):17014-17023.

84.   Wang LW*, et al.* (2007) O-fucosylation of thrombospondin type 1 repeats in ADAMTS-like-1/punctin-1 regulates secretion: implications for the ADAMTS superfamily. *The Journal of biological chemistry* 282(23):17024-17031.

85.   Hofsteenge J*, et al.* (2001) C-mannosylation and O-fucosylation of the thrombospondin type 1 module. *The Journal of biological chemistry* 276(9):6485-6498.

86.   Leonhard-Melief C & Haltiwanger RS (2010) O-fucosylation of thrombospondin type 1 repeats. *Methods in enzymology* 480:401-416.

87.   Sato T*, et al.* (2006) Molecular cloning and characterization of a novel human beta1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain. *Glycobiology* 16(12):1194-1206.

88.   Kozma K*, et al.* (2006) Identification and characterization of abeta1,3-glucosyltransferase that synthesizes the Glc-beta1,3-Fuc disaccharide on thrombospondin type 1 repeats. *The Journal of biological chemistry* 281(48):36742-36751.

89.	Heinonen TY & Maki M (2009) Peters'-plus syndrome is a congenital disorder of glycosylation caused by a defect in the beta1,3-glucosyltransferase that modifies thrombospondin type 1 repeats. *Annals of medicine* 41(1):2-10.

90.	Reis LM*, et al.* (2008) Mutation analysis of B3GALTL in Peters Plus syndrome. *American journal of medical genetics. Part A* 146A(20):2603-2610.

91.	Lesnik Oberstein SA*, et al.* (2006) Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase. *American journal of human genetics* 79(3):562-566.

92.	Chen CI*, et al.* (2012) Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. *The EMBO journal* 31(14):3183-3197.

93.	Rostovtsev VV, Green LG, Fokin VV, & Sharpless KB (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie (International ed. in English)* 41(14):2596-2599.

94.	Tornoe CW, Christensen C, & Meldal M (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *The Journal of organic chemistry* 67(9):3057-3064.

95.	Laughlin ST, Baskin JM, Amacher SL, & Bertozzi CR (2008) In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* 320(5876):664-667.

96.	Laughlin ST & Bertozzi CR (2009) In vivo imaging of Caenorhabditis elegans glycans. *ACS chemical biology* 4(12):1068-1072.

97.	Hsu TL*, et al.* (2007) Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proceedings of the National Academy of Sciences of the United States of America* 104(8):2614-2619.

98.	Dehnert KW*, et al.* (2012) Imaging the sialome during zebrafish development with copper-free click chemistry. *Chembiochem : a European journal of chemical biology* 13(3):353-357.

99.	Besanceney-Webler C, Jiang H, Wang W, Baughn AD, & Wu P (2011) Metabolic labeling of fucosylated glycoproteins in Bacteroidales species. *Bioorganic & medicinal chemistry letters* 21(17):4989-4992.

100.	Rabuka D, Hubbard SC, Laughlin ST, Argade SP, & Bertozzi CR (2006) A chemical reporter strategy to probe glycoprotein fucosylation. *Journal of the American Chemical Society* 128(37):12078-12079.

101.	Liu TW*, et al.* (2012) A chemoenzymatic approach toward the identification of fucosylated glycoproteins and mapping of N-glycan sites. *Glycobiology* 22(5):630-637.

102.   Sawa M, *et al.* (2006) Glycoproteomic probes for fluorescent imaging of fucosylated glycans in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 103(33):12371-12376.

103.   Zandberg WF, Kumarasamy J, Pinto BM, & Vocadlo DJ (2012) Metabolic inhibition of sialyl-Lewis X biosynthesis by 5-thiofucose remodels the cell surface and impairs selectin-mediated cell adhesion. *The Journal of biological chemistry* 287(47):40021-40030.

104.   Rana NA, *et al.* (2011) O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. *The Journal of biological chemistry* 286(36):31623-31637.

105.   Stanley P, Caillibot V, & Siminovitch L (1975) Selection and characterization of eight phenotypically distinct lines of lectin-resistant Chinese hamster ovary cell. *Cell* 6(2):121-128.

106.   Lin AI, Philipsberg GA, & Haltiwanger RS (1994) Core fucosylation of high-mannose-type oligosaccharides in GlcNAc transferase I-deficient (Lec1) CHO cells. *Glycobiology* 4(6):895-901.

107.   Yurchenco PD & Atkinson PH (1975) Fucosyl-glycoprotein and precursor polls in HeLa cells. *Biochemistry* 14(14):3107-3114.

108.   Liu T-W, *et al.* (2011) Functional expression of l-fucokinase/guanosine 5′-diphosphate-l-fucose pyrophosphorylase from Bacteroides fragilis in Saccharomyces cerevisiae for the production of nucleotide sugars from exogenous monosaccharides. *Glycobiology* 21(9):1228-1236.

109.   Chen CI, *et al.* (2012) Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. *The EMBO journal*.

110.   Xu X, *et al.* (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature biotechnology* 29(8):735-741.

111.   Rampal R, *et al.* (2005) Lunatic fringe, manic fringe, and radical fringe recognize similar specificity determinants in O-fucosylated epidermal growth factor-like repeats. *The Journal of biological chemistry* 280(51):42454-42463.

112.   Jiang H, *et al.* (2014) Monitoring dynamic glycosylation in vivo using supersensitive click chemistry. *Bioconjugate chemistry* 25(4):698-706.

113.   Clynes RA, Towers TL, Presta LG, & Ravetch JV (2000) Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. *Nature medicine* 6(4):443-446.

114. Arnold JN, Wormald MR, Sim RB, Rudd PM, & Dwek RA (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual review of immunology* 25:21-50.

115. Xue J, Zhu LP, & Wei Q (2013) IgG-Fc N-glycosylation at Asn297 and IgA O-glycosylation in the hinge region in health and disease. *Glycoconjugate journal* 30(8):735-745.

116. Lefranc M-PaL, Gérard (2001) *The Immunoglobulin FactsBook*.

117. Jensen PH, Karlsson NG, Kolarich D, & Packer NH (2012) Structural analysis of N- and O-glycans released from glycoproteins. *Nature protocols* 7(7):1299-1310.

118. Gudihal RaW, Keith (2010) Glycopeptide and glycan analysis of monoclonal antibodies using a microfluidic-based HPLC-Chip coupled to an Agilent Accurate-Mass Q-TOF LC/MS. *Agilent Technologies, Inc.*