Stony Brook University

Structural and Mechanistic Insights into Lunatic Fringe

A Dissertation Presented

by

Kelvin Brent Luther

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

August 2008

Stony Brook University

The Graduate School

Kelvin Brent Luther

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 co-Advisor
Professor and Interim Chair, Department of Biochemistry and Cell Biology

Dr. Hermann Schindelin - Dissertation co-Advisor
Professor, Rudolf-Virchow-Zentrum,
DFG-Forschungzentrum für Experimentelle Biomedizin
Universität Würzburg
Adjunct Professor, Department of Biochemistry and Cell Biology
Stony Brook University

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

Dr. William J. Lennarz
Distinguished Professor, Department of Biochemistry and Cell Biology
Director, Institute for Cell and Developmental Biology

Dr. Pamela Stanley
Professor, Department of Cell Biology
Horace W. Goldsmith Professor
Albert Einstein College of Medicine of Yeshiva University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the dissertation

Structural and Mechanistic Insights into Lunatic Fringe

by

Kelvin Brent Luther

Doctor of Philosophy

in
Molecular and Cellular Biology

Stony Brook University

2008

Disrupted Notch signaling causes lethality in an embryo and is implicated in many disease states postnatally. Interaction between Notch and ligands of the DSL family can be modulated by unusual O-fucose glycans on the tandem EGF-like repeats of the extracellular domain of Notch. The *O*-fucose can be further elongated to a tetrasaccharide beginning with N-acetylglucosamine (GlcNAc), added by one of three fringe β1,3-GlcNAc transferases (Lunatic (Lfng), Manic, or Radical) in mammals. Lfng is involved in vertebrate segmentation, and disruption of its function results in the human genetic disorder, Spondylocostal Dysostosis. A recent crystal structure for Mfng allowed us to produce a homology model for Lfng. Using the program HEX, we docked an *O*-fucosylated EGF repeat onto the Lfng model. We manually culled inappropriate solutions and found the *O*-fucose of the docked EGF clustered in two groups between the putative catalytic Aspartate 289 and the GlcNAc of the donor nucleotide sugar. Based on this model we chose residues in Lfng to mutate and analyzed their activity with an in vitro assay. We categorized the mutants as V_{max} or K_M defects and attempted to determine their role in enzyme function. Based on these results, we propose that one of the *O*-fucose clusters is the most likely orientation. We also have evidence that a small loop not observed in the crystal structure may become ordered upon substrate binding, closing one side of the catalytic pocket. We propose a model whereby the closing of this loop alters the conformation of the catalytic aspartate, increasing catalytic efficiency. This model may be generally applicable to the mechanism of glycosyltransferases outside the Fringe family. Finally, we have analyzed the donor nucleotide-sugar specificity of Lfng and have identified a number of residues responsible for UDP-GlcNAc specificity.

Table of Contents

List of Symbols

^{3}H	tritium
ADAM	A Disintegrin And Metalloprotease
ADAMTS	A Disintegrin And Metalloprotease with Thrombospondin
AJ	adherens junction
asp	aspartate
ARC	American Radiolabeled Chemicals
β3GlcT	beta-1,3-glucosyltransferase
β3GlcNAcT	beta-1,3-N-acetylglucosaminyltransferase
C	Celsius
CADASIL	Cerebral Autosomal Dominant Arteriopathy with Sub-cortical Infarcts and Leukoencephalopathy
CAZy	Carbohydrate Active enZymes
CD	circular dichroism
CD2	cluster of differentiation 2
CD36	cluster of differentiation 36
CDG	congenital disorder of glycosylation
CHO	Chinese hamster ovary
cm	centi-meter
CmanT	*C*-mannosyltransferase
C-terminus	carboxy-terminus
dol-P-man	dolichol-P-mannose
DEL	*Drosophila* Delta
Dll	mammalian Delta-like
DMEM	Dulbecco's modification of Eagle's medium
DMSO	dimethyl-sulfoxide
DNA	deoxy-ribonucleic acid
DSL	Delta/Serrate/Lag-2
ECD	extracellular domain
EDTA	ethylene-diamine tetra-acetic acid
EGF	epidermal growth factor-like
EGF-*O*-fucose	*O*-fucosylated epidermal growth factor-like repeat
endo-free	endotoxin-free
EndoH	endoglycosidase H
EPOR	erythropoietin receptor
ER	endoplasmic reticulum
ES	embryonic stem
et al.	*et alia*
Fng	fringe
Fuc	fucose

g	acceleration by gravity
Gal	galactose
GDP	guanosine-5'-diphosphate
Glc	glucose
GlcA	glucuronate
GlcNAc	N-acetylglucosamine
glu	glucose
Gmd	GDP-mannose 4,6 dehydratase
GnT1	N-acetylglucosaminyltransferase I
Golgi	Golgi body
GT	glycosyltransferase
GTS	Glycosyltransferase Super-family
HCl	hydrochloric acid
HEK293T	human embryonic kidney 293T cells
HEK293 TREx	human embryonic kidney 293TREx cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hes7	hairy enhancer of split-7
HMM	hidden Markov model
HPAEC	high pH anion-exchange chromatography
HPLC	high performance liquid chromatography
ICD	intracellular domain
IgG	immuno-globulin G
IL12	interleukin 12
J1	Jagged 1
J2	Jagged 2
Jag	Jagged
JNK	C-jun N-terminal kinase
kb	kilobase
k_{cat}	catalytic constant
kDa	kilo-Dalton
K_i	inhibition constant
K_M	Michaelis constant
L	Liter
Lfng	lunatic fringe
LPS	lipopolysaccharide
M	molar
Man	mannose
µCi	micro-Curie
MD	molecular dynamics
Mesp2	mesoderm posterior 2
Mfng	manic fringe
µg	micro-gram
mg	milli-gram
min	minute
µL	micro-Liter
mL	milli-Liter

µm	micro-meter
µM	micro-Molar
mM	milli-Molar
µmol	micro-mole
mmol	milli-mole
$MnCl_2$	manganese chloride
MQ	MilliQ purified
N1	Notch 1
N2	Notch 2
N^{12f}	notch with the *O*-fucose site at EGF-repeat 12 mutated to alanine
NaCl	sodium chloride
NaN_3	sodium azide
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NFκB	nuclear factor kappa B
ng	nano-gram
nm	nano-meter
nM	nano-Molar
Ni-NTA	Ni^{2+}-nitrilotriacetic acid
N-linked	asparagine-linked
nmol	nano-mole
NMR	nuclear magnetic resonance
NP-40	nonidet P-40
N-terminus	amino-terminus
OFut1	*Drosophila* protein *O*-fucosyltransferase 1
$OFut1^{R245A}$	*Drosophila* protein *O*-fucosyltransferase 1 with arginine 245 to alanine mutation
p	*para*
PAD	pulsed amperometric detection
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
Pfu	*Pyrococcus furiosus*
pH	negative logarithm of the proton concentration
pM	pico-Molar
pmol	pico-mole
PMP-C	*Pars intercerebralis* major peptide C
PNGase F	peptide *N*-glycanase F
pNP	*para*-nitrophenyl
Pofut1	mammalian protein *O*-fucosyltransferase 1
Pofut2	mammalian protein *O*-fucosyltransferase 2
Poglut	mammalian protein *O*-glucosyltransferase
PSM	pre-somitic mesoderm
Rfng	radical fringe
RMSD	root mean squared deviation
RNAi	RNA interference
RNAse	ribonuclease

s	second
S2	Schneider 2
SAC	sub-apical compartment
SCD	spondylocostal dysostosis
SCO	sub-commissural organ
SDS	sodium dodecyl sulfate
SER	serrate
Sia	sialic acid
TAK1	transforming growth factor beta-activated kinase 1
TFA	trifluoro-acetic acid
TGF-β	transforming-growth factor beta
TNF-α	tumor necrosis factor alpha
Tris	trishydroxymethylaminomethane
TSR	thrombospondin type-1 repeat
Tween	polyoxyethylene (20) sorbitan monolaurate
U	units
UDP	uridine-5'-diphosphate
UDP-Glc	uridine-5'-diphospho-glucose
UDP-GlcNAc	uridine-5'-diphospho-N-acetylglucosamine
UMP	uridine-5'-monophosphate
V_{max}	maximal velocity
WT	wild-type
$^{v}/_{v}$	volume over volume
$^{w}/_{v}$	weight over volume
Xyl	xylose
XylT	xylosyltransferase
$ZnCl_2$	zinc chloride

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my co-Advisors Bob Haltiwanger and Hermann Schindelin for agreeing to take me into their labs and serve as my mentors. Even when the project was not progressing as it should I received only encouragement and support which is appreciated far more than I can say. You have both enriched my scientific understanding immeasurably and have been instrumental in making these years enjoyable, even when project related frustration threatened to mar the experience. I could not have chosen better mentors for my graduate education.

Thank-you to Erwin London for first of all giving me a place to do one of my rotations and teaching me a great deal in a few short months. I had the enviable position of wanting to work in the Haltiwanger, London and Schindelin labs after my rotations and not knowing where to go until rescued by a joint project. Thank-you for making that choice so hard. Thank-you for agreeing to be on my committee and your invaluable contributions to my graduate education in that role.

Thank-you to Bill Lennarz for being a member of my committee and always having nothing but words of encouragement. You have created an environment of collegiality in the department that made working here a wonderful experience. There is no way to overstate how fundamental a contribution this has been to my success, and the success of all of the students, post-docs, and faculty in Biochemistry and Cell Biology. I was doubly blessed to have your contributions and insights during committee meetings and throughout my time spent here.

Thank you to Pamela Stanley for agreeing to serve as the external member of my committee. Thank-you for all of the comments, suggestions, and insights you have provided me over the years during Fringe meetings and in front of my posters at the Glycobiology conferences. Thank-you for making the trip to attend my defense.

Thank-you to Neta Dean who served on my committee for the majority of my time here and made many insightful and helpful contributions in that role. I could always count on your friendly inquiry into how things were going to let me know that you would contribute in any fashion you could to help make my project a success.

Thank-you to Caroline Kisker for serving in unofficial capacity as yet another mentor. Everyone from the Schindelin and Kisker labs enjoyed the unique advantage of having the insights and encouragement of both Hermann and Caroline throughout their graduate school experience. I am one of only a very few students to have been lucky enough to have your support in addition to two advisors.

Thanks to Dr. Gerald Hart at Johns Hopkins University for a very generous gift of a large amount of UPD-hexanolamine agarose for the Lfng project. Thanks to Debbie Brown for many questions answered over the years, and to Jamie Konopka who almost indoctrinated me into the world of yeast biology on the lower floors of the building. Thanks to Greg Caputo, Fred Grau, Mike Lake, Bill Parrish, Mike Rudolph, and Amy Czura who made significant contributions to enjoyable rotations in the Haltiwanger,

Konopka, London, and Schindelin labs. Thanks to Liqun Wang for all of the plasmids you prepared that contributed to my projects. Thanks to the Hofsteenge, Ferry, and Majerus labs for contributions to successful collaborations and in particular Krisztina Kozma and Sarah Lawrence. Thanks also to Kanagalaghatta Rajashankar for contributing greatly to my first opportunity to work with a crystal structure by providing a MAD data set for the phosphotransacetylase project. Thanks to Sanford Simon for many illuminating conversations about enzyme kinetics and for reading and commenting on the Lunatic Fringe manuscript. Thank-you to Todd Miller for reading the Lunatic Fringe paper and helping us to bring some much needed focus to the manuscript. Thanks to Hongshik Ahn for helpful discussions.

Thanks to all of the students and post-docs in the department who have provided friendship and help over the years. In particular of course, a thank-you goes to all of the members of the Haltiwanger, Kisker, and Schindelin labs. Thanks in particular to Raajit Rampal, Ola Nita-Lazar, Mihai Nita-Lazar, Maggie Gabanyi, and Megha for friendship and encouragement over the years.

Thanks to Carol Juliano, Janet Koenig, and Claire Foley in the MCB office and Eileen Dowd in the CSB for providing me with help whenever I needed it. Thanks to Pilar Bartley for being a great landlord and a friend and providing me with a place to live that became a home.

Most importantly and fundamentally, I want to thank my family. Thanks to my sisters and their families for support and encouragement and making being this far from home harder. Thanks most to my parents for providing the environment that allowed me to learn as much as I wanted and become interested in so many topics as a child. With hindsight, I now know that they started and then nurtured this journey into the world of Academia even if they don't realize it. It goes without saying that without their unqualified support none of this would have happened.

I would like to acknowledge the NIH for funding that allowed this research to be conducted.

Introduction

Protein Glycosylation

Glycosylation of proteins has been observed in all organisms from bacteria to man (Wacker, *et al.*, 2002, Young, *et al.*, 2002). Many functions for glycans have been proposed (reviewed by (Varki, 1993)), from stabilization of proteins (Imperiali and Rickert, 1995), to antigenic masking (Wyatt, *et al.*, 1998), to direct functional consequences on ligand binding (Bruckner, *et al.*, 2000, Moloney, *et al.*, 2000a) to name just a few. In the last 15 years, knowledge of the biological relevance and the impact of glycosylation on human disease has ballooned. The dramatic and continuing increase in the number of identified congenital disorders of glycosylation (CDG), all of which result directly from altered glycan structures is one example (reviewed by (Freeze and Aebi, 2005)). Genetic ablation of many glycosyltransferase (GT) enzymes show embryonic lethal phenotypes. This includes protein *O*-fucosyltransferase 1 (Pofut1) in mice (Shi and Stanley, 2003) and flies (Okajima and Irvine, 2002, Sasamura, *et al.*, 2003), protein *O*-fucosyltransferase 2 (Pofut2) (Haltiwanger and Holdener, manuscript in preparation), N-acetylglucosaminyltransferase 1 (GnT1) in mice (Ioffe and Stanley, 1994), and Rumi, a protein *O*-glucosyltransferase (Poglut) in flies (Acar, *et al.*, 2008) as just a few examples. Structurally, there are two main classes of protein glycosylation depending on whether the chain is initiated by formation of an amide bond with an asparagine (*N*-linked), or *O*-linked through a hydroxyl bearing side-chain (serine, threonine, hydroxy-lysine) (Varki, *et al.*, 1999). This introduction will focus mainly on the structures of various *O*-linked glycans on cysteine-rich protein motifs, and the effects they exert on proteins involved in intracellular signaling. I will also touch on the potential effects on signaling of a rare form of glycosylation resulting in carbon-carbon bond formation between a sugar and a tryptophan, known as *C*-mannosylation.

Glycosylation of EGF-like Repeats

Epidermal Growth Factor-like (EGF)-repeats are small (~40 amino acid) cysteine-rich motifs with six conserved cysteines forming three conserved disulfide bonds (Appella, *et al.*, 1988, Davis, 1990). EGF-repeats containing an appropriate consensus sequence (C^2-$X_{(4-5)}$-[S/T]-C^3) between the second and third cysteine will be co- and/or post-translationally modified by fucose (Fuc) on a serine or threonine during their transit through the endoplasmic reticulum (ER) by Pofut1 (OFut1 in *Drosophila*) (Figure 1A) (Luo and Haltiwanger, 2005, Shao, *et al.*, 2003, Wang, *et al.*, 2001). Both OFut1 and Pofut1 contain an ER localization signal. The Fuc can be further elongated in the Golgi body (Golgi) by one of the Fringe (Fng) β1,3N-acetylglucosaminyltransferases (β3GlcNAcT) to form a disaccharide (Bruckner, *et al.*, 2000, Moloney, *et al.*, 2000a). While there does not appear to be elongation past the disaccharide in flies

Figure 1

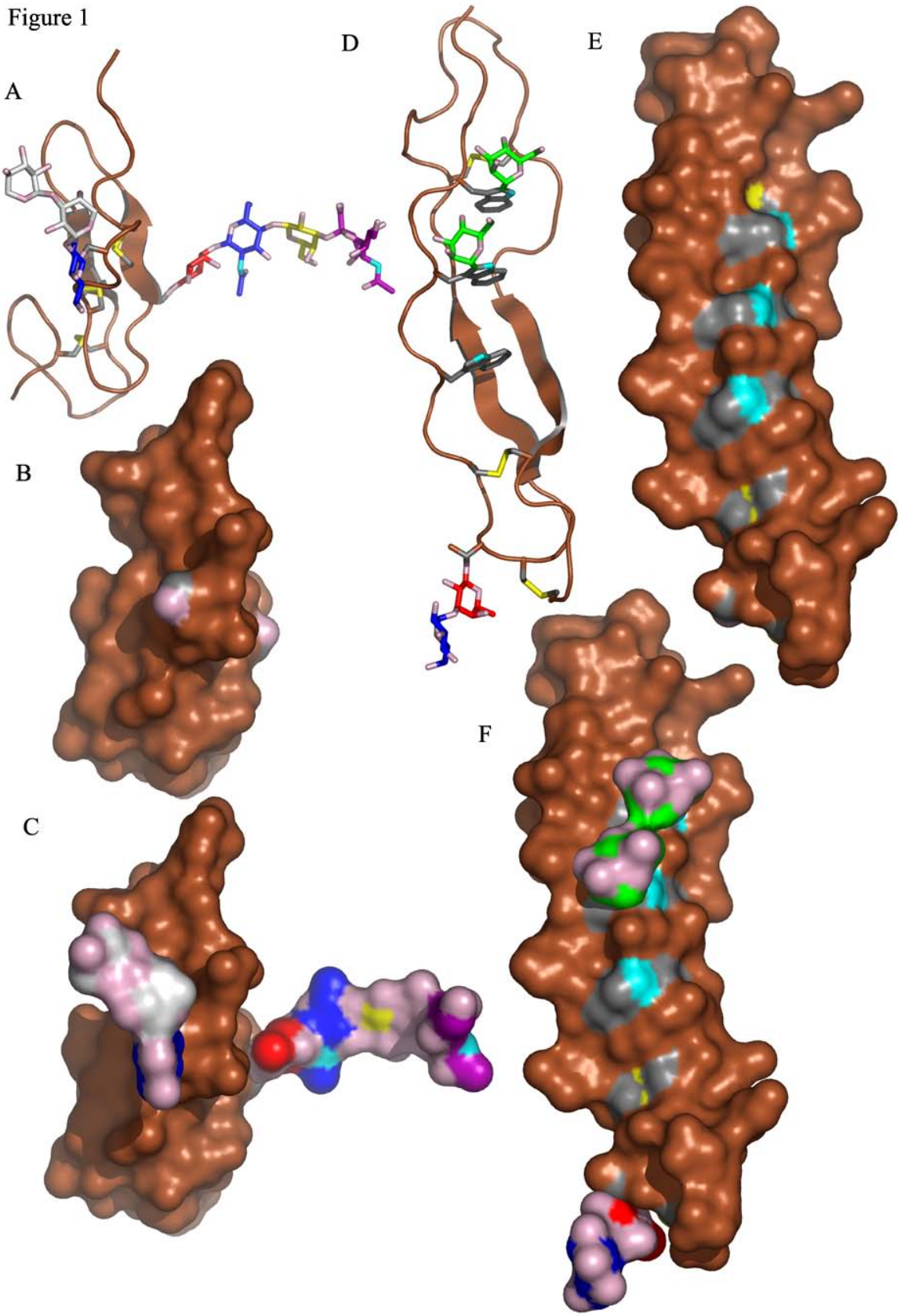


Figure 1. *O*-Glycan Containing Epidermal Growth Factor-like and Thrombospondin Type-1 Repeats

The carbons of sugar residues are colored according to the Consortium for Functional Glycomics (http://www.functionalglycomics.org) recommended nomenclature. The carbon atoms of Fucose are red, glucose and N-acetylglucosamine are blue, galactose is yellow, xylose is white, sialic acid is purple, and mannose is green. Oxygens are colored pink, nitrogen is cyan, carbons of amino acid side-chains involved in disulfide bonding or glycosylation are grey, and the sulfur of cysteines are colored yellow. All other atoms are brown. **A)** Cartoon representation of fully glycosylated Factor IX EGF-repeat containing both an *O*-fucose tetrasaccharide and an *O*-glucose trisaccharide. The tetrasaccharide containing EGF-repeat was described previously (Luo, 2007). The *O*-glucose trisaccharide was modeled by taking a Xyl-α1,3-Xyl-α1,3-Glc trisaccharide solution structure model from the Sweet website (http://www.glycosciences.de/modeling/sweet2/doc/index.php) and adding the modification to the appropriate serine. The serine side-chain was then cycled through known rotamer conformations and the position judged most appropriate by eye was chosen. **B)** Surface representation of panel A without glycosylation. **C)** Surface representation of panel A. **D)** Cartoon representation of a fully glycosylated TSR containing a Glc-β1,3-Fuc disaccharide and two mannosylated tryptophans. The disaccharide was modeled by taking a Glc-β1,3-Fuc solution structure model from the Sweet website, and aligning the fucose from the disaccharide with the fucose of TSR3 from the structure of the Thrombospondin 1 TSRs (pdb accession 1LSL) (Tan, *et al.*, 2002). The mannose on the tryptophans was modeled by superimposing the planar ring of 4-nitrophenyl-α-D-mannose with the tryptophan ring, removing the oxygen on C1 of the mannose, and placing the mannose at an approximated carbon bond length from the appropriate indole carbon. **E)** Surface representation of panel D without glycosylation. **F)** Surface representation of panel D.

(Xu, *et al.*, 2007), in mammals, once the disaccharide is formed, the glycan is further elongated to a tetrasaccharide by addition of a galactose (Gal) and a sialic acid (Sia) (Chen, *et al.*, 2001, Moloney, *et al.*, 2000b).

In addition to the presence of *O*-fucose glycans, an *O*-glucose glycan, will be added to a serine by Poglut when the appropriate consensus sequence (C^1-X-S-X-P-C^2) is present between the first and second cysteine of the EGF-repeat (Figure 1A) (Acar, *et al.*, 2008, Moloney, *et al.*, 2000b). Poglut appears to be localized in the ER as well (Acar, *et al.*, 2008). The *O*-glucose glycan can be further elongated by consecutive addition of two xylose (Xyl) sugars (Hase, *et al.*, 1990). The addition of these Xyl sugars is effected by two separate, as yet uncloned, xylosyltransferase (XylT) enzymes (Minamida, *et al.*, 1996, Omichi, *et al.*, 1997). The intracellular localization of xylosylation is likewise unclear at this point.

Glycosylation of Thrombospondin Type 1 Repeats

Similar to the EGF-repeat, a second type of cysteine-rich motif known as a thrombospondin type 1 repeat (TSR) contains six conserved cysteines, and three disulfide bonds (Adams and Tucker, 2000). Unlike the EGF-repeat however, there is some flexibility in the disulfide-bonding pattern in the TSR resulting in two structural groups termed class 1 and 2 (Tan, *et al.*, 2002, Tossavainen, *et al.*, 2006, Tucker, 2004). The difference in disulfide bonding pattern results from one cysteine being present on separate, neighboring strands in the two classes, although the overall fold is very similar between the two (Tossavainen, *et al.*, 2006). Two unusual forms of glycosylation are known to occur on TSRs. A homologue of Pofut1 known as Pofut2 fucosylates TSRs (Luo, *et al.*, 2006a, Luo, *et al.*, 2006b) on a serine or threonine between the first and second conserved cysteine in the consensus sequence (C^1-S-X-[S/T]-C^2-G) (Hofsteenge, *et al.*, 2001) (Figure 1B). This fucosylation event also appears to occur in the ER despite the lack of an ER localization sequence in the Pofut2 sequence (Luo, *et al.*, 2006a). The *O*-fucose can then be elongated by a β1,3-glucosyltransferase (β3GlcT) to a disaccharide (Kozma, *et al.*, 2006, Sato, *et al.*, 2006). In addition to *O*-fucose, tryptophans in the consensus sequence (W-X-X-W) can be *C*-mannosylated (Figure 1B) (Hofsteenge, *et al.*, 2001). This glycosylation is performed by an enzyme in the ER membrane using dolichol-P-mannose (dol-P-man) as a donor substrate (Doucey, *et al.*, 1998, Hofsteenge, *et al.*, 1994). This is an extremely unusual form of glycosylation in that it involves a carbon-carbon bond formed between the tryptophan and the mannose (Man). Additionally, it is not known whether both tryptophans can be modified in a single consensus sequence.

Importance of Epidermal Growth Factor-like Repeat Glycosylation in Signaling

Notch Signaling

Notch (N) was first reported as an X-linked lethal trait in *Drosophila* (Morgan, 1917) and was ultimately identified as a membrane anchored signaling receptor (Wharton, *et al.*, 1985) with four homologues in mammals. The Notch pathway is important throughout development. Defects in Notch pathway components lead to

development related disorders such as spondylocostal dysostosis (SCD) (Bulman, *et al.*, 2000, Sparrow, *et al.*, 2006), cerebral autosomal dominant arteriopathy with sub-cortical infarcts and leukoencephalopathy (CADASIL) (Arboleda-Velasquez, *et al.*, 2005, Joutel, *et al.*, 1996), Alagille syndrome (Krantz, *et al.*, 1999, Li, *et al.*, 1997), and tetralogy of Fallot (Eldadah, *et al.*, 2001, Krantz, *et al.*, 1999) in humans. Apart from its role in development, the Notch pathway is important for a myriad of cell-fate decisions throughout life including regulation of oligodendrocyte precursor cell decisions (John, *et al.*, 2002) and T-cell development (reviewed by (Radtke, *et al.*, 2004)) as just two out of dozens of examples potentially related to disease. Disruption of Notch signaling during neuronal development in flies leads to a classic neurogenic Notch phenotype. Normally, Notch signaling, through a process called lateral inhibition, would prevent more than one cell in a cluster of neuronal precursors from becoming a neuron. Disrupted Notch signaling leads to an overproliferation of neurons, thus termed a neurogenic phenotype. Not surprisingly, considering the pivotal role of Notch in developmental and cell lineage processes, aberrant Notch signaling contributes to many diseases including cancers, and multiple sclerosis (for a review see (Rampal, *et al.*, 2007)).

Notch becomes activated upon interaction with membrane anchored ligands of the Delta/Serrate/Lag-2 (DSL) family on adjacent cells. The DSL ligands are subdivided into Delta (DEL) (Delta-like (Dll) 1, 3 and 4 in mammals) and Serrate (SER) (Jagged (Jag) 1 and 2 in mammals) subfamilies (for a recent review see (Bray, 2006)). In mammals, the nascent Notch receptor is subject to cleavage by a furin-like protease (S1 cleavage) in the Golgi. This creates a Notch heterodimer with the extracellular domain (ECD) tethered to a transmembrane-intracellular domain (ICD). The ECD is composed mainly of tandem EGF-repeats, with three Notch/Lin12 repeats at the C-terminus forming a negative regulatory region (see Figure 2). S1 processing does not appear to occur in *Drosophila* (Kidd and Lieber, 2002). Ligand binding induces a proteolytic event (termed S2 cleavage) where the majority of the extracellular portion of Notch is released from the cell surface. S2 cleavage can be catalyzed by either A Disintegrin And Metalloprotease (ADAM) 10 or ADAM17 in mammals, while Kuzbanian performs the cleavage in *Drosophila* (Mumm, *et al.*, 2000). Recent structural studies suggest that the S2 cleavage site is masked by the Notch/Lin12 domains in the resting state. Thus, ligand binding must induce a conformational rearrangement in the extracellular domain of Notch in order for S2 cleavage to occur (Gordon, *et al.*, 2007) (see Figure 2A-C). The ECD is then endocytosed with the ligand by the signal sending cell (Parks, *et al.*, 2000). After release of the Notch ECD, a third cleavage (S3) by the presenilin/γ-secretase complex occurs just inside the inner leaflet of the membrane, releasing the ICD (De Strooper, *et al.*, 1999), which transits to the nucleus and regulates expression of downstream targets (see Figure 2C and E). There are various intracellular modulators of Notch signaling including E3 ubiquitin ligases, E3 ligase inhibitors, coactivators, corepressors, DNA binding proteins, and signaling inhibitors (reviewed by (Bray, 2006, Stanley, 2007)).

Fringe Modulates Notch by Elongating *O*-Fucose on EGF Repeats

The importance of *O*-fucosylation for Notch function was first realized when the Fng family of genes were demonstrated to catalyze addition of N-acetylglucosamine (GlcNAc) to *O*-fucose on the EGF repeats of the Notch ECD (Bruckner, *et al.*, 2000, Moloney, *et al.*, 2000a). Fringe was originally described in *Drosophila* for its role in

Figure 2

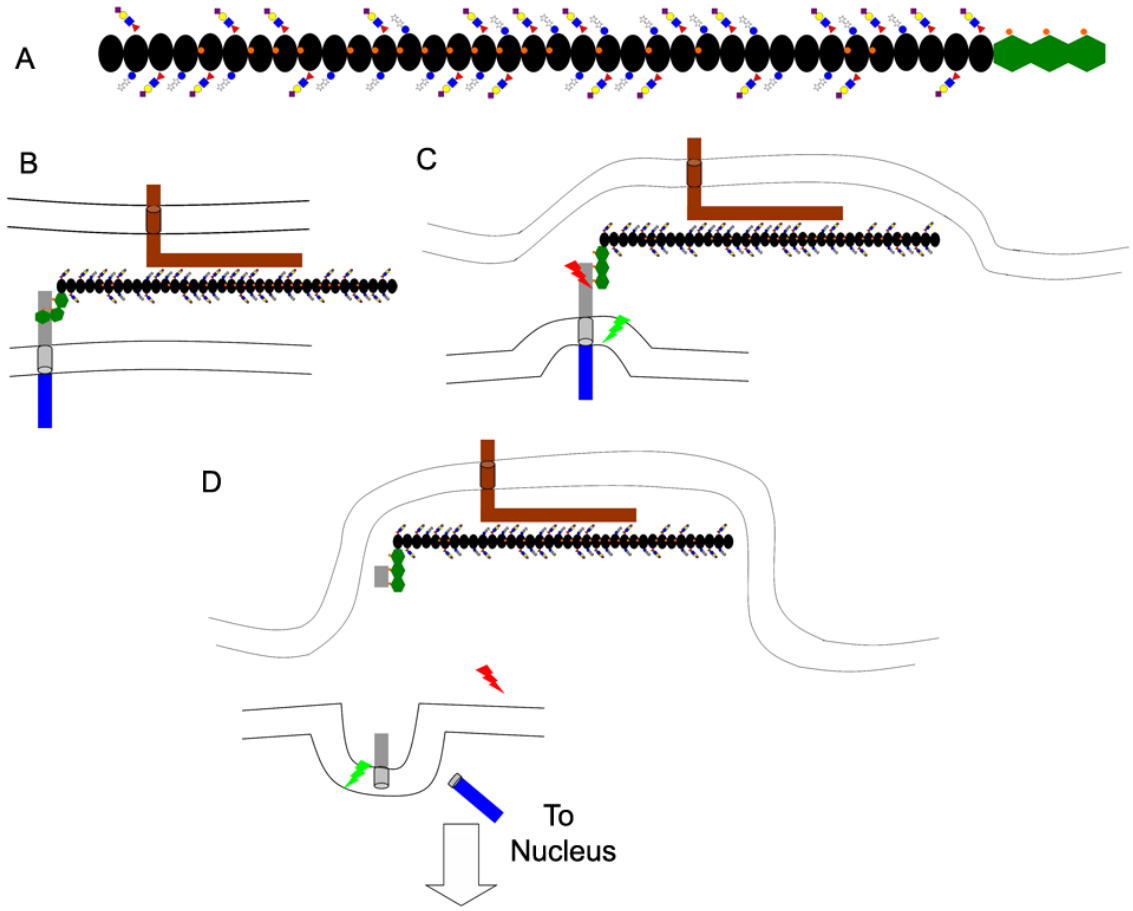


Figure 2. Model of Notch Receptor Activation
The Notch receptor EGF-repeats are indicated with black ovals, the Notch/Lin12 repeats of the negative regulatory region as green hexagons. The transmembrane portion of Notch is in grey. Intracellular Notch is shown in abbreviated fashion in blue. A simplified Delta ligand is shown on an adjacent cell in brown. The glycans on Notch are indicated with the suggested shapes and colors of the Consortium for Functional Glycomics (http://www.functionalglycomics.org). Glucose is a blue circle, N-acetylglucosamine a blue square, fucose a red triangle, galactose a yellow circle, sialic acid a purple diamond, and xylose a white star. Calcium ions are shown as orange spheres associated with EGF-repeats and the Notch/Lin12 repeats. ADAM10/ADAM17 is shown as a red lightning bolt. The presenilin/γ-secretase complex is shown as a bright green lightning bolt. **A)** Diagram of the Notch EGF-repeats. The glycan modifications and calcium binding EGF-repeat pattern is based on human Notch 1. Note that the glycans on consecutive EGF-repeats will be pointing out from the page on one EGF-repeat and pointing into the page on the neighboring EGF-repeat which is why the *O*-fucose tetrasaccharide and *O*-Glucose trisaccharide glycans switch sides on neighboring EGF-repeats in the diagram (see Figure 1A, C). The diagram indicates a maximal level of glycosylation at all sites. Mass spectral analysis suggests that this is not the case (Haltiwanger lab unpublished results, (Nita-Lazar and Haltiwanger, 2006a, b, Xu, *et al.*, 2007), but that it varies depending on Notch subtype, species, and cell type. The *in vivo* glycosylation pattern could involve monosaccharide *O*-fucose or *O*-glucose on some repeats. **B)** The Delta ligand interacts with the Notch receptor on a neighboring cell. **C)** Subsequent to Delta/Notch interaction, Delta begins to be endocytosed by the sending cell. This causes a conformational change in the Notch/Lin12 repeats exposing the S2 cleavage site. **D)** After S2 cleavage, the presenilin/γ-secretase complex cleaves the remaining transmembrane Notch inside the inner leaflet of the membrane. The Notch intracellular domain transits to the nucleus to transactivate downstream targets.

boundary formation (Irvine and Wieschaus, 1994). Although Notch and its ligands are expressed throughout the imaginal disc, Notch is only activated at the dorsal/ventral boundary. Fringe functions to position Notch activation at the boundary by inhibiting Notch activation from SER while activating activation from DEL (for a review see (Irvine and Rauskolb, 2001)). Fringe performs a similar positioning of Notch activation during development of limbs and in the eye (for a review see (Irvine, 1999)). Vertebrates have a family of three Fng enzymes named Lunatic (Lfng), Manic (Mfng) and Radical Fringe (Rfng). Knocking out Lfng in mice produces segmentation and somitogenesis defects (Evrard, *et al.*, 1998, Zhang and Gridley, 1998, Zhang, *et al.*, 2002). The Lfng homozygous mutants have severely disrupted skeletal patterning, with shortened, missing, fused, and bifurcated ribs, while some ribs were disconnected from the vertebrae (Zhang and Gridley, 1998). The animals exhibited shortened tails and disrupted vertebral patterning (Evrard, *et al.*, 1998, Zhang and Gridley, 1998) due to effects on the Notch dependent process of somitogenesis in these mutants. Some homozygotes die post-natally, apparently due to respiratory problems from their malformed rib cages, although most survive into adulthood (Zhang and Gridley, 1998). Disruption of Lfng results in altered patterns of expression of the Notch pathway genes Dll1, Dll3, Notch 1 (N1), and Notch 2 (N2) in the developing somites. The Rfng knockout did not show any obvious phenotype (Zhang, *et al.*, 2002). In order to test the possibility of a synergistic effect between Lfng and Rfng, mouse double knockouts were produced with no obvious differences between the double mutant and the Lfng homozygous mutant mice (Zhang, *et al.*, 2002). There have been no reports of a Mfng knockout.

The fact that Rfng knockouts show no obvious phenotype raises the question of functional redundancy for the Fng enzymes. As for their catalytic activity, all three mammalian Fng enzymes catalyze the transfer of GlcNAc to *O*-fucose on EGF-repeats in *in vitro* assays, although with differing efficiency (Rampal, *et al.*, 2005b). Lfng and Rfng both modified a Factor IX EGF-repeat with similar efficiency, while Mfng was three-fold less efficient with this substrate (Rampal, *et al.*, 2005b). In the case of EGF-repeat 26 from mouse N1 however, Lfng activity was more than six-fold greater than Rfng and more than one hundred and fifty fold greater than Mfng (Rampal, *et al.*, 2005b). Comparison of the sequences of the EGF-repeats from human Factor IX (a good substrate for Lfng) and Factor VII (a poor substrate for Lfng) revealed two amino acids which are necessary for efficient Lfng recognition (Rampal, *et al.*, 2005b). Mutation of these two amino acids in Factor IX EGF-repeat to the corresponding amino acids from Factor VII EGF-repeat significantly reduced recognition by Lfng. Performing the reverse (conversion of the two amino acids in Factor VII EGF-repeat to the corresponding residues in Factor IX EGF-repeat) did not cause Factor VII EGF-repeat to become as good a substrate as Factor IX EGF-repeat, indicating that they are not sufficient for Lfng recognition (Rampal, *et al.*, 2005b). These two amino acids were important for recognition of the EGF-repeat by all three Fng enzymes in *in vitro* assays, although to varying degrees. These patterns seem to indicate that while the Fng enzymes all catalyze the same reaction, their specificity for any given EGF-repeat is significantly different. The three enzymes may have evolved optimum specificities for different substrates and may be promiscuous to varying degrees with sub-optimal substrates. These data suggest the Fng enzymes are not functionally redundant. More detailed analyses of the knockout mice, and evaluation of a Mfng knockout, are necessary to resolve these issues.

Role of Lunatic Fringe During Somitogenesis

The formation of vertebrate somites is regulated by waves of gene expression termed the somitogenesis clock (for recent reviews see (Andrade, *et al.*, 2007, Kageyama, *et al.*, 2007)). The Notch pathway is essential via cell-to-cell communication for synchronizing the waves of gene expression originating in the presomitic mesoderm (PSM). These unidirectional waves of expression originate in the caudal PSM and travel to the rostral PSM where they eventually terminate when encountering mesoderm posterior 2 (Mesp2) expression (Andrade, *et al.*, 2007, Morimoto, *et al.*, 2005). This sets the boundary for the developing somite (Andrade, *et al.*, 2007). Lfng turns Notch on and off in this context, and its expression is regulated in concert with periodic somite formation, and as such, is now recognized as an integral component of the somitogenesis clock. In computer simulations, Zhu and Dhar showed that transient blockage of Notch signaling involving N1, Lfng and hairy enhancer of split (Hes)7, a downstream target of Notch, could maintain a unidirectional wave of signaling (Zhu and Dhar, 2006). Hes7 negatively regulates its own expression and that of Lfng (Chen, *et al.*, 2005) with the result that Hes7 and Lfng transcription is initiated, the proteins are expressed, and then Hes7 and Lfng transcription is eliminated resulting in waves of expression (reviewed in (Kageyama, *et al.*, 2007)). Disruption of the clock by elimination or misexpression of Lfng results in somitogenesis defects.

Lunatic Fringe knockout mice share a striking resemblance to mice lacking Dll3, suggesting that disruption of these genes has similar effects *in vivo* (Kusumi, *et al.*, 1998). The Dll3 ligand has significant sequence divergence from the other ligands. Dll3 has recently been shown to negatively regulate Notch signaling in a cell autonomous manner, while being incapable of activating the Notch signaling pathway (Ladi, *et al.*, 2005). This is in contrast to the other ligands which have been shown to activate Notch signaling. Dll3 was shown to bind to Notch only when expressed in the same cell, and Lfng was able to reverse the inhibition (Ladi, *et al.*, 2005). Lfng reversal of Dll3 inhibition appears to result solely from the ability to increase signaling through Dll ligands since coexpression of Lfng in the Notch/Dll3 expressing cell had no effect on the Dll3/N interaction (Ladi, *et al.*, 2005). Thus both Lfng and Dll3 appear to function to regulate Notch activation during somite formation. The genetic disease SCD in humans is caused by mutations in Dll3 (Bulman, *et al.*, 2000) Lfng (Sparrow, *et al.*, 2006), or Mesp2 (Whittock, *et al.*, 2004), presumably due to the effects these Notch pathway components have on somitogenesis during development (for a review see (Sparrow, *et al.*, 2007)).

Mechanisms For the Effects of Fringe on Notch Function

The majority of data suggests that Fng-mediated elongation of *O*-fucose on Notch results in a change in the binding between Notch and its ligands. This is clearly the case in the *Drosophila* system. The *in vivo* effects of Fng can be recapitulated using purified components *in vitro* (Xu, *et al.*, 2007). Addition of GlcNAc to *O*-fucose using Fng causes an increase in DEL binding and a decrease in SER binding (Table 1) (Xu, *et al.*, 2007). Although the data using mammalian components also suggests that Fng modulates the binding between Notch and its ligands, the results are complicated by the increased number of receptors, ligands, and Fng enzymes. Weinmaster and coworkers found that with N1, Lfng increased signaling from Dll1 and inhibited Jag1 mediated

Table 1. Results of Fringe Effects on Various Notch/Ligand Cell-based Signaling and Binding Assays

Dll1: Delta-like1, J1: Jagged1, J2: Jagged2, N1: Notch1, N2: Notch2, DEL: Delta, SER: Serrate. Arrows indicate an increase or decrease in signaling through the indicated ligand or changes in measured binding. n.c. indicates no change. The bottom two rows are *Drosophila* experiments while the remainder of the table refers to data from mammalian systems. The table contains data from references where the complement of Notch pathway components is clearly known except for the studies in CHO cells, where endogenous Notch (mainly Notch1) was used (Chen, *et al.*, 2001, Stahl, *et al.*, 2008a). There is considerable data from other references where there is less certainty of the exact complement of components present (Shimizu, *et al.*, 2000a, Shimizu, *et al.*, 1999, Shimizu, *et al.*, 2000b).

	Cell-based Signaling					**Cell-based Binding**				
	N1		**N2**			**N1**		**N2**		**References**
	Dll1	**J1**	**Dll1**	**J1**	**J2**	**Dll1**	**J1**	**Dll1**	**J1**	
Lfng			n.c.	↓↓	n.c.			n.c.	↓	(Shimizu, *et al.*, 2001)
Mfng			n.c.	↓	n.c.			n.c.	↓	
Lfng	↑	↓	↑	↑		↑	n.c.			(Hicks, *et al.*, 2000, Yang, *et al.*, 2005)
Mfng	↑(weak)	↓				↑	n.c.			
Rfng	↑	↑				↑	n.c.			
Lfng		↓				↑	↓			(Chen, *et al.*, 2001, Moloney, *et al.*, 2000a, Stahl, *et al.*, 2008a)
Mfng		↓								
						DEL	**SER**			
Fng						↑	↓			(Xu, *et al.*, 2007)

signaling using either NIH3T3 or C2C12 myoblast cells (see Table 1) (Hicks, *et al.*, 2000, Yang, *et al.*, 2005). This mirrors the situation known to occur in *Drosophila* where Fng increases signaling from DEL and inhibits signaling from SER (Fleming, *et al.*, 1997, Panin, *et al.*, 1997). The Weinmaster group has shown that Mfng mirrors the effects of Lfng on N1, although Mfng appears to be less efficient (Hicks, *et al.*, 2000). Importantly, although Lfng causes an increase in binding of Dll1 to N1, it does not appear to significantly alter the binding of Jag1 to N1 (see Table 1) (Hicks, *et al.*, 2000) (Yang, *et al.*, 2005). Thus, unlike in *Drosophila* where Fng decreases SER binding, there is no immediately obvious mechanism for Lfng mediated inhibition of Jag1 signaling. Weinmaster and coworkers also found that the effects of Rfng on signaling through N1 are different from those of Lfng and Mfng. Rfng causes increased signaling with both Dll1 and Jag1 (see Table 1) (Yang, *et al.*, 2005).

The effect of Fng on N2 shows different effects. The Hirai group reported no effect of Lfng or Mfng on the ability of Dll1 or Jag2 to activate N2 using Chinese hamster ovary (CHO) cells (see Table 1) (Shimizu, *et al.*, 2001). With Jag1, they reported that both Lfng and Mfng cause a decrease in N2 signaling, with the level of decrease significantly higher from Mfng (see Table 1) (Shimizu, *et al.*, 2001). In contrast, the Weinmaster group reported (using C2C12 myoblasts) that Lfng causes increased activation of N2 with either the Dll1 or Jag1 ligands (see Table 1) (Hicks, *et al.*, 2000). Clearly, the results for Lfng modulation of N2 signaling will require further investigation. These differences may reflect further layers of complexity caused by the presence or absence of various Notch modulators such as Dll3 that are present in some cells but not others. The data in Table 1 shows that the effect of Fng on activation of Notch by Delta-family ligands appears to be at the level of binding, while that on the activation of Notch by the SER/Jag-family of ligands may be more complicated.

The molecular details for how elongation of *O*-fucose by Fng can change the binding between Notch and DEL (and SER) is not at all clear. Fng appears to modify *O*-fucose on many of the EGF-repeats in the ECD of Notch (Rampal, *et al.*, 2005a, Shao, *et al.*, 2003, Xu, *et al.*, 2007). EGF-repeats 11 and 12 are believed to be essential for ligand binding, and mutation of the *O*-fucose site in EGF 12 decreases N1 function both in cell-based assays (Rampal et al., 2005b) and *in vivo* in mice (Ge and Stanley, 2008). Nonetheless, this is clearly not the entire story since other EGF-repeats in Notch not directly implicated in ligand binding are known to dramatically affect function (Perez, *et al.*, 2005, Rampal, *et al.*, 2005a). For instance the *Abruptex* mutations cluster in EGF-repeats 24-29 outside the ligand binding domain (Perez, *et al.*, 2005). These mutations result in hyperactivated Notch that is refractory to Fng (De Celis and Bray, 2000). Several EGF-repeats in the *Abruptex* region are modified by Fng, and mutation of the *O*-fucose sites in EGF-repeat 26 or 27 in mouse N1 alters Notch signaling in cell based assays (Rampal, *et al.*, 2005a). These data suggest that Fng may mediate its effects on Notch activity not just at EGF-repeat 12, but at numerous sites along the ECD.

One possible explanation for how modification by Fng at numerous sites scattered across the ECD could modulate activity would be effects on the conformation. Structural studies on a short region of N1 (EGF-repeats 11-13) revealed that the presence of calcium binding EGF-repeats results in a fairly rigid structure, while the absence of calcium binding allows flexibility (Hambleton, *et al.*, 2004). Hambleton *et al.* note a helical pattern in the packing of their EGF-repeat crystals suggesting that strings of calcium

binding EGF-repeats may take on conformations resembling helices capable of stacking against one another (Hambleton, *et al.*, 2004). The pattern of calcium binding EGF-repeats in various Notch proteins is highly conserved (Xu, *et al.*, 2005), and as such, may play a crucial role in the function of the protein. This suggests a model whereby the regions containing calcium binding EGF-repeats are rigid, with more flexible "hinge" regions between them, allowing the rigid regions to stack against each other. The elongation of *O*-fucose by Fng could inhibit interactions between neighboring rigid sections of Notch, or conversely, interactions between elongated glycans could facilitate a new interaction between the neighboring rigid regions (Figure 3). The DEL class of ligands would prefer the conformation produced by elongated glycans in the Notch ECD, and the SER/Jag class would prefer the Notch ECD conformation in the absence of Fng elongation. Recently, Pei & Baker showed that the *Abruptex* region of Notch can compete with DEL in binding assays using the ligand binding region of Notch as a target. These data, coupled with knowledge that *Abruptex* mutations activate Notch signaling suggests that the *Abruptex* and ligand binding regions may interact *in vivo* (Pei and Baker, 2008). A conformational change disengaging this interaction would allow DEL ligand binding. It may be that Fng elongated *O*-fucose glycans are modulating the interaction between the *Abruptex* region and ligand binding region of Notch. Due to the size, structural, and conformational heterogeneity possible with extracellular Notch, classical approaches to studying Notch extracellular domain structure may not be feasible. Answers to these questions may await innovative new approaches to the problem.

O-Fucose on Notch EGF-like Repeats

While elongation of *O*-fucose by Fng modulates Notch function, addition of *O*-fucose itself appears to be essential for proper Notch function. Elimination of OFut1 in *Drosophila* results in severe Notch-like phenotypes. An *Ofut1* mutant called *neurotic* results in a classic neurogenic phenotype where there is an overabundance of neurons due to failure of Notch dependent lateral inhibition (Sasamura, *et al.*, 2003). Knockdown of *Ofut1* using RNAi also results in lateral inhibition defects in sensory organ precursor cells, as well as in wing formation, both of which require Notch function (Okajima and Irvine, 2002). Similarly, elimination of Pofut1 in mice produces a severe embryonic lethal phenotype that is more dramatic than single Notch receptor knockouts in mice. This is presumably due to the wholesale destruction of signaling through all four of the Notch receptors present in mammals. The mice die in mid-gestation with defects in neurogenesis, vasculogenesis, and cardiogenesis due to disrupted Notch signaling, while Notch expression does not appear to be affected (Shi and Stanley, 2003).

Although the importance of Pofut1 (or OFut1) for Notch function in both flies and mice is clear, the mechanism by which it affects Notch is not. Irvine and coworkers have shown that OFut1 appears to function as a chaperone for Notch (Okajima, *et al.*, 2005). Their data suggest that loss of OFut1 causes a loss of cell-surface expression of Notch. Interestingly, an enzymatically inactive form of OFut1 ($OFut1^{R245A}$) can rescue the secretion defect, suggesting that the chaperone activity is distinct from the *O*-fucosyltransferase activity. They recently showed that the $OFut1^{R245A}$ mutant can rescue the *Ofut1* null Notch-like neurogenic phenotype (Okajima, *et al.*, 2008). Additionally, flies null for GDP-mannose 4,6 dehydratase (Gmd), which is required for the production

Figure 3

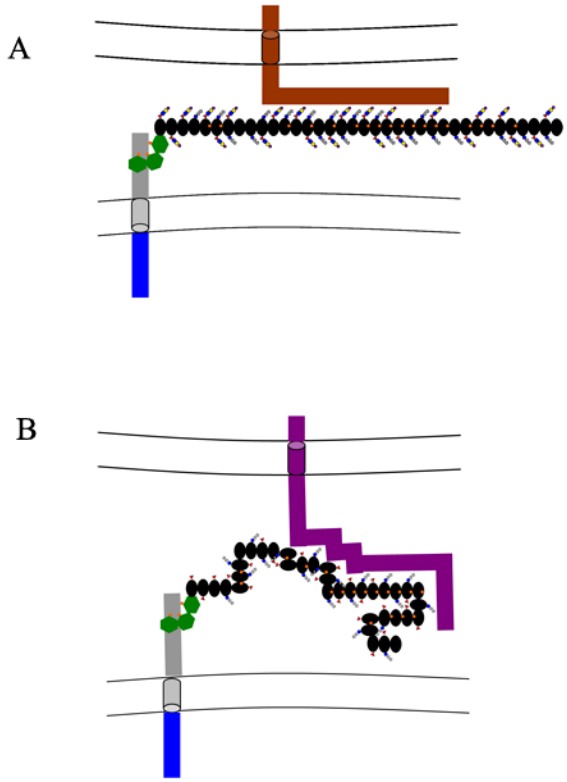


Figure 3. Model of Notch Conformational Change Based on Fringe Modification Coloring as in Figure 2 with a Jagged/Serrate ligand represented as a simplified purple structure and a Delta ligand represented as a simplified brown structure. Intracellular Notch is shown simplified in blue. **A)** Notch modified by Fringe might be more rigid, and have larger steric clashes with elongated glycans preventing Notch from folding back on itself. This conformation would favor Delta ligand binding. **B)** Notch with only *O*-fucose monosaccharide might be less rigid and have less steric hindrance to fold back on itself. This conformation would favor interaction with Jagged ligands. The optimal Notch conformation for each ligand would not interact well with the other ligand.

of GDP-β-L-fucose, do not show a neurogenic phenotype as would have been expected with non-functional Notch (Okajima, *et al.*, 2008). The major phenotype observed in clones of cells expressing OFut1^{R245A} is similar to that of Fng mutants, consistent with the loss of the substrate for Fng. These results suggest that *O*-fucose is essential as a substrate for Fng, but not for other Notch activities. (Okajima, *et al.*, 2008).

The situation in mice appears to be quite different. Stanley and coworkers have shown convincingly that embryonic stem (ES) cells lacking Pofut1 have wild type levels of Notch receptors on their cell surfaces (Stahl, *et al.*, 2008b). Notch activity (either ligand binding or Notch activation) is severely compromised in these cells, suggesting that Notch requires the addition of *O*-fucose for full activity. Overexpression of an enzymatically inactive Pofut1 (equivalent to the R245A mutant in OFut1) in these cells partially restores Notch activity. However, overexpression of another ER protein has the same effect, suggesting the "chaperone" activity of Pofut1 may be a bulk protein effect. Notch activity (both ligand binding and Notch activation) in CHO cells lacking GDP-fucose is also compromised, again suggesting the importance of *O*-fucose in Notch function (Chen, *et al.*, 2001, Moloney, *et al.*, 2000a, Stahl, *et al.*, 2008a).

Experiments utilizing *O*-fucose site mutants in Notch reveal that *O*-fucose is important for optimal Notch function in both flies and mice, but they do not resolve whether the *O*-fucose is important by itself, or is important as a Fng substrate. Expression of Notch lacking Fuc in EGF-repeat 12 of the ligand binding region (N^{12f}) in flies showed a reduction in the ability to respond to Fng (Lei, *et al.*, 2003). Expression of the same mutant (N^{12f}) of mouse N1 in place of endogenous N1 produce mice that were viable and fertile, indicating that this site is not essential for Notch function (Ge and Stanley, 2008). Nonetheless, the mice showed growth defects and T cell abnormalities (T cell development is N1 dependent), suggesting subtle changes in Notch activity. While *O*-fucose on EGF-repeat 12 can be modified by Fng in some contexts (Shao et al., 2003; Xu et al., 2007), it is not know whether the effects observed in the N^{12f}/ N^{12f} mice are due to loss of *O*-fucose or to the lack of ability of Fng to modify the *O*-fucose at that site.

Other studies support additional roles for OFut1 in flies. Sasamura *et al.* proposed that extracellular OFut1 is necessary for proper cycling of cell surface Notch through endosomes and on to lysosomes in a Fuc-dependent manner (Sasamura, *et al.*, 2007). Sasaki *et al.* provided evidence that Notch is localized to the sub-apical complex (SAC)/adherens junction (AJ) in *Drosophila* epithelial cells, and that Notch localization in this context is dynamin dependent and thus likely dependent on transcytosis (Sasaki, *et al.*, 2007). In both papers, the authors suggest a role for OFut1 and fucosylation in these processes. Unfortunately there are technical flaws with the experiments as performed leaving the issue of an OFut1 or *O*-Fucose requirement in these processes unresolved. A major concern of these studies is whether OFut1, which is known to be retained in the ER by virtue of a C-terminal KDEL-like sequence, is actually found at significant levels in extracellular spaces *in vivo*.

All of the experiments performed in the absence of OFut1/POFUT1, as well as those suggesting a non-enzymatic function for OFut1 suffer from the lack of any analysis of the glycosylation state of Notch. The "enzymatically inactive" mutants of OFut1 (e.g. R245A) could retain small amounts of activity allowing partial *O*-fucosylation of Notch. In the absence of GDP-fucose, it is possible that other substrates (e.g. GDP-mannose)

could substitute as a donor substrate for the enzyme. This might explain why Notch without *O*-fucose could maintain at least some of its Fng-independent functions.

There are several potential explanations for the apparent divergence between the role of *O*-fucose in the *Drosophila* and mouse models. First, flies and mice may simply be different. Perhaps to function in flies, Notch does not require *O*-fucose, while in higher organisms it does. It is known that mammals and more primitive organisms use Notch signaling at different stages in development. Most dramatically, mammals do not use Notch for early stages in development such as germ layer formation whereas *C. elegans* and sea urchins do. Shi and Stanley have suggested that adoption of Notch signaling for earlier processes in development may be a later adaptation, rather than the ancestral function (Shi and Stanley, 2006). A second possibility is that the robustness of developmental processes in the two model systems may differ. Perhaps *Drosophila* development provides less strenuous demands upon Notch signaling than mice, and as such, despite Notch signaling at a less than peak efficiency in *Ofut1* null flies, it is sufficient. This may be testable if there are differences in the mutant flies with regard to the number of larvae that hatch, and/or the time it takes them to reach this stage. Additionally, placing the developing larvae under a stress such as higher temperature may reveal a phenotype not apparent under normal developmental conditions. Indeed, if OFut1 is functioning as a classical chaperone, it is exactly this type of stress that the OFut1^{R245A} mutant enzyme should be able to rescue. Thirdly, the differences between mice and flies may not involve differences in the Notch signaling pathway *per se*, but differences in the robustness of the protein expression machinery in the two species. If mice are simply more capable of handling inefficiently folded Notch, perhaps due in part to a slower developmental timescale, then the fly system would be considerably more susceptible to disruption. In this case, even an inefficient rescue effect might be enough to achieve a Notch signaling threshold to permit fly development to continue. Differences in the complement of chaperones in different cells could also help to explain the differences observed.

It is likely that glycosylation significantly affects the physical dynamics of peptides and protein sequences in solution. The presence of *N*-linked glycans on peptides derived from β-turn motifs in hemaglutinin A and a strand from a β-sheet in the Fab fragment of IgA altered the conformational space sampled by these peptides (Imperiali and Rickert, 1995). In the case of the β-turns, glycosylation produced more compact structures (Imperiali and Rickert, 1995). Glycosylation of the β-strand did not have a large effect on how compact the peptide was, but did noticeably narrow the breadth of conformational space sampled by the peptide (Imperiali and Rickert, 1995). Thus, while different peptide/protein structures may react differently to glycosylation, the presence of the glycan appears to act as a sort of anchor in solution, decreasing the conformational space sampled by the molecule. It is not hard to imagine the effect this could have on protein folding. Several groups analyzing the effect of *N*-linked glycans on the stability and folding characteristics of ribonuclease (RNAse) determined that the glycan creates a more stable unfolded molecule that is less dynamic in solution (Arnold, *et al.*, 1999, Choi, *et al.*, 2008). Importantly, they have shown that only the first GlcNAc is likely necessary for this effect on the thermal stability of the molecule (Arnold, *et al.*, 1999, Choi, *et al.*, 2008). Wyss *et al.* reported that removal of all but the first GlcNAc from a high mannose *N*-linked glycan on human CD2 by Endoglycosidase H (EndoH) had no

effect on function or on the circular dichroism (CD) spectra of the protein (Wyss, *et al.*, 1995). However, removal of the entire glycan by peptide *N*-glycanase F (PNGase F) eliminates function, antibody recognition, causes protein aggregation, and creates significant alterations in the CD spectra (Wyss, *et al.*, 1995). A nuclear magnetic resonance (NMR) study of an *O*-fucosylated form of the *Pars intercerebralis* major peptide C (PMP-C) from *Locusta migratoria* showed a strikingly similar result in that there was no structural effect on the backbone fold of the protein due to glycosylation, but the thermal stability of the protein was substantially increased (Mer, *et al.*, 1996). This could even be observed as a decrease in the rate of deuterium exchange for amide protons on residues remote from the site of fucosylation (Mer, *et al.*, 1996). The presence of an *O*-fucose glycan on a Factor VII EGF-repeat was shown to have no effect on the backbone structure of the motif, while still showing NMR chemical shift effects on regions distant from the site of glycosylation (Kao, *et al.*, 1999). Interestingly, the presence of the *O*-fucose glycan increased the affinity of the N-terminal calcium binding domain for calcium (Kao, *et al.*, 1999). While the effect was a modest two-fold increase in affinity, considering that calcium is known to produce rigidity between EGF-repeats (Hambleton, *et al.*, 2004, Rao, *et al.*, 1995), and that this was an NMR study on a single EGF-repeat, the effect could be greater in proteins with tandem EGF-repeats. Analysis of β-*O*-glucose on serine and threonine peptides by NMR and molecular dynamics (MD) calculations agree with these studies in that the glycosylated peptides preferentially adopt conformational space resembling more folded structures as opposed to conformational space resembling the extended conformations of non-glycosylated forms (Corzana, *et al.*, 2006). The evidence seems to support the statement that a monosaccharide glycan on an unfolded peptide or protein will either bias the backbone toward more compact structures or decrease the breadth of conformational space sampled in the ensemble of states. It seems likely that the glycosylation effect on a small motif like an EGF-repeat facilitates a predisposition toward proper folding, more stable structures, and possibly, in the particular example of cysteine-rich motifs, proper cysteine pairing for disulfide bond formation (see Figure 4). In the case of the Notch ECD as a whole, *O*-fucosylation might alter the affinity of calcium binding EGF-repeats for calcium, making some regions more rigid (Figure 3). It is interesting that Pei & Baker found that DEL binding appeared to be stronger at lower pH and that this effect was reversed on a return to neutral pH (Pei and Baker, 2008). Perhaps this is due to alteration of Notch structure through effects on calcium chelation in the N-terminus of calcium binding EGF-repeats.

Galactose Elongated Glycans on Notch EGF-like Repeats

Recent results suggest there is no elongation past the GlcNAc-β1,3-Fuc disaccharide on *Drosophila* Notch in S2 cells (Xu, *et al.*, 2007). The effect of Fng on Notch-ligand binding can be recapitulated with purified components *in vitro*, suggesting that further elongation is not required for Fng to modulate Notch activity (Xu, *et al.*, 2007). In contrast, in mammals further elongation of the glycan is significant. Stanley and coworkers showed that Lfng didn't inhibit Jag1-mediated Notch activation in CHO cells incapable of adding Gal to the GlcNAc-β1,3-Fuc disaccharide (Chen, *et al.*, 2001). Detailed analysis of mice lacking the β1,4-galactosyltransferase 1 (β4GalT1) enzyme responsible for addition of this Gal revealed a mild segmentation defect, suggesting some involvement of the Gal in the ability of Lfng to modulate Notch function *in vivo*

Figure 4

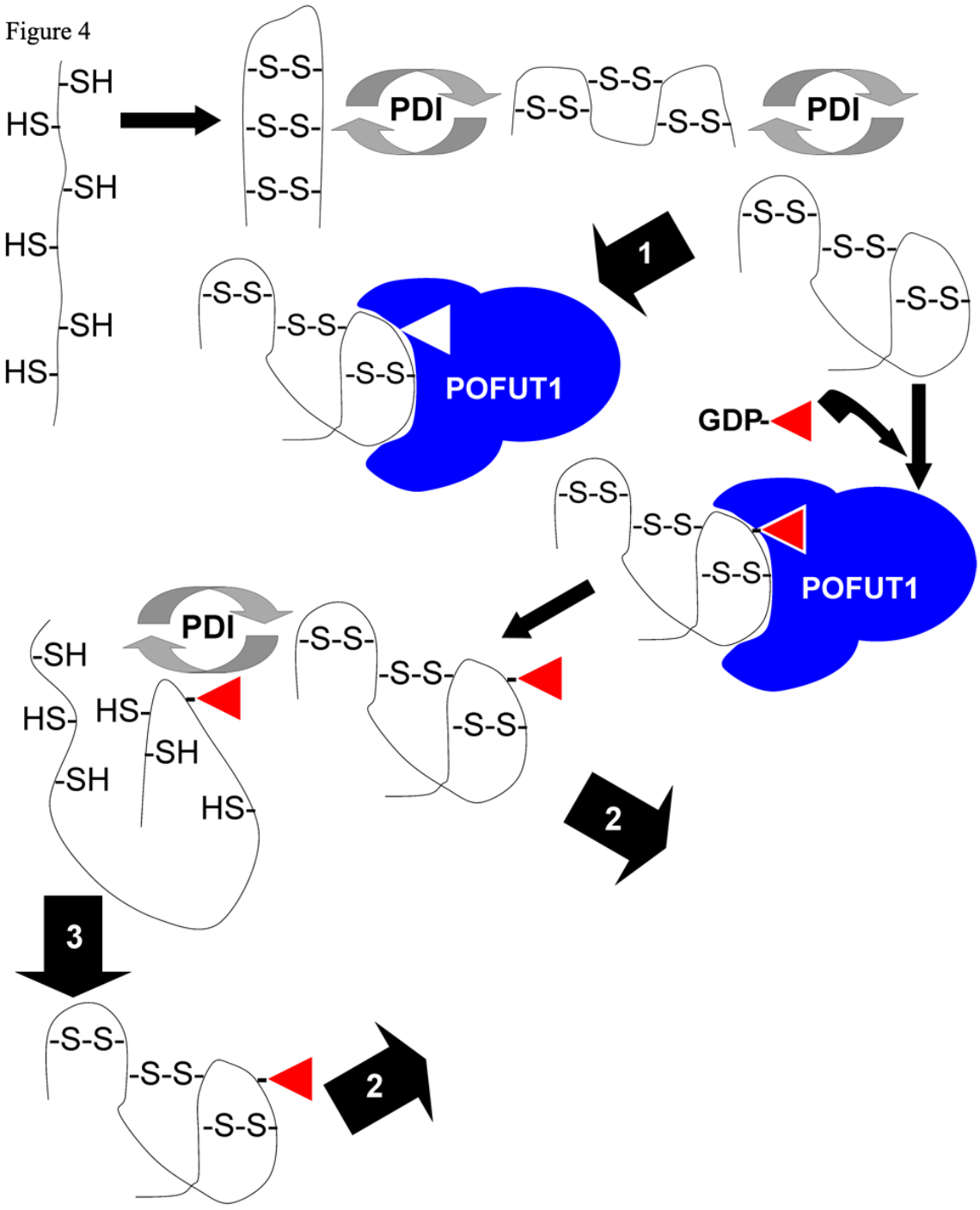

-SH
HS-
-SH
HS-
-SH
HS-
-S-S-
-S-S-
-S-S-
PDI
-S-S-
-S-S-
-S-S-
PDI
1
-S-S-
-S-S-
-S-S-
POFUT1
-S-S-
-S-S-
-S-S-
GDP-
-S-S-
-S-S-
-S-S-
POFUT1
PDI
-SH
HS-
HS-
-SH
-SH
HS-
-S-S-
-S-S-
-S-S-
2
3
-S-S-
-S-S-
-S-S-
2

Figure 4. Model For OFut1/Pofut1 Effects on Notch EGF-repeat Folding in the Endoplasmic Reticulum

The EGF-repeat is shown as a line with the cysteine side chains indicated as –SH in reduced form and –S-S- for the disulfide. OFut1/Pofut1 is shown as a blue globular structure. *O*-linked fucose is shown as a red triangle, and GDP-fucose is represented as GDP-connected to a red triangle. **Arrow #1)** In *Drosophila*, OFut1 in the absence of GDP-fucose chaperones the folded EGF-repeat out of the endoplasmic reticulum. **Arrow #2)** When a properly folded EGF-repeat is recognized by OFut1/Pofut1 in the presence of GDP-fucose, the enzyme transfers fucose to the EGF-repeat. At this point the EGF-repeat is folded and ready for export. **Arrow#3)** In the model where a monosaccharide alters the solution conformation of an unfolded protein or peptide, once a folded EGF-repeat is *O*-fucosylated, if it partially or completely unfolds subsequent to glycosylation, it remains in a state capable of refolding much more efficiently than the original unfolded and extended structure present in the nascent peptide.

(Chen, *et al.*, 2006). In contrast, Stanley and coworkers have shown that the presence or absence of Sia does not affect Fng modulation of Notch signaling from Jag1 (Chen, *et al.*, 2001). The additional affects of the Gal on Notch function may represent an additional level of fine tuning acquired through evolution.

O-Glucose on Notch EGF-like Repeats

While the importance of *O*-fucose modifications in Notch biology has been clear for several years, the importance of *O*-glucose glycans has only recently been realized. Acar *et al.* have shown that the *rumi* gene encodes the Poglut (Acar, *et al.*, 2008). Mutations in *rumi* exhibit a temperature sensitive Notch phenotype that is suggestive of a folding effect on the receptor (Acar, *et al.*, 2008). The authors reported that loss of Rumi results in cell-autonomous defects affecting the extracellular domain of Notch. The authors observed accumulation of Notch at the cell surface, but no apparent ER related folding defects for Notch in the *rumi* mutants (Acar, *et al.*, 2008). Thus, Rumi does not appear to be a chaperone analogous to OFut1. Acar *et al.* proceed to show that RNAi-mediated knockdown of *rumi* in Schneider 2-cells (S2-cells) does not affect DEL binding to Notch (Acar, *et al.*, 2008). Thus, loss of Rumi appears to affect a step between ligand binding and the S2 cleavage. These data strongly suggests that the rumi effect on Notch is a conformational effect on the extracellular domain of the receptor that prevents cleavage to activated Notch from occurring (see Figure 4). This supports the conclusion that *O*-glucose glycans are necessary for the function of Notch in flies, and that the *rumi* temperature-sensitive phenotype is due to impaired *O*-glucosylation of the Notch receptor.

Little is known about the biological function of elongation of *O*-glucose by two Xyl (Figure 1). Enzymatic activities for the two XylT responsible for their addition have been identified (Minamida, *et al.*, 1996, Omichi, *et al.*, 1997), and the β-glucose α1,3-xylsosyltransferase has been partially purified (Omichi, *et al.*, 1997). The biochemical data suggests that each Xyl is added by a separate enzyme. Determination of whether it is *O*-glucose alone, or the elongated glycan that is required for Notch function will require the identification of the genes encoding these enzymes.

Thrombospondin Repeats in Signaling

O-Fucose Glycans on Thrombospondin Repeats

Little is know about the function of *O*-fucose on TSRs, although database searches using the consensus sequence for modification suggest more than 40 proteins may be modified in mammals. TSRs are known to interact with fibronectin, heparin, glycosaminoglycans (including heparan sulfate), and CD36 (reviewed by (Adams and Lawler, 2004, Adams and Tucker, 2000)). TSR containing proteins have been shown to promote neurite outgrowth (Bamdad, *et al.*, 2004) and possess anti-angiogenic qualities (Iruela-Arispe, *et al.*, 1999, Lee, *et al.*, 2006, Tolsma, *et al.*, 1993). A Disintegrin And Metalloprotease with ThromboSpondin 1 (ADAMTS1) has been reported to produce anti-angiogenic peptides from Thrombospondin 1 and 2, suggesting that TSRs may function not solely as folded protein motifs, but in an activated glyco-peptide form as well (Lee, *et al.*, 2006). The interaction between the TSR motifs of thrombospondin 1 or 2 and CD36 are reported to encompass the peptide that is *O*-fucosylated (Iruela-Arispe, *et*

al., 1999, Silverstein and Febbraio, 2007). These peptides have been shown to exhibit anti-angiogenic activity alone, presumably through interactions with CD36 (Iruela-Arispe, *et al.*, 1999). Whether the presence of *O*-fucose on the peptide would have an effect on this anti-angiogenic effect is as yet unknown. The presence of the glycan would presumably have a greater effect on the structure and function of a small peptide in comparison to a large protein motif making this an intriguing avenue of inquiry. Studies investigating the efficacy of TSR peptides to combat cancer are under way. Meiniel and coworkers have shown that a peptide encompassing the *O*-fucose site of a subcommissural organ (SCO)-spondin TSR can promote neurite outgrowth in culture (Meiniel, *et al.*, 2003). The neurite outgrowth is stimulated through $\alpha 1/\beta 1$ integrin, although it remains to be determined whether the TSR peptide is binding directly to the integrin or is associated with some other integrin activating factor (Meiniel, *et al.*, 2003). As with the other peptide studies mentioned, the effect of glycosylation on their function has yet to be determined.

Recent studies suggest that addition of *O*-fucose to TSRs may play a role in quality control. Both ADAMTS13 and ADAMTS-like 1 are modified on multiple TSRs with *O*-fucose (Ricketts, *et al.*, 2007, Wang, *et al.*, 2007). Majerus and coworkers have shown that *O*-fucose is necessary for the secretion of ADAMTS13 (Ricketts, *et al.*, 2007). They showed that at least six of the seven TSRs containing a consensus sequence for *O*-fucosylation contained the modification, and that elimination of glycosylation sites by mutagenesis had an additive effect on the secretion defect (Ricketts, *et al.*, 2007). The enzymatic activity of Pofut2 was necessary to rescue the defect as an enzymatically dead Pofut2 was unable to restore secretion of ADAMTS13 (Ricketts, *et al.*, 2007). Apte and co-workers reported a similar secretion defect with ADAMTS-like 1/Punctin 1 (Wang, *et al.*, 2007). They report the presence of both monosaccharide and disaccharide on the TSRs of Punctin-1, and that elimination of *O*-fucosylation sites in three of the four TSRs of Punctin-1 have a dramatic cumulative effect on the secretion of Punctin-1 from a number of cell types in culture (Wang, *et al.*, 2007). Lec13 CHO cells lack the ability to fucosylate due to a defect in the Gmd gene (Ripka, *et al.*, 1986). Expression of Punctin-1 lacking *O*-Fucose sites in all three TSRs in the Lec13 cells showed identical secretion defects to other cell types. Supplementation of the Lec13 medium with fucose allowing a salvage pathway to produce the GDP-fucose donor largely rescued the phenotype, definitively showing that the defect was due to the missing glycans (Wang, *et al.*, 2007). Interestingly, Pofut2 only modifies properly folded TSRs *in vitro* (Luo, *et al.*, 2006b). The fact that Pofut2 is ER localized, together with its ability to differentiate between folded and unfolded structures, has raised the possibility that Pofut2 plays an important role in quality control or folding of TSR containing proteins.

As an initial step towards understanding the function of TSR *O*-fucosylation, we have generated a mouse lacking Pofut2. Preliminary analysis suggests that homozygous mutants die during early embryogenesis (Haltiwanger and Holdener, manuscript in preparation). Although the specific defect in these mice is not known, the data suggests that *O*-fucosylation of one or more of the target proteins is essential for early embryogenesis. Studies to further define the defect are underway.

β1,3-Glucosyltransferase Elongated *O*-Fucose on Thrombospondin Repeats

Mutations in the β3GlcT that modifies *O*-fucose on TSRs results in a human genetic deficiency known as Peters Plus syndrome (Lesnik Oberstein, *et al.*, 2006). Patients with Peters Plus syndrome display developmental delay, cleft lip or palate, defects in the anterior eye chamber, and short stature (for a review see (Maillette de Buy Wenniger-Prick and Hennekam, 2002)). Recently, the Hofsteenge group has shown by mass spectrometry that the TSR containing protein properdin from the serum of Peters Plus patients does not contain glucosylated *O*-fucose glycans (Hess, *et al.*, 2008). The disaccharide on TSRs is produced in the ER owing to the localization of the β3GlcT in the ER along with Pofut2 (Kozma, *et al.*, 2006, Sato, *et al.*, 2006). The fact that Properdin was found in the serum of Peters Plus individuals suggests that addition of the glucose (Glc) is not essential for secretion of the protein. Further studies will be necessary to determine how loss of Glc from TSRs results in the defects observed in these individuals.

C-*Mannosylation*

The *C*-mannosyltransferase (CmanT) activity is present in a number of animal cells and ultimately in insects as well (Doucey, *et al.*, 1999, Hofsteenge, *et al.*, 2001, Munte, *et al.*, 2008). The gene(s) for the enzyme or enzymes have not yet been identified and it is currently unclear whether there is more than one transferase, and whether the animal and insect enzyme activities possess the same specificity. The Man is transferred in alpha linkage typically to the first tryptophan side-chain in motifs resembling the W-X-X-W sequence from the amino terminus of TSRs (Doucey, *et al.*, 1998). The CmanT apparently only requires the tryptophan containing portion of the TSR sequence, and as such, many proteins that do not contain TSRs are also potentially modified. This includes the erythropoietin receptor (EPOR) (Furmanek, *et al.*, 2003). The absence of *C*-mannosylation in EPOR has been implicated in functional defects for the receptor (Hilton, *et al.*, 1996). It appears that this is not due to direct effects on receptor ligand interaction, but mirroring the *O*-fucose modification of TSRs apparently affects folding and trafficking of the EPOR (Hilton, *et al.*, 1996). Ligands can be modified as well as receptors with a hypertrehalosaemic hormone from the stick insect *Carausius morosus* shown to contain *C*-mannosylated tryptophans, as was recombinant human interleukin 12 (IL12) (Doucey, *et al.*, 1999, Munte, *et al.*, 2008). While analysis of endogenously produced IL12 was not practical, microsomes prepared from the cells producing the cytokine were shown to be efficient at *C*-mannosylating peptide substrates (Doucey, *et al.*, 1999).

C-mannosylation of TSRs has the potential to affect a number of cell signaling events, although few details have yet been reported. Ihara and coworkers have found that *C*-mannosylated peptides from TSRs enhance the production of tumor necrosis factor alpha (TNF-α) by lipopolysaccharide (LPS) in mouse macrophage-like RAW264.7 cells (Muroi, *et al.*, 2007). Production of TNF-α was not caused by the unglycosylated peptides, and they determined that TNF-α production was enhanced partly through C-jun N-terminal kinase (JNK) activation by transforming growth factor beta (TGF-β)-

activated kinase 1 (TAK1) (Muroi, *et al.*, 2007). Nuclear factor kappa B (NFκB) was also reported to play a smaller role in the LPS induced TNF-α increase. Peptides from TSRs have also been shown to possess anti-angiogenic effects, including a peptide comprised of the tryptophan containing *C*-mannosylation sequence (Iruela-Arispe, *et al.*, 1999). What effect the presence or absence of *C*-mannosylation might have on the activity of these peptides is as yet unknown. The SCO-spondin TSR peptide used by Meiniel and coworkers to promote neurite outgrowth in culture was comprised of both the *C*-mannosylation and *O*-fucose consensus sequences (Meiniel, *et al.*, 2003). As such, the relative contribution of either consensus sequence, with or without glycosylation awaits further investigation.

Glycosyltransferase Structure and Mechanism

Glycosyltransferase Families

It has been estimated that most organisms utilize approximately 1% of their genome to code for GTs (Coutinho, *et al.*, 2003, Davies, *et al.*, 2005). The number and diversity of recognized GT families has climbed steadily in the last decade. In 2001 there were less than 50 recognized GT families, and by the end of 2003 approximately 70 families were annotated in the Carbohydrate Active enZymes CAZy database. Today that number is almost 90. The majority of these enzymes have been classified based on sequence similarity to one of two folds, GT-A or GT-B. While the majority of these classifications are undoubtedly correct due to the ability to compare the sequences of related genes, many GTs sharing the same fold, in the same family do not share significant sequence similarity. Attempting to overcome this problem, Kikuchi *et al.* used the profile hidden Markov model (HMM) method to classify known GT families into super-families based on their predicted fold in 2003 (60 CAZy families) (Kikuchi, *et al.*, 2003). They concluded that there were at least four super-families they termed GTS-A through D. GTS-A and B corresponded broadly to the previously determined GT-A and GT-B folds, while GTS-C is comprised of several CAZy families that use dolichol-linked donors and one family of bacterial arabinosyltransferases. GTS-D comprises CAZy families of α1,2-/α1,6-fucosyltransferases (Kikuchi, *et al.*, 2003). There were a number of CAZy families (13 of 60) that did not fit into the GTSA-D classification. This includes four families of sialyltransferases forming four separate clusters suggesting they may be unique. Also, the α1,3-/α1,4-fucosyltransferases previously thought to be GT-B enzymes were suggested to be a different fold than the GTS-D family. Interestingly, Kikuchi *et al.* noted that the GTS-A family appeared to be related to nucleotidyl-transferases while the GTS-B family was related to the uridine-5'-diphosphate (UDP)-N-acetylglucosamine-2-epimerase family. They hypothesize that ancestral glycosyltransferases were of the GTS-B family supplied with nucleotide-sugar donor substrates by proteins of the GTS-A fold. They suggest many glycosyltransferase enzymes have since evolved from these nucleotidyl transferases (Kikuchi, *et al.*, 2003). The Pofut enzymes have been suggested to be distantly related to the GTS-D super-family, although the Pofut CAZy families were not included in the Kikuchi *et al.* analysis. The Fng family of enzymes have always been classified as having either a GT-A or GTS-A fold, and their membership in this group has been confirmed structurally (Jinek, *et al.*, 2006).

Structure and Mechanism

Glycosyltransferases can transfer a glycan to the acceptor substrate with either inversion or retention of the anomeric configuration (Unligil and Rini, 2000). Both the GTS-A and GTS-B superfamilies are known to contain members of each mechanistic class. Fng is an inverting enzyme, the mechanism of which is presumed to proceed through the abstraction of a proton from the 3'-OH of the fucose-*O*-R by a catalytic base (asp or glu) in the active site. This creates a nucleophile that attacks the anomeric carbon of the GlcNAc in UDP-GlcNAc (Sinnott, 1990, Unligil and Rini, 2000). The enzyme then releases the GlcNAc-β1,3-Fuc-*O*-R and UDP products. The mechanism is often dependent on the presence of manganese (as it is for Fng), although this is due to the requirement for a manganese ion to help coordinate the phosphates of the nucleotide-sugar donor in the active site, rather than to a catalytic role (Jinek, *et al.*, 2006, Qasba, *et al.*, 2005, Ramakrishnan, *et al.*, 2004).

There are no structures of GTS-C or GTS-D superfamily members at the moment. The GTS-A fold is comprised primarily of a Rossman-like fold containing the nucleotide binding DXD motif and the catalytic center of the enzyme, and a second smaller C-terminal α/β domain presumed to interact with acceptor substrates (Breton, *et al.*, 2006, Qasba, *et al.*, 2005). The GTS-B fold is comprised of two Rossman-like folds linked together with the donor substrate binding to one domain in a cleft between the two Rossman-like folds (Breton, *et al.*, 2006, Qasba, *et al.*, 2005, Unligil and Rini, 2000). A recent structure of mouse Mfng confirmed the prediction from previous GT structures and sequence alignments that the DXD motif found in most GTS-A super-family members, and present in the Fng enzymes is coordinating a manganese ion, which in turn coordinates the phosphates of the UDP-GlcNAc donor substrate (Figure 6) (Jinek, *et al.*, 2006). A not uncommon occurrence in GT crystal structures in the absence of an acceptor substrate is that density for the sugar portion of the bound nucleotide-sugar donor is absent, indicating conformational flexibility (Gastinel, *et al.*, 1999, Morera, *et al.*, 1999, Vrielink, *et al.*, 1994). This was the case with Mfng (Jinek, *et al.*, 2006). The Mfng structure is comprised of the catalytic core of the enzyme including the C-terminal domain presumed to be involved in acceptor substrate binding. Most glycosyltransferases exhibit flexible loops (Breton, *et al.*, 2006, Unligil and Rini, 2000). In the case of Mfng, these flexible loops were not visible in the crystal structure and presumably become ordered upon substrate binding (Jinek, *et al.*, 2006). A number of proposals have been made for the function of these loops, from a role in forming a binding platform for the acceptor substrate, to functioning as a switch to expel substrates after glycan transfer and prevent substrate inhibition (Qasba, *et al.*, 2005, Unligil and Rini, 2000). Unfortunately there is no structural information for a Fng enzyme with a bound acceptor substrate.

Aim of This Work

Chapter One

The Notch receptor is a critical component during development and in numerous cell-fate decisions. Understanding the Notch signaling pathway and components may one day have profound influences on how we intervene in numerous disease states. The Fng β3GlcNAcT enzymes, as regulators of this pathway, are potential drug targets.

Understanding the structural aspects of the interaction between Notch and the Notch pathway components is a critical step toward a future pharmacologic intervention in cases where Notch is a contributor to disease. The recent crystal structure of the Mfng enzyme (Jinek, *et al.*, 2006) was a first step toward the development of drugs to inhibit the Fng enzymes. Unfortunately, there is currently no structure with the acceptor substrate bound to the enzyme. This is the crucial information necessary to begin a structure based approach to inhibitor design.

Thus, in chapter one, we designed a study to provide insight into enzyme-substrate specificity, as well as mechanistic insights into Lfng catalysis. Results of docking a model of a fucosylated EGF-repeat (EGF-*O*-fucose) acceptor substrate onto a homology model of Lfng, and a multiple sequence alignment of Fng proteins, was used to guide a mutagenesis strategy for the Lfng enzyme. We targeted four areas, namely two conserved loops not visible in the Mfng structure, two clusters of docked EGF-*O*-fucose substrates, residues involved in UDP-GlcNAc donor specificity, and the F187L mutation associated with the human disease SCD. All represent areas for which there is little or no structural information. The majority of these results have been submitted to the *Journal of Biological Chemistry* for publication.

Chapter Two

In chapter two, the characterization of three separate glycan structures is described. In the first case, we showed conclusively that the Rfng enzyme is a β3GlcNAcT like its two sister Fng enzymes. These data were published in the *Journal of Biological Chemistry* (Rampal, *et al.*, 2005b). In the second case, in collaboration with Hofsteenge and coworkers, we showed that the glycan product produced on TSRs by the β3GlcT was the disaccharide Glc-β1,3-Fuc-*O*. These data were published in the *Journal of Biological Chemistry* (Kozma, *et al.*, 2006). In the third case, in collaboration with Majerus and coworkers, we confirmed that the glycans released from ADAMTS13 were the disaccharide Glc-β1,3-Fuc-*O*. These data were published in the *Journal of Biological Chemistry* (Ricketts, *et al.*, 2007). Here we report the additional observation of an as yet uncharacterized *O*-fucose glycan structure released from the TSRs of ADAMTS13.

Chapter Three

In chapter three we report the results of database searches for potential *O*-fucose and *O*-glucose containing proteins based on previously published consensus sequences for glycan addition. The results of the *O*-fucose consensus search was published in *Current Molecular Medicine* (Rampal, *et al.*, 2007).

Chapter One

A Kinetic Analysis of Lunatic Fringe Mutants

Summary

We report the catalytic effects of a mutagenesis screen spanning the active site region of Lfng using the small molecule 4-nitrophenyl-α-L-fucopyranoside (pNP-fucose) acceptor substrate. We utilized a kinetic analysis of mutant enzyme activity toward the small molecule pNP-fucose to judge the effect on Lfng activity. We have found evidence that one loop shields the active site coincident with, or subsequent to, substrate binding. We propose a mechanism whereby the ordering of this short loop may alter the conformation of the catalytic aspartate. We report that Lfng may have a slow-on component to its mechanism and hypothesize about the role such a mechanism may play in the vertebrate segmentation clock. We identify several residues near the UDP-GlcNAc binding site, which are specifically permissive toward UDP-GlcNAc utilization. Finally, we report that the F187L Lfng mutant identified as a genetic cause for the human disease SCD may be catalytically active, contrary to previously published conclusions (Sparrow, *et al.*, 2006). In light of these data we propose that the defect with Lfng in SCD is not due to a catalytic defect but to enzyme stability, folding, trafficking and/or localization.

Methods

Homology Models of Lunatic Fringe and β1,3Glucosyltransferase

Lfng was aligned with Mfng using ClustalW (Thompson, *et al.*, 1994). The alignment and the Mfng structure (pdb 2J0B) were then submitted to the SwissModel server (Schwede, *et al.*, 2003). When we modeled the β3GlcT enzyme onto the Mfng structure using the SwissModel server the result did not produce significant differences in the active site. However, we found that the result from submitting the mouse β3GlcT sequence to the ESyPred3D server (Lambert, *et al.*, 2002) along with the Mfng structure (pdb 2J0B) produced significant differences near the donor-substrate binding site which we could survey by mutation.

Docking of EGF-O-fucose to the Lunatic Fringe Homology Model

The EGF-*O*-fucose model was prepared as previously described (Rampal, *et al.*, 2005b). We used the program HEX4.5 (Ritchie, 2003) to dock the EGF-*O*-fucose onto the Lfng model. The crystal structure of the Factor IX EGF-repeat showed dual side-chain conformations for a number of residues. We first docked the structure containing both conformations of all side chains, and determined that three amino acid side-chains (a phenylalanine, tryptophan, and lysine) were likely to be oriented toward the Lfng protein when docked. We prepared eight separate pdb files for all combinations of the three amino acid side-chain conformations, and docked each one separately. Any result that

positioned the fucose near the catalytic asp was saved. We saved the first 10 to 30 solutions (lowest energy) for each of the eight docked EGF-*O*-fucose ligands resulting in 161 solutions, which were then loaded into the molecular visualization software PyMOL (Delano, 2002). It should be noted that the HEX program clusters similar solutions based on a user specified root mean squared deviation (RMSD) cutoff. In our first round of manual culling, we accepted only the first (lowest energy) solution for any given HEX cluster, and ignored subsequent solutions that belonged to the same cluster. Thus when we describe clustering of solutions in the Results section we are not referring to the original clusters of solutions produced by the HEX program. We further manually culled inappropriate solutions where the fucose and aspartate 289 were too far apart using PyMOL, resulting in a final total of 80 solutions.

Preparation of Lunatic Fringe Enzyme Mutants

Primer pairs in Table 2 were polymerase chain reaction amplified by *Pfu* Turbo polymerase (Stratagene) with the program summarized in Table 2 with pSecTag2CLfng (Rampal, *et al.*, 2005b) as template. Eight µg of endo-free maxi-prep DNA of each construct was transfected into a 10 cm dish of approximately 50% confluent HEK293T cells using 40 µL of GenePorter reagent (Gene Therapy Systems). The transfected cells were grown for three days in DMEM (Gibco) supplemented with 9% Bovine calf serum (Hyclone) and 0.9% penicillin and streptomycin (Gibco). One tablet of complete-Mini EDTA-free protease inhibitor cocktail (Roche) was dissolved in 10 mL of MilliQ system purified (MQ) H_2O. Ten mL of media was harvested into 1 mL of the protease inhibitor on ice. The media was centrifuged at approximately 4500 × *g* for 10 minutes to remove cell debris. The supernatant was then mixed with 275 µL of 1 M Tris-HCl (Fisher) pH 8.0 on ice. Ni^{2+}-nitrilotriacetic acid (Ni-NTA) beads (Qiagen) were equilibrated in 150 mM NaCl (Fisher), 10 mM Tris-HCl pH 7.8 (wash 2). A 120 µL aliquot of the wash 2 equilibrated 50% bead slurry was added to the harvested media on ice. The media and beads were placed on a rotator at 4° C for 1 hour, and then poured into a small disposable column in a cold room. The beads were washed with 10 bed volumes of 500 mM NaCl, 10 mM Tris-HCl pH 7.8, ten bed volumes of wash 2, three times with one bed volume of 50 mM imidazole (Sigma), 150 mM NaCl 10 mM Tris-HCl pH 7.8, and then eluted five times with one bed volume of 250 mM imidazole, 150 mM NaCl, 10 mM Tris-HCl pH 7.8. The elutions were consolidated and glycerol (J. T. Baker) was added to 20% by volume. A sample was removed and mixed with 5 × SDS gel loading buffer for western immuno-blot quantification, and the remainder was aliquoted and immediately frozen at -80° C.

Enzyme Quantification

The enzyme gel samples to be quantified were electrophoresed by 10% SDS-PAGE adjacent to lanes containing 400, 200, 100, 50, 25, and 12.5 ng samples of Lfng standard previously quantified by densitometry of Coomassie stained gels (Rampal, *et al.*, 2005b). The gel was then blotted onto nitrocellulose (BioRad), and blocked at 4° C overnight in 5% non-fat dried milk powder in 150 mM NaCl, 10 mM phosphate buffered saline pH 7.4 (PBS). The blot was incubated at room temperature with gentle rocking with a 1:500 dilution of his-probe rabbit polyclonal IgG (H15) antibody (Santa Cruz) in PBS plus 0.1% Tween 20 (Sigma) (PBS-Tween) for 45 minutes. The blot was rinsed in

Table 2. PCR Primer Pairs Used For Lunatic Fringe Mutagenesis
The underlined sequence corresponds to the portion that is complementary to the paired primer. The sequences are in capital letters with the exception of the nucleotides that are being mutated. The bottom of the table contains the mutagenesis program for the PCR cycler.

Primer Pairs	
Mutant	Primer Sequence
S168A	CACCAACTGCgCCTCGGCCCACAGCCGC
	GGCCGAGGcGCAGTTGGTGAGCACCACATTGC
S168V	CCAACTGCgtCTCGGCCCACAGCCGCCAGG
	GGGCCGAGacGCAGTTGGTGAGCACCACATTGCC
H171A	CTCGGCCgcCAGCCGCCAGGCTCTGTCCTG
	GGCGGCTGgcGGCCGAGGAGCAGTTGGTGAG
H171D	CTCGGCCgACAGCCGCCAGGCTCTGTCC
	GGCGGCTGTcGGCCGAGGAGCAGTTGGTG
R173A	CCACAGCgcCCAGGCTCTGTCCTGCAAGATGG
	GAGCCTGGgcGCTGTGGGCCGAGGAGC
A175V	CAGCCGCCAGGtTCTGTCCTGCAAGATGGCTGTG
	CAGGACAGAaCCTGGCGGCTGTGGGCCGAG
L176A	CGCCAGGCTgccTCCTGCAAGATGGCTGTGGAGTATG
	TTGCAGGAggcAGCCTGGCGGCTGTGGGCCGAG
S177C	CAGGCTCTGTgCTGCAAGATGGCTGTG
	TCTTGCAGcaCAGAGCCTGGCGGCTG
S177D	CAGGCTCTGgaCTGCAAGATGGCTGTGGAG
	TCTTGCAGtcCAGAGCCTGGCGGCTGTG
F187L	GTATGACCGAcTgATTGAGTCTGGGAAGAAGTGG
	GACTCAATcAgTCGGTCATACTCCACAGCC
S228A	GCAAGCCCgcCCTGGACAGGCCCATCCAGG
	CTGTCCAGGgcGGGCTTGCCGATGTACACGTC
S228L	GCAAGCCCctCCTGGACAGGCCCATCCAGG
	CTGTCCAGGagGGGCTTGCCGATGTACACGTC
S228T	GCAAGCCCAcCCTGGACAGGCCCATCCAGG
	CTGTCCAGGgTGGGCTTGCCGATGTACACGTC
S228Y	GCAAGCCCtaCCTGGACAGGCCCATCCAGG
	CTGTCCAGGtaGGGCTTGCCGATGTACACGTC
L229Q	AAGCCCAGCCaGGACAGGCCCATCCAGGCCAC
	GGCCTGTCCtGGCTGGGCTTGCCGATGTAC
R231A	CCTGGACgccCCCATCCAGGCCACAGAACGG
	GGATGGGggcGTCCAGGCTGGGCTTGCCG
I233A	CAGGCCCgcCCAGGCCACAGAACGGATCAGC
	GTGGCCTGGgcGGGCCTGTCCAGGCTG
A235Y	CCCATCCAGtaCACAGAACGGATCAGCGAG
	CGTTCTGTGtaCTGGATGGGCCTGTCCAGG

E237A	CAGGCCACAGccCGGATCAGCGAGCACAAAGTG
	CTGATCCGggCTGTGGCCTGGATGGGCCTG
H242N	GATCAGCGAGaACAAAGTGAGACCTGTC
	CACTTTGTtCTCGCTGATCCGTTCTGTG
F251S	CACTTTTGGTcTGCCACCGGAGGAGCTGG
	CCGGTGGCaGACCAAAAGTGGACAGGTCTC
F251Y	CACTTTTGGTaTGCCACCGGAGGAGCTGG
	CCGGTGGCAtACCAAAAGTGGACAGGTCT
T253A	TGGTTTGCCgCCGGAGGAGCTGGCTTCTG
	CTCCTCCGGcGGCAAACCAAAAGTGGACAG
G254A	GTTTGCCACCGcAGGAGCTGGCTTCTGCATC
	CCAGCTCCTgCGGTGGCAAACCAAAAGTGG
D288A	GGCTCCCCGcTGACTGCACCATTGGCTAC
	GTGCAGTCAgCGGGGAGCCGGATGCG
D288E	GGCTCCCCGAgGACTGCACCATTGGCTAC
	GTGCAGTCcTCGGGGAGCCGGATGCG
D288S	GCTCCCCagcGACTGCACCATTGGCTACATTG
	GCAGTCgctGGGGAGCCGGATGCGCTC
D289A	TCCCCGATGcCTGCACCATTGGCTACATTG
	ATGGTGCAGgCATCGGGGAGCCGGATG
D289E	TCCCCGATGAgTGCACCATTGGCTACATTG
	ATGGTGCAcTCATCGGGGAGCCGGATG
D289S	TCCCCGATtcCTGCACCATTGGCTACATTG
	ATGGTGCAGgaATCGGGGAGCCGGATG
C290A	CCGATGACgcCACCATTGGCTACATTGTAGAG
	AATGGTGgcGTCATCGGGGAGCCGGATG
C290S	CCGATGACTcCACCATTGGCTACATTG
	CCAATGGTGgAGTCATCGGGGAGC
S312Q	CTCTTCCACcagCACCTAGAGAACCTGCAGCAGGTG
	CTCTAGGTGctgGTGGAAGAGGCCGCTCCGG
H313A	CTTCCACTCCgcCCTAGAGAACCTGCAGCAGGTG
	GTTCTCTAGGgcGGAGTGGAAGAGGCCGCTCC
L314R	CCACTCCCACagAGAGAACCTGCAGCAGGTG
	GGTTCTCTctGTGGGAGTGGAAGAGGCCGCTC
G334H	CTGAGCTATcaCATGTTTGAGAACAAGCGGAACG
	CTCAAACATGtgATAGCTCAGGGTCACCTGCTC
H242N + V244T	GAGAACAAAaccAGACCTGTCCACTTTTGGTTTGC
	ACAGGTCTggtTTTGTTCTCGCTGATCCGTTCTG
S312Q + H313A	CTTCCACCAGgcCCTAGAGAACCTGCAGCAGGTG
	GTTCTCTAGGgcCTGGTGGAAGAGGCCGCTCC
S312Q + H313A + L314R	CCACCAGGCCagAGAGAACCTGCAGCAGGTG
	GGTTCTCTctGGCCTGGTGGAAGAGGCCGC

PCR Mutagenesis Program	
1	95° C 30 seconds
2	95° C 30 seconds
3	60° C 1 minute
4	68° C 8 minutes (1 minute/kb)
5	(Repeat steps 2-4) × 18
6	68° C 10 minutes

PBS-Tween and then washed three times for 15, 5, and 5 minutes in PBS-Tween, followed by incubation in a 1:10,000 dilution of the Alexa-Fluor 660 goat anti-rabbit IgG (H+L) antibody (Molecular Probes) in PBS-Tween for 1 hour in the dark, after which the blot was rinsed and washed as before. The blot was then exposed using the Odyssey infrared imaging system (LI-COR) using the 700 nm channel, 169 µm resolution setting, set to medium quality, and intensity 5.0. The bands were quantified using the Odyssey software with a reciprocal fit to the standard curve.

Enzyme Assays

Lfng assays were adapted from previously published assays (Rampal, *et al.*, 2005b). The pNP-fucose substrate was prepared in 50 mM HEPES pH 6.8, 10 mM $MnCl_2$, 30% $^{v}/_{v}$ DMSO. Then 3.5 µL (3.5 U) of calf intestinal alkaline phosphatase (Roche) was added. Finally, 200 ng of Lfng enzyme in 20% $^{v}/_{v}$ glycerol was added to initiate the assay, with a final volume of 50 µL. The samples were eluted from the C18 columns by washing three times with 500 µL of 80% methanol. The eluted samples were then vortexed with 5 mL of Scintiverse II (Fisher), and scintillation counted in a Beckman-Coulter LS6500 scintillation counter for 1 minute per sample. The pNP-fucose saturation curves were performed with concentrations of pNP-fucose between 5 and 100 mM and 400 µM UDP-GlcNAc (0.5 µCi of UDP-[6-^{3}H]-GlcNAc (American Radiolabeled Chemicals (ARC) 60 Ci/mmol). The UDP-GlcNAc saturation curve was measured with UDP-GlcNAc concentrations of 1, 2, 5, 10, 20, 100, 250, and 500 µM with a pNP-fucose concentration of 100 mM. UDP inhibition curves were performed with UDP-GlcNAc concentrations of 5, 20, 100, 250 and 500 µM with UDP concentrations of 0, 2, 5, and 10 µM and pNP-fucose concentrations of 5 and 100 mM. UMP inhibition curves were performed with UDP-GlcNAc concentrations of 5, 20, 100, 250, and 500 µM, UMP concentrations of 0, 20, 50, and 100 µM and a pNP-fucose concentration of 100 mM. UDP-Glc utilization assays were performed with 100 µM UDP-Glc (0.5 µCi of UDP-[6-^{3}H]-Glc (ARC 60 Ci/mmol) and 100 mM pNP-fucose. The V_{max} values for each enzyme variant were normalized using a standard source of Lfng with saturating pNP-fucose in triplicate.

Fitting of Enzyme Curves

The efficiency of ^{3}H counting was determined by preparation of a calibration curve from known amounts of ^{3}H. The equation of this curve was used to determine the disintegrations per minute, which were converted into velocities. Velocities in pmol/L/s and substrate concentrations in pM were fit to Michaelis-Menten curves using the program EnzFitter (Biosoft®). UDP and UMP inhibition data were fit to various inhibition curves using EnzFitter [36]. All fits used the Marquardt-Levenberg algorithm (Levenberg, 1944, Marquardt, 1963).

UDP-hexanolamine Agarose Assay

Between 400 ng and 1 µg of enzyme was incubated 30 minutes on ice with approximately an 11.87-fold excess of UDP-hexanolamine agarose beads by volume in 50 mM HEPES pH 6.8, 10 mM $MnCl_2$ with and without 100 mM EDTA. The UDP-hexanolamine agarose was a generous gift of Dr. Gerald Hart at Johns Hopkins University, and the agarose was estimated to be conjugated with hexanolamine at a

concentration of 11 μmol/mL. The beads were equilibrated in 50 mM HEPES pH 6.8, 10 mM $MnCl_2$ prior to use. Subsequent to the incubation on ice, the beads were pelleted, the supernatant removed, the bead volume brought back to an equivalent of the supernatant, and samples removed and mixed with 5 × SDS gel loading buffer. In the case of some mutants, we could not produce 1 μg of enzyme in a small enough volume. For this reason, the amount of each sample loaded on the gel was normalized to the lowest concentration sample so that in each case an equivalent amount of total enzyme was present. The electrophoresis and western immuno-blotting was performed as described under enzyme quantification.

Results

Generation of a Lunatic Fringe Homology Model

Elimination of Lfng causes a significant somitogenesis phenotype when knocked out in mice (Evrard, *et al.*, 1998, Zhang and Gridley, 1998), while Rfng has no known phenotype (Zhang, *et al.*, 2002) and the phenotype of a Mfng knockout has not yet been published. Additionally, a mutation in human Lfng has recently been shown to cause the human genetic disorder SCD (Sparrow, *et al.*, 2006). Finally, previous work has shown that Mfng and Rfng show significantly lower activity toward EGF-*O*-fucose *in vitro* than Lfng (Rampal, *et al.*, 2005b). For these reasons, we decided to focus our efforts on understanding the catalysis and substrate recognition of mouse Lfng. Mouse Lfng has a stretch of 285 residues with 49.8% identity and 71.6% similarity to Mfng, and these enzymes are 52.5% identical and 72.2% similar over the stretch of sequence for which density was observed in the Mfng structure (Figure 5). This high degree of identity made it possible for us to create a homology model for Lfng on the basis of the Mfng structure (Figure 6).

Development of Conditions to Allow Acceptor Substrate Saturation

Before detailed kinetic analysis of Lfng could be performed, conditions for saturation of the enzyme with both donor and acceptor substrates needed to be achieved. Although saturation of Lfng with the donor substrate UDP-GlcNAc has been previously reported (Rampal, *et al.*, 2005b), saturation with acceptor substrate had not. This is mainly because of the difficulty in obtaining high enough concentrations of acceptor substrates to achieve saturation. We have discovered that Fng enzymatic activity is surprisingly resistant to high concentrations of DMSO. Indeed, when utilizing a small molecule acceptor such as pNP-fucose, assays containing DMSO concentrations as high as 60% show activity above what can be observed in the absence of DMSO (Figure 7). We have found that Lfng activity with the small molecule acceptor peaks between 30% and 50% DMSO. Thus, we have chosen 30% for our standard assays (Figure 7).

Using these conditions we have now been able to saturate with the acceptor substrate pNP-fucose (Figure 8). Interestingly, the activity curve of Lfng decreased at saturating concentrations of pNP-fucose, suggesting product inhibition (Figure 8A). Since UDP is a potent inhibitor of many glycosyltransferases, we repeated the assay in the presence of alkaline phosphatase to degrade any UDP formed during the assay. Addition of alkaline phosphatase prevented the decrease in activity at high pNP-fucose

Figure 5

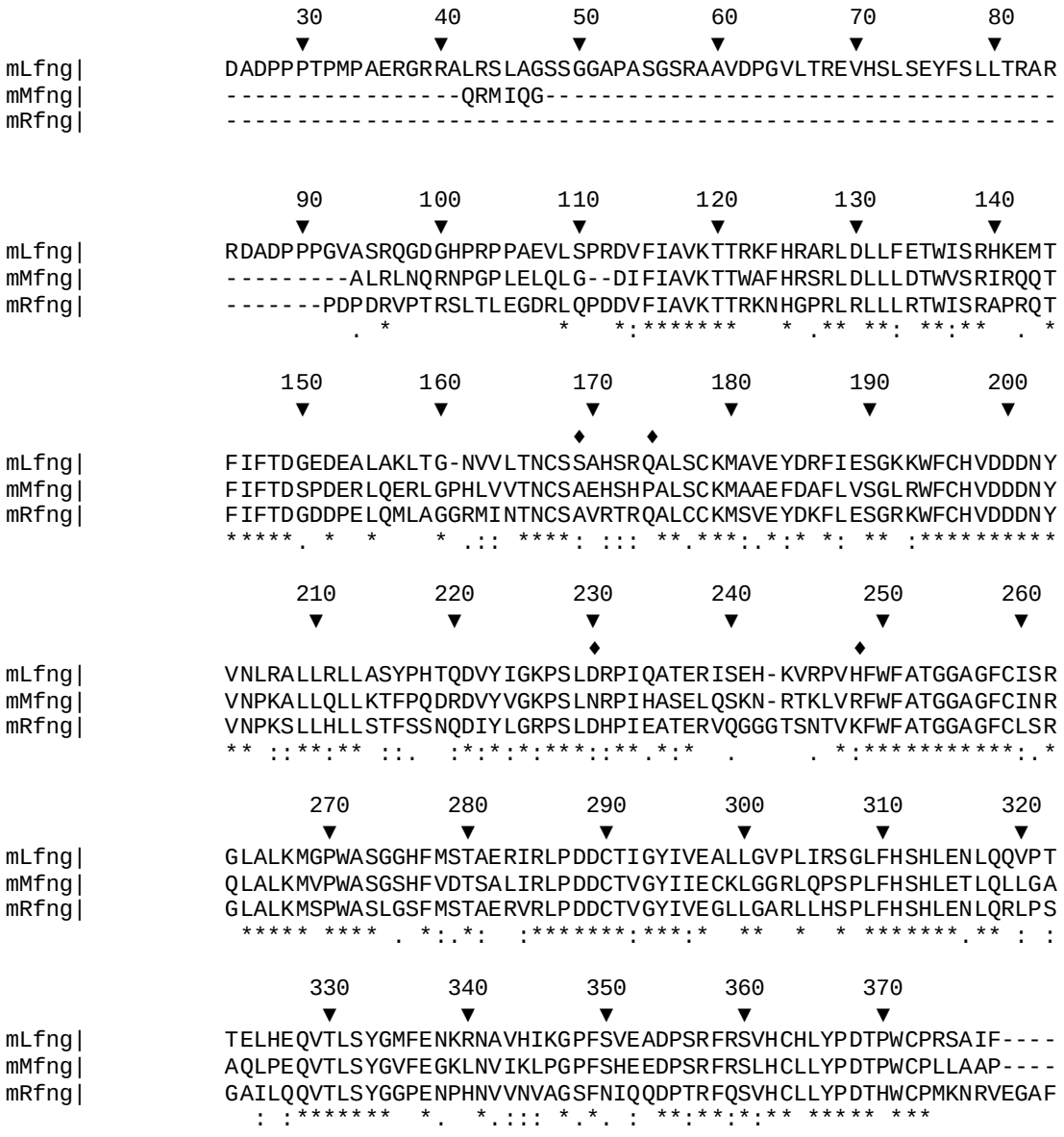


Figure 5. ClustalW Sequence Alignment of Mouse Fringe Enzymes
The loops that were not observed in the Mfng structure (Jinek, *et al.*, 2006) are indicated by diamonds at the first and last disordered residue above the sequence. Sequence conservation is indicated below the sequence with an asterisk indicating identical residues, two dots strongly type-conserved, and a single dot weakly type-conserved residues. The numbers above the sequence alignment indicate the Lfng sequence position.

Figure 6

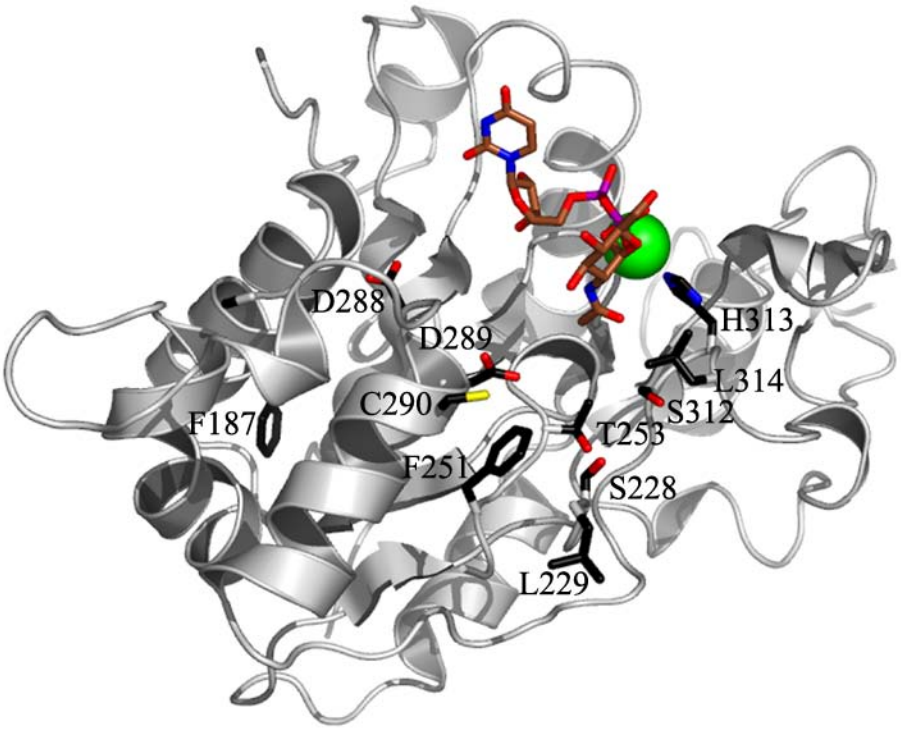


Figure 6. Homology Model of Lunatic Fringe
The Lfng homology model is shown in a white cartoon representation. The amino acid side-chains of residues mutated in this study are shown in black sticks and all-bonds coloring. Residues present in the two loops that were mutated are not observable in this figure. The UDP-GlcNAc donor is shown in brown sticks and all-bonds coloring. The Mn^{2+} ion is shown as a green sphere.

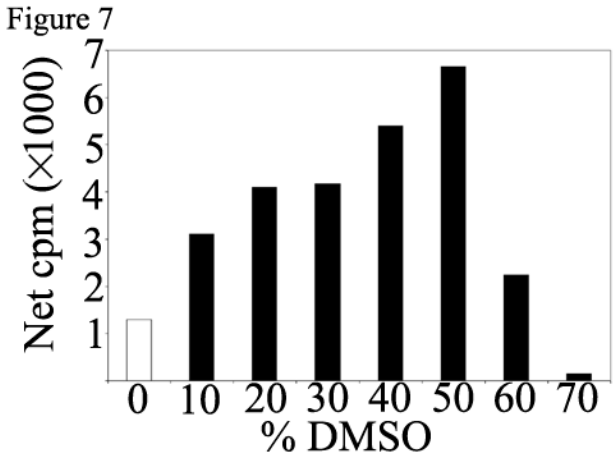


Figure 7. The Effect of DMSO on Lunatic Fringe Activity With 4-Nitrophenyl-α-L-fucose

Lfng activity with 10 mM pNP-fucose acceptor substrate, 100 μM UDP-[6-^{3}H]-GlcNAc donor substrate as a function of increasing DMSO concentrations. Lfng activity peaks at 50% and then drops precipitously. The bar for 0% DMSO is colored white because this value was calculated from a separate experiment requiring the use of a pNP-fucose stock dissolved in H_2O rather than DMSO.

Figure 8

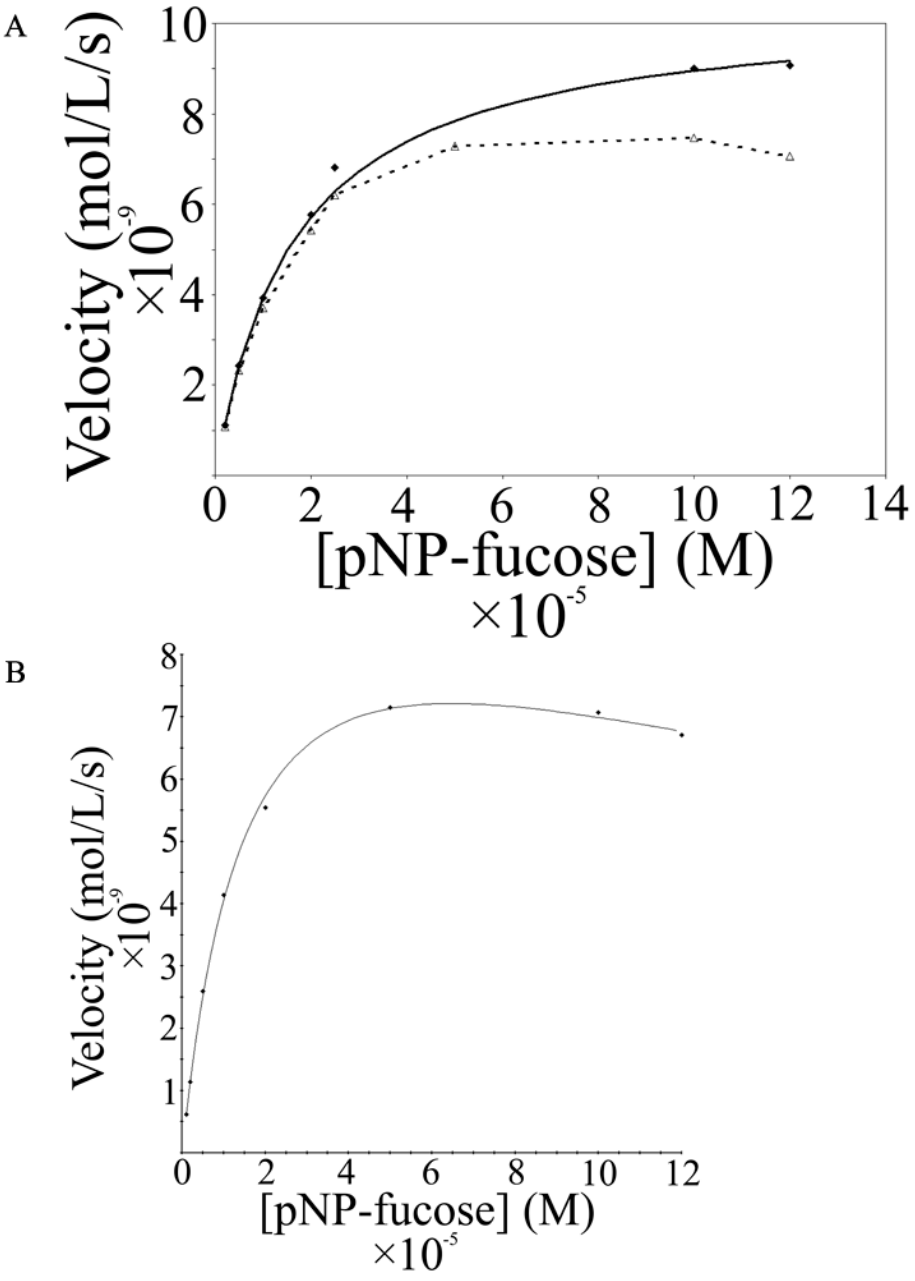


Figure 8. Michaelis-Menten Curves For Lunatic Fringe Activity with 4-Nitrophenyl-α-L-fucose

A) pNP-fucose saturation curve for WT Lfng with sub-saturating 100 μM UDP-GlcNAc concentration with and without alkaline phosphatase present in the assay. The data for the assays with alkaline phosphatase are shown with solid diamonds, and the Michaelis-Menten fit from the program EnzFitter (BioSoft®) is shown with a solid line. The data for the assays without alkaline phosphatase are shown with open triangles and a broken line. **B)** The substrate inhibition fit from the program EnzFitter (Biosoft®) of a pNP-fucose saturation curve with saturating (400 μM) UDP-GlcNAc.

concentrations, and the data could be fit to a hyperbolic Michaelis-Menten curve (Figure 8A). Thus, alkaline phosphatase was included in our standard assay.

Kinetic Characterization of the Wild-type Lunatic Fringe Enzyme

Initially we performed a thorough kinetic analysis of wild type Lfng using the conditions described above (Table 3). The Mfng structure showed no density for the GlcNAc portion of the UDP-GlcNAc substrate suggesting that in the absence of acceptor substrate, Mfng interacts mainly with the nucleotide portion of the donor (Jinek, *et al.*, 2006). The Rossman-like fold of glycosyltransferases is known as a nucleotide binding domain so it is not surprising that glycosyltransferases are strongly inhibited by the nucleotide product generated after glycan transfer. In the Golgi, a nucleoside diphosphatase rapidly degrades UDP to UMP to eliminate this inhibition (Novikoff and Goldfischer, 1961). We characterized the inhibition of Lfng by UDP and UMP and found both to be competitive inhibitors of the enzyme (Figure 9), with a K_i of 11.04 μM for UDP and K_i of 96.35 μM for UMP (Table 3).

When we repeated the pNP-fucose saturation curve at high levels of UDP-GlcNAc (approximately 10 fold higher than the K_M) we again noticed a dip in the activity curve at the highest concentrations of pNP-fucose acceptor (Figure 8B), suggesting product inhibition as we had seen before (Figure 8A). In an attempt to remove the GlcNAc-β1,3-Fuc-α1,*O*-pNP (pNP-disaccharide) product from the reaction, we added UDP-galactose and β1,4-galactosyltransferase as a coupling enzyme to the assay. This had no effect on the inhibition seen at high pNP-fucose levels (data not shown). However, these data could be fit to a substrate inhibition model, thus suggesting that at very high levels of UDP-GlcNAc the enzyme is inhibited by the donor (Figure 8B). The substrate inhibition fit gives a value of 257.3 μM for the competitive inhibition constant of UDP-GlcNAc (Table 3).

The structural information and docking results suggest that the mechanism must be ordered, with the donor substrate binding first since the UDP-GlcNAc donor-substrate is likely buried beneath the EGF-*O*-fucose acceptor-substrate (Figure 12A,B). The observed inhibition by UDP suggests that a dead-end complex forms in the absence of a coupling enzyme (alkaline-phosphatase or nucleoside diphosphatase) to eliminate the UDP product.

We were unable to saturate the enzyme with an EGF-*O*-fucose acceptor substrate despite reaching a maximum concentration in excess of 190 μM in the assay (Figure 10A). We estimate the K_M for EGF-*O*-fucose from these data to be approximately 2 mM from a Hanes-Woolf plot (data not shown). This is unexpectedly high, given that the EGF-*O*-fucose is a much better substrate for the enzyme than the pNP-fucose (Rampal, *et al.*, 2005b). However, the extrapolated k_{cat} for the enzyme with these data would be approximately 2 s^{-1}, which is considerably higher than the value of 0.08 that we see with pNP-fucose. We had previously published a specific activity for Lfng with 5 μM Factor IX EGF-*O*-fucose of 16 nmol/min/mg (Rampal, *et al.*, 2005b). In the current study we found a similar value of 13.5 nmol/min/mg. We see activity of approximately 309 nmol/min/mg at approximately 190 μM EGF-*O*-fucose compared to approximately 132 nmol/min/mg when saturated with pNP-fucose.

We also observed a distortion in the data when using the EGF-*O*-fucose acceptor substrate. If we linearize the data for a Hanes-Woolf plot, at low EGF-*O*-fucose

Table 3. Kinetic Data for Wild-type and Mutant Lunatic Fringe

All K_M and V_{max} values reported here were produced from the appropriate fit using EnzFitter (Biosoft®). The V_{max} values reported here are per mg of enzyme. All assays were performed with 110 nM enzyme in a 50 µL volume. All errors reported are standard error values from the corresponding fit produced with EnzFitter (Biosoft®). Due to variation in Lfng activity from day to day, all V_{max} values were normalized to a standard Lfng assay using a standard Lfng source. The percent wild-type activity values for the donor specificity mutants were produced by dividing mutant enzyme activity utilizing UDP-GlcNAc or UDP-Glc by the wild-type Lfng activity with the appropriate donor. †In some cases activity was too low to measure below saturating pNP-fucose concentrations and a Michaelis-Menten curve could not be produced. In these cases we report the velocity at 100 mM pNP-fucose rather than V_{max}, and no value for K_M which could not be determined.

Wild Type						
V_{max} (µmol/L/s)	43.93 ±0.31	**UDP GlcNAc**		**pNP-fucose**		
k_{cat} (s^{-1})	0.08	**K_M (µM)**	37.94 ±2.47	**K_M (mM)**	10.91 ±0.65	
UDP Inhibition		**UMP Inhibition**		**UDP-GlcNAc Inhibition**		
K_i (µM)	11.04 ±2.50	**K_i (µM)**	96.35 ±22.68	**K_i (µM)**		257.3 ±40.5
Mutants That Did Not Express						
Mutant	R173A	D289N	D289S	C290A	S312Q	
Catalytically Inactive Active Site Mutants						
Mutant	S228L	S228Y	F251S	T253A	D289E	
Active Site Mutant to Short Loop Side of Aspartate 289						
Mutant	G254A	D288A				
V_{max} (µmol/L/s)	1.06 ±0.01	0.3†				
pNP-Fuc K_M (mM)	17.7 ±0.6	--------				
Active Site Mutants to Long Loop Side of Aspartate 289						
Mutant	S228A	S228T	F251Y	C290S		
V_{max} (µmol/L/s)	28.67 ±0.12	3.83 ±0.02	24.6 ±2.6	42.25 ±0.30		
pNP-Fuc K_M (mM)	117.7 ±0.7	59.5 ±0.7	129 ±20	111.1 ±1.3		
Short Loop Mutants						
Mutant	S168A	S168V	H171A	H171D	A175V	L176A
V_{max} (µmol/L/s)	5.29 ±0.03	4.27 ±0.03	0.31 ±0.04	2.96 ±0.08	1.48 ±0.01	0.2†
pNP-Fuc K_M (mM)	20.4 ±0.3	11.6 ±0.4	50.7 ±11.1	18.2 ±1.6	27.7 ±0.5	--------

UDP-GlcNAc Saturation Curves for Short Loop Mutants					
Mutant	S168A	S168V	H171D	A175V	
UDP-GlcNAc K_M (μM)	34.33 ±2.56	49.73 ±1.44	60.9 ±7.2	70.7 ±6.4	
Long Loop Mutants					
Mutant	L229Q	I233A	A235Y	E237A	
V_{max} (μmol/L/s)	15.4 ±1.5	10.4 ±0.1	8.34 ±0.11	4.315 ±0.115	
pNP-Fuc K_M (mM)	113 ±17	57.7 ±0.4	61.4 ±1.6	36.33 ±2.31	
Donor Specificity Mutants					
Mutant	S312T	H313A	L314R	G334H	H313A + G334H
% of wt UDP-GlcNAc Utilization	1.9 ±0.1	6.2 ±0.2	21.7 ±1.0	94.6 ±0.4	3.1 ±0.3
% of wt UDP-Glc Utilization	18.1 ±3.5	75.9 ±5.3	119.5 ±11.5	78.0 ±2.2	34.7 ±2.6
Activity Comparison	Glc > GlcNAc	Glc > GlcNAc	Glc > GlcNAc	GlcNAc > Glc	Glc > GlcNAc

Figure 9

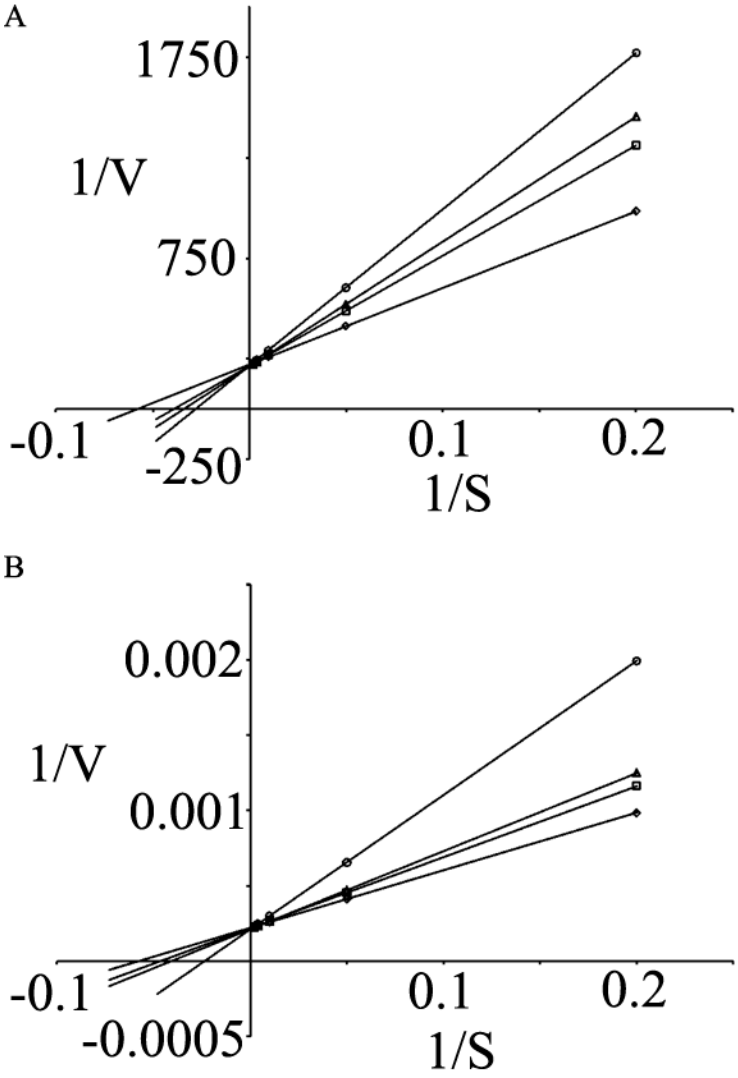


Figure 9. Reciprocal plots of UDP and UMP Inhibition of Lunatic Fringe
A) Lineweaver-Burk plot of UDP competitive inhibition of Lfng with varying UDP-GlcNAc concentration and saturating (100 mM) pNP-fucose concentration. **Diamond**: 0 μM UDP; **Square**: 2 μM UDP; **Triangle**: 5 μM UDP; **Circle**: 10 μM UDP **B)** Lineweaver-Burk plot of UMP competitive inhibition of Lfng with varying UDP-GlcNAc concentration and saturating (100 mM) pNP-fucose concentration. **Diamond**: 0 μM UMP; **Square**: 20 μM UMP; **Triangle**: 50 μM UMP; **Circle**: 100 μM UMP

Figure 10

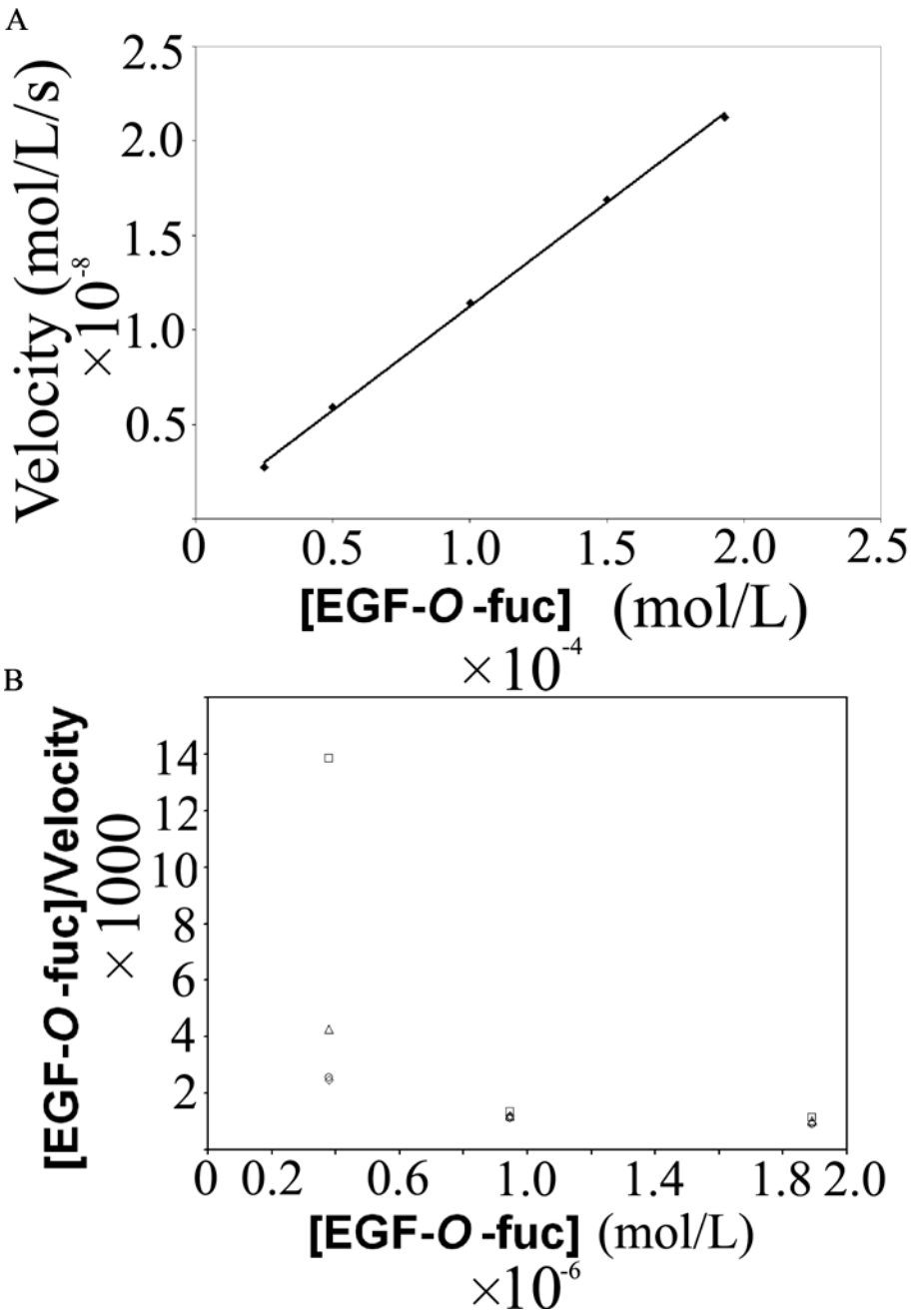


Figure 10. Lunatic Fringe Assays With Factor IX EGF-*O*-fucose as Acceptor Substrate.

A) Assay of Lfng activity with increasing EGF-*O*-fucose concentrations from 25 μM to 193 μM with 378 nM Lfng. **B)** A Hanes-Woolf plot of Lfng activity with increasing enzyme concentration of 662 nM (**square**), 1.32 μM (**triangle**), 3.97 μM (**diamond**), and 9.26 μM (**circle**).

concentrations (below approximately 25 μM) the data are non-linear, and have a negative slope, rather than the expected positive slope (Figure 10B). We observed that with increasing enzyme concentration, the data gradually trends toward the expected linear relationship, and the slope switches from negative to positive (Figure 10B). These distortions from the expected linear relationship could suggest some form of slow-on/slow-off kinetics with the EGF-*O*-fucose acceptor substrate (Bieth, 1995). The small molecule pNP-fucose acceptor substrate does not show these non-linear affects at low concentrations indicating this is a consequence of using the EGF-*O*-fucose substrate and not a more generalized effect.

Because of the low k_{cat} with pNP-fucose, we were concerned that our purified Lfng might be of low quality, and the catalytic efficiency of a given aliquot of our enzyme was artificially depressed for that reason. To address this, we used UDP-hexanolamine agarose to pellet the catalytically competent enzyme. Visualization of the supernatant and pellet by western immuno-blot indicates that all enzymatic activity can be removed from the enzyme aliquot only when all Lfng protein is in the agarose pellet (Figure 11). This confirms that the vast majority of our purified enzyme is catalytically active and properly folded. We have performed this experiment on enzyme that has been stored at -80° C for over a year and achieved the same result indicating that the enzyme is stable for long-term storage under our freezing conditions in glycerol.

Docking Factor IX EGF-O-fucose Onto the Lunatic Fringe Homology Model

In order to gain an understanding of how Lfng interacts with an *O*-fucosylated EGF-repeat, we docked a previously described (Rampal, *et al.*, 2005b) model of Factor IX EGF-repeat bearing *O*-fucose onto the Lfng model using the HEX program (Ritchie, 2003). This program docks the ligand based primarily on shape complementarity. After manually culling the lowest energy solutions of any docked EGF-repeats where the Fuc was not in proximity to the active site we retained 80 solutions. The docked EGF-repeats were observed to cluster in roughly two groups, with an approximate 180° rotation in orientation (Figure 12A, B). The different orientations placed the Fuc on opposite sides of the catalytic aspartate 289 (Figure 12C, D). Interestingly, both docking solutions reveal a cavity for the donor substrate UDP-GlcNAc between the enzyme and docked EGF-repeat, adding weight to the reliability of these models (Figure 12C, D).

Analysis of Lunatic Fringe Mutants

We used the docked EGF-repeat solutions on the Lfng homology model and sequence alignment data to inform our choice of mutants. We aligned all sequences annotated as Fng or Fng-like enzymes in the NCBI protein database excluding any plant sequences (Figure 13). Although proteins annotated as Fng-like proteins may not be Fng enzymes, we found that when we included the more distantly related Fng-like sequences that several residues stood out as extremely conserved. The most obvious example was the DXD motif, which is conserved across many glycosyltransferases (Wiggins and Munro, 1998). Mutating this well characterized motif is known to abrogate enzymatic activity in a number of glycosyltransferases including both *Drosophila* Fng (Moloney, *et al.*, 2000a) and mouse Mfng (Chen, *et al.*, 2001). However, serine 228, and serine 312 were two other residues in the vicinity of the active site that also showed this very high

Figure 11

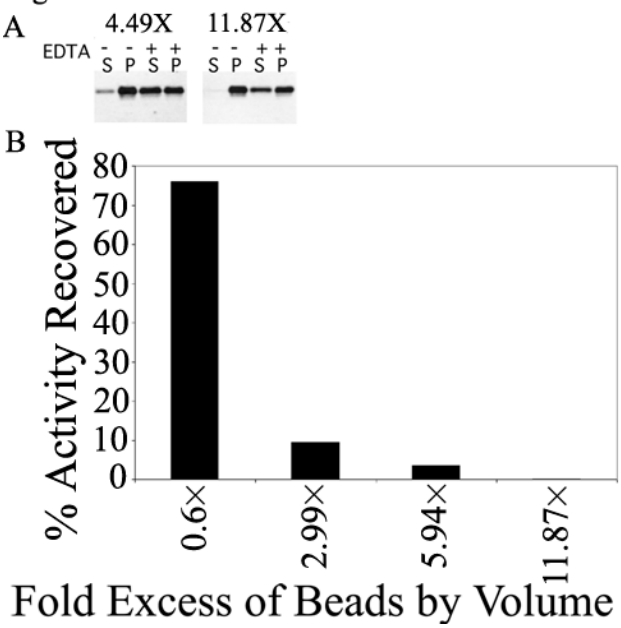


Figure 11. UDP-hexanolamine Experiment With Wild-type Lunatic Fringe
A) Western immuno-blot of Lfng incubated with UDP-hexanolamine agarose (S supernatant, P pellet). The experiment was performed with and without 100 mM EDTA which will prevent UDP-hexanolamine from interacting with Lfng. In the absence of EDTA, all of the Lfng is pulled down by the beads at the higher concentration. EDTA inhibits the process, thus demonstrating its specificity. **B)** A plot of Lfng activity remaining in the supernatant after incubation with increasing amounts of UDP-hexanolamine agarose. At the highest concentration of UDP-hexanolamine agarose, no Lfng remains in the supernatant as demonstrated by a complete absence of activity. At lower concentrations, residual Lfng activity remains in the supernatant.

Figure 12

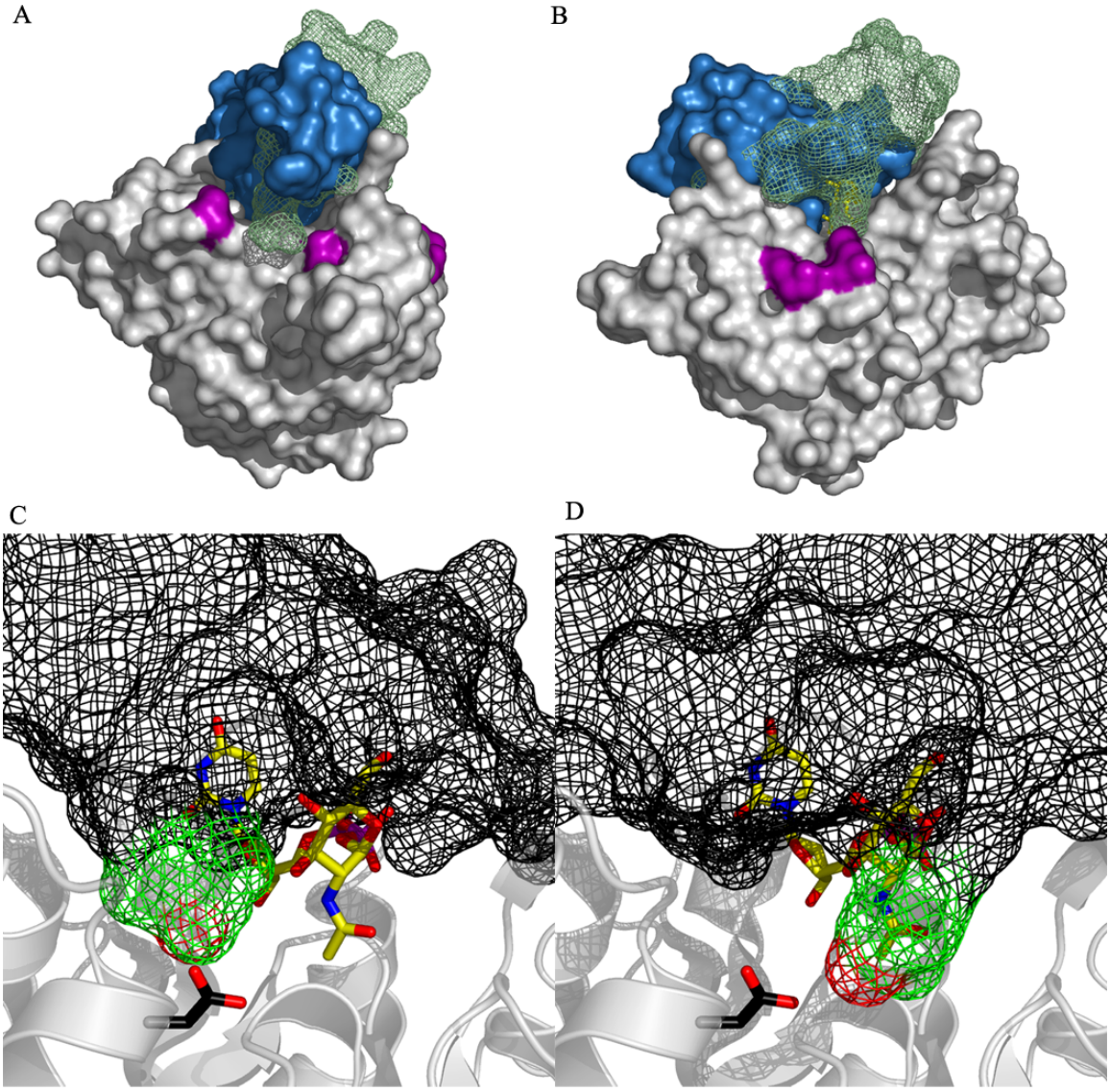
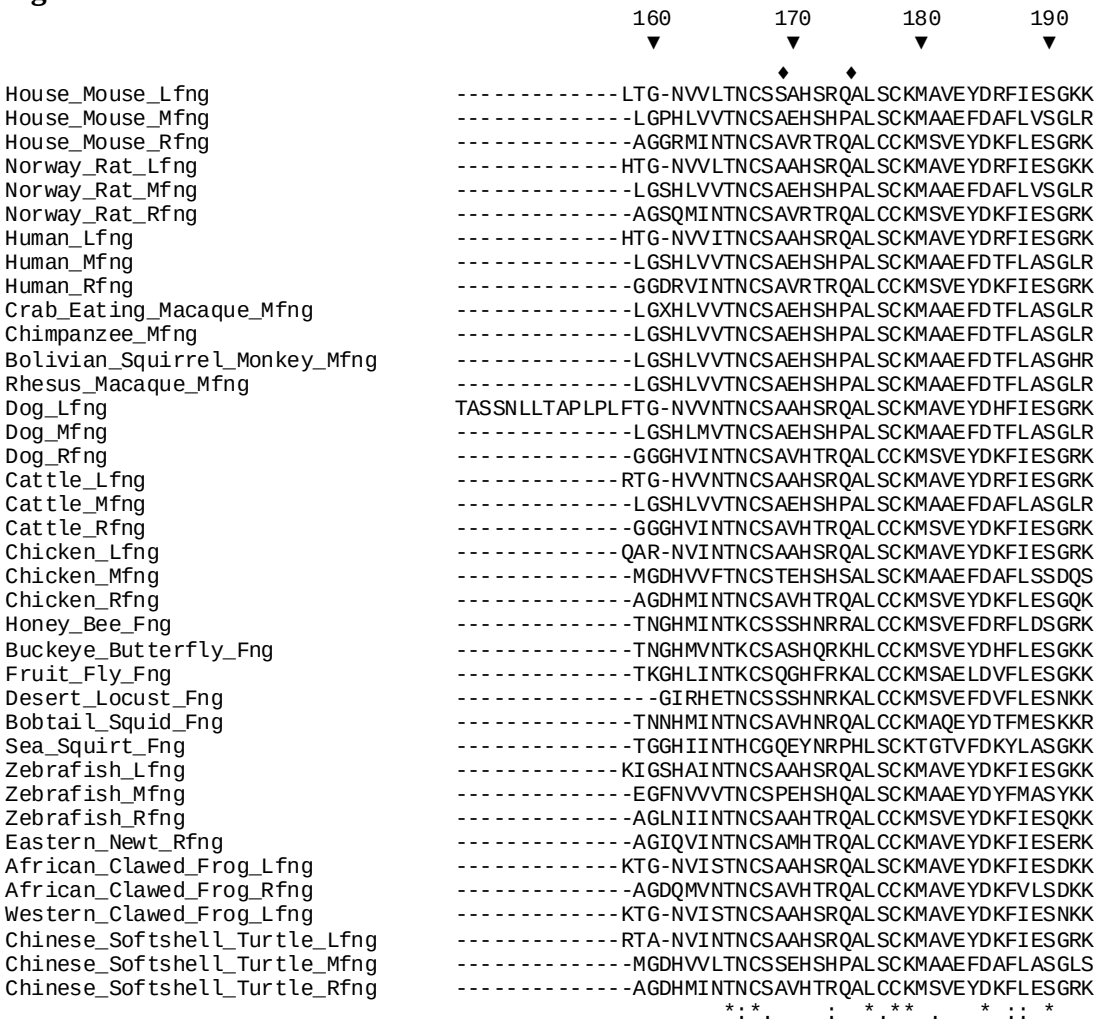


Figure 12. Docking of an EGF-*O*-fucose Onto the Lunatic Fringe Homology Model A and B) The docked EGF-*O*-fucose solutions form roughly two clusters. The Lfng surface is in white, termini of the two loops not observed in the Mfng structure are colored purple on the Lfng surface. One EGF-*O*-fucose cluster is in surface representation in blue, the other in green mesh. (Panel B is rotated approximately 90° clockwise with respect to A). **C and D)** The Fuc is observed to cluster on either side of the catalytic aspartate in the docked solutions. EGF is in black mesh, with the Fuc colored green, and O3 of the Fuc colored red. Lfng is shown as a grey cartoon, while the catalytic aspartate is colored black in stick form with all-bonds coloring. UDP-GlcNAc (yellow) is in stick representation and all-bonds coloring. Panel C corresponds to the solid blue EGF-*O*-fucose orientation in panels A and B and panel D corresponds to the green mesh EGF-*O*-fucose orientation in panels A and B.

Figure 13


160
170
180
190
House_Mouse_Lfng
House_Mouse_Mfng
House_Mouse_Rfng
Norway_Rat_Lfng
Norway_Rat_Mfng
Norway_Rat_Rfng
Human_Lfng
Human_Mfng
Human_Rfng
Crab_Eating_Macaque_Mfng
Chimpanzee_Mfng
Bolivian_Squirrel_Monkey_Mfng
Rhesus_Macaque_Mfng
Dog_Lfng
Dog_Mfng
Dog_Rfng
Cattle_Lfng
Cattle_Mfng
Cattle_Rfng
Chicken_Lfng
Chicken_Mfng
Chicken_Rfng
Honey_Bee_Fng
Buckeye_Butterfly_Fng
Fruit_Fly_Fng
Desert_Locust_Fng
Bobtail_Squid_Fng
Sea_Squirt_Fng
Zebrafish_Lfng
Zebrafish_Mfng
Zebrafish_Rfng
Eastern_Newt_Rfng
African_Clawed_Frog_Lfng
African_Clawed_Frog_Rfng
Western_Clawed_Frog_Lfng
Chinese_Softshell_Turtle_Lfng
Chinese_Softshell_Turtle_Mfng
Chinese_Softshell_Turtle_Rfng

```
                                    200       210       220       230       240
                                     ▼         ▼         ▼         ▼         ▼
                                                                   ♦
House_Mouse_Lfng                    WFCHVDDDNYVNLRALLRLLASYPHTQDVYIGKPSLDRPIQATERI-SHE
House_Mouse_Mfng                    WFCHVDDDNYVNPKALLQLLKTFPQDRDVYVGKPSLNRPIHASELQ-SKN
House_Mouse_Rfng                    WFCHVDDDNYVNPKSLLHLLSTFSSNQDIYLGRPSLDHPIEATERVQGGG
Norway_Rat_Lfng                     WFCHVDDDNYVNLRALLRLLASYPHTQDVYIGKPSLDRPIQATERI-SEH
Norway_Rat_Mfng                     WFCHVDDDNYVNPKALLQLLRTFPQDHDVYVGKPSLNRPIHASELQ-SKN
Norway_Rat_Rfng                     WFCHVDDDNYVNPKSLLHLLSTFSSNQDIYLGRPSLDHPIEATERVQGGG
Human_Lfng                          WFCHVDDDNYVNLRALLRLLASYPHTRDVYVGKPSLDRPIQAMERV-SEN
Human_Mfng                          WFCHVDDDNYVNPRALLQLLRAFPLARDVYVGRPSLNRPIHASEPQ-PHN
Human_Rfng                          WFCHVDDDNYVNARSLLHLLSSFSPSQDVYLGRPSLDHPIEATERVQGGR
Crab_Eating_Macaque_Mfng            WFCHVDDDNYVNPRALLQLLKGFPLDSDVYVGRPSLNRPIRASEPQ-PHN
Chimpanzee_Mfng                     WFCHVDDDNYVNPRALLQLLRAFPLAHDVYVGRPSLNRPIHASEPQ-PHN
Bolivian_Squirrel_Monkey_Mfng       WFCHVDDDNYVNARALLQLLRAFPLARDVYVGRPSLNRPIHASEPR-PHN
Rhesus_Macaque_Mfng                 WFCHVDDDNYVNPRALLQLLKGFPLDSDVYVGRPSLNRPIRASEPQ-PHN
Dog_Lfng                            WFCHVDDDNYVNVRALLRLLASYPHTQDVYIGKPSLDRPIQATERV-SEN
Dog_Mfng                            WFCHVDDDNYVNPRALLRLLKTFPQTRDVYLGRPSLNRPIRASEPR-PHN
Dog_Rfng                            WFCHVDDDNYVNPKGLLQLLATFSPSQDVYLGRPSLDHPIEAAERVQGGG
Cattle_Lfng                         WFCHVDDDNYVNVRALLRLLGSYPHTQDVYLGKPSLDRPIQATERV-SEN
Cattle_Mfng                         WFCHVDDDNYLNPRALLKLLKTFPQTRDVYVGRPSLNRPIHASEPQ-PHN
Cattle_Rfng                         WFCHVDDDNYVNPKGLLQLLSTFSPSQDIYLGRPSLDHPIEATERIQGGG
Chicken_Lfng                        WFCHVDDDNYVNVRTLVKLLSSYPHTQDIYIGKPSLDRPIQATERI-SEN
Chicken_Mfng                        WFCHLDDDNYLNPEALLKLLSSYSAMKDVYVGKPSLNRPIRASETL-PNN
Chicken_Rfng                        WFCHVDDDNYVNPRTLLRLLSAFSPSQDVYVGRPSLDHPIEAADHVQSDG
Honey_Bee_Fng                       WFCHFDDDNYVNVPRLLKLLDNYNPREDWYLGRPSIPAPLEIIRQDKKKE
Buckeye_Butterfly_Fng               WFCHFDDDNYVNIPRLISVLQTYNHQEDWYLGRTSVYEPVKIY---KKPT
Fruit_Fly_Fng                       WFCHFDDDNYVNVPRLVKLLDEYSPSVDWYLGKPSISSPLEIHLDSKNTT
Desert_Locust_Fng                   WFCHVDDDNYVNVPRLVRVLSGYNPQQDWYLGKPSIRAPLEILN---RDN
Bobtail_Squid_Fng                   WFCHVDDDTYVNVPKLVTVLQKYNHTKDWYLGKPSLRHPIEIMD---RDN
Sea_Squirt_Fng                      WWCRFDDDNYVNPPRLVNLVNGYNWTQDICIGKLSVPS-------YTANY
Zebrafish_Lfng                      WFCHVDDDNYVNTKTLVKLLSNYPHTQDMYIGKPSLDRPIEATERL-GDN
Zebrafish_Mfng                      WLCHVDDDNYLNPGALLSLLMAFPADGDIYVGKPSLDRPMRAQELL-EGN
Zebrafish_Rfng                      WFCHVDDDNYVILPSLLELLSSYSHTQDVYLGRPSLDHPIEAAERVKSDG
Eastern_Newt_Rfng                   WFCHVDDDNYVNLFSLRHLLASFSHSQDVYLGRPSLDHPIEAIERVKSDG
African_Clawed_Frog_Lfng            WFCHVDDDNYVNVRTLVKLLSRYSHTNDIYIGKPSLDRPIQATERI-SES
African_Clawed_Frog_Rfng            WFCHLDDDNYLNLHALLDLLSTFSHSTDVYVGRPSLDHPVETVDRMKGDG
Western_Clawed_Frog_Lfng            WFCHVDDDNYVNVQTLVKLLSRYSHTNDIYIGKPSLDRPIQATERI-SEN
Chinese_Softshell_Turtle_Lfng       WFCHVDDDNYVNVRTLVKLLSGYPHTQDIYIGKPSLDRPIQATERI-SEN
Chinese_Softshell_Turtle_Mfng       WFCHVDDDNYLNPQALLKLLSSYSPAQDVYIGKPSLNRPIRASEMM-PNN
Chinese_Softshell_Turtle_Rfng       WFCHVDDDNYVNPQTLLRLLSAFSHSQDVYVGRPSLDHPIEAADHVQSDG
                                    * *:.***.*:    *  ::  :    *  :*: *:
```

```
                                    250       260       270       280       290
                                    ▼         ▼         ▼         ▼         ▼
                                    ◆
House_Mouse_Lfng                    KVRPVHFWFATGGAGFCISRGLALKMGPWASGGHFMSTAERIRLPDDCTI
House_Mouse_Mfng                    RTKLVRFWFATGGAGFCINRQLALKMVPWASGSHFVDTSALIRLPDDCTV
House_Mouse_Rfng                    TSNTVKFWFATGGAGFCLSRGLALKMSPWASLGSFMSTAERVRLPDDCTV
Norway_Rat_Lfng                     RVRPVHFWFATGGAGFCISRGLALKMGPWASGGHFMSTAERIRLPDDCTI
Norway_Rat_Mfng                     RTRLVRFWFATGGAGFCINRQLALKMVPWASGSHFVDTSALIRLPDDCTV
Norway_Rat_Rfng                     TSNTVKFWFATGGAGFCLSRGLALKMSPWASLGSFMSTAERVRLPDDCTV
Human_Lfng                          KVRPVHFWFATGGAGFCISRGLALKMSPWASGGHFMNTAERIRLPDDCTI
Human_Mfng                          RTRLVQFWFATGGAGFCINRKLALKMAPWASGSRFMDTSALIRLPDDCTM
Human_Rfng                          TVTTVKFWFATGGAGFCLSRGLALKMSPWASLGSFMSTAEQVRLPDDCTV
Crab_Eating_Macaque_Mfng            RTRLVQFWFATGGAGXCINRKLALKMAPWASGSRFMDTSALIRLPDDCTM
Chimpanzee_Mfng                     RTRLVQFWFATGGAGFCINRRLALKMAPWASGSRFMDTSALIRLPDDCTM
Bolivian_Squirrel_Monkey_Mfng       RTRLVQFWFATGGAGFCINRKLALKMAPWASGSRFMDTSALIRLPDDCTV
Rhesus_Macaque_Mfng                 RTRLVQFWFATGGAGFCINRKLALKMAPWASGSRFMDTSALIRLPDDCTM
Dog_Lfng                            KVRPVHFWFATGGAGFCISRGLALKMSPWASRGHFMSTAERIRLPDDCTV
Dog_Mfng                            RTRQVQFWFATGGAGFCINRKLALKMAPWASGSRFVETSALIRLPDDCTV
Dog_Rfng                            TVTTVKFWFATGGAGFCLSRGLALKMSPWASLGSFMSTAERVRLPDDCTV
Cattle_Lfng                         KVRPVHFWFATGGAGFCISRGLALKMSPWASGGHFMSTAERIRLPDDCTI
Cattle_Mfng                         RTKLVQFWFATGGAGFCINRKLALKMAPWASGSHFMDTSALIRLPDDCTV
Cattle_Rfng                         TVTTVKFWFATGGAGFCLSRGLALKMSPWASLGGFMSTAERVRLPDDCTV
Chicken_Lfng                        KMHPVHFWFATGGAGFCISRGLALKMSPWASGGHFMSTAEKIRLPDDCTI
Chicken_Mfng                        QMKSVRFWFATGGAGFCISRKLARKMMPWASGKNFLSTSELIRLPDDCTI
Chicken_Rfng                        SKTSVKFWFATGGAGFCISRGLALKMSPWASLGNFISTAERVRLPDDCTI
Honey_Bee_Fng                       TNTKVKFWFATGGAGFCISRALAMKMTPVAGGGKFITVGDRIRLPDDVTM
Buckeye_Butterfly_Fng               NKLLFSFWFATGGAGFCISRSLALKMLPVASGGRFISICEGIRLPDDVSV
Fruit_Fly_Fng                       TNKKITFWFATGGAGFCLSRALTLKMLPIAGGGKFISIGDKIRFPDDVTM
Desert_Locust_Fng                   TAQKISFWFATGGAGFCLSRALALKMMPVASGGKFISIGEKIRLPDDVTM
Bobtail_Squid_Fng                   PGQKISFWFATGGAGFCISRSLALKMMPHAGDGRLMTVGEKIRLPDDCTV
Sea_Squirt_Fng                      RGRPEIYQFAHGGAGCCISRPLALKMQPWCGREKLVETTEDVGRHEDCTL
Zebrafish_Lfng                      KMRPVNFWFATGGAGFCISRGLALKMSPWASGGHFMNTAEKIRLPDDCTI
Zebrafish_Mfng                      KTRDVHFWFATGGAGFCLSRNLAERMAPWASGPRFEQTSAVIMLPDDCTV
Zebrafish_Rfng                      S-VSVKFWFATGGAGFCISRGLALKMSPWASLGNFITTAEKIRLPDDCTI
Eastern_Newt_Rfng                   S-ASVRFWFATGGAGFCISRGLALKMSPWASMGNFITTAELVRLPDDCTI
African_Clawed_Frog_Lfng            NMRPVNFWFATGGAGFCISRGLALKMSPWASGGHFMNTAEKIRLPDDCTI
African_Clawed_Frog_Rfng            S-GSLKFWFATGGAGFCISRGLALKMSPWASMGNFISTAEKVRLPDDCTI
Western_Clawed_Frog_Lfng            NMRPVNFWFATGGAGFCISRGLALKMSPWASGGNFMNTAEKIRLPDDCTI
Chinese_Softshell_Turtle_Lfng       KMHPVHFWFATGGAGFCISRGLALKMSPWASGGHFMSTAERIRLPDDCTI
Chinese_Softshell_Turtle_Mfng       QTRSVHFWFATGGAGFCISRRLATKMAPWASGSHFLSTSDLIRLPDDCTV
Chinese_Softshell_Turtle_Rfng       S-TTVKFWFATGGAGFCISRGLALKMSPWASLGNFISTAERVRLPDDCTI
                                          : ** **** *:.* *: :* * ..   :      :   :* ::
```

```
                                  300       310       320       330       340
                                   ▼         ▼         ▼         ▼         ▼
House_Mouse_Lfng                  GYIVEALLGVPLIRSGLFHSHLENLQQVPTTELHEQVTLSYGMF-ENKRN
House_Mouse_Mfng                  GYIIECKLGGRLQPSPLFHSHLETLQLLGAAQLPEQVTLSYGVF-EGKLN
House_Mouse_Rfng                  GYIVEGLLGARLLHSPLFHSHLENLQRLPSGAILQQVTLSYGGP-ENPHN
Norway_Rat_Lfng                   GYIVEALLGVPLIRSGLFHSHLENLQQVPTTELHEQVTLSYGMF-ENKRN
Norway_Rat_Mfng                   GYIIECKLGGRLQPSPLFHSHLETLQLLGTAQLLEQVTLSYGVF-EGKLN
Norway_Rat_Rfng                   GYIVEGLLGARLLHSPLFHSHLENLQKLPSGAVLQQVTLSYGGP-ENPHN
Human_Lfng                        GYIVEALLGVPLIRSGLFHSHLENLQQVPTSELHEQVTLSYGMF-ENKRN
Human_Mfng                        GYIIECKLGGRLQPSPLFHSHLETLQLLRTAQLPEQVTLSYGVF-EGKLN
Human_Rfng                        GYIVEGLLGARLLHSPLFHSHLENLQRLPPDTLLQQVTLSHGGP-ENPHN
Crab_Eating_Macaque_Mfng          GYIIECKLGGRLQPSPLFHSHLETLQLLRTAQLPEQV------------
Chimpanzee_Mfng                   GYIIECKLGGRLQPSPLFHSHLETLQLLRTAQLPEQVTLSYGVF-EGKLN
Bolivian_Squirrel_Monkey_Mfng     GYIIECKLGGRLQPSPRFHSHLETLQLLRTAELPEQVTLSYGVF-EGKLN
Rhesus_Macaque_Mfng               GYIIECKLGGRLQPSPLFHSHLETLQLLRTAQLPEQVTLSYGVF-EGKLN
Dog_Lfng                          GYIVEALLGVPLIRSGLFHSHLENLQQVPASELREQVTLSYGMF-ENKRN
Dog_Mfng                          GYIVECKLRGHLQPSPLFHSHLETLQLLGAAQLPEQVTLSYGVF-EGKLN
Dog_Rfng                          GYIVEGLLGARLLPSSLFHSHLENLQRLPPDAVLRQVTLSYGGP-DNPHN
Cattle_Lfng                       GYIVEALLGVPLVRCGLFHSHLENLQQVPASELHEQVTLSYGMF-ENKRN
Cattle_Mfng                       GYIVECKLGGHLQSCPLFHSHLETLQLLEAAQLPEQVTLSYGVF-EGKLN
Cattle_Rfng                       GYIVEGLLGARLLHSSLFHSHLENLQKLSPDTLLQQVTLSYGGP-ENPQN
Chicken_Lfng                      GYIIESVLGVKLIRSNLFHSHLENLHQVPKTEIHKQVTLSYGMF-ENKRN
Chicken_Mfng                      GYIIECKVGGQLLPNRLFHSHLENLQLIPTSDLMQQVTLSYGVF-ENKLN
Chicken_Rfng                      GYIIEGLLEVKLLHSPLFHSHLENLQRLQGESVLQQVTLSYGDP-ENKHN
Honey_Bee_Fng                     GYIIEHLLKKQLTVVEQFHSHLEPMKFLNKDTFDEQVSFSYSKGPRDEWN
Buckeye_Butterfly_Fng             GFIIEHLMKKNLTRVPEFHSHLEQMKLLMPETFRDQISFSYSK-SKNGWN
Fruit_Fly_Fng                     GFIIEHLLKVPLTVVDNFHSHLEPMEFIRQDTFQDQVSFSYAHM-KNQWN
Desert_Locust_Fng                 GYIIEHMLQKPLTVIEQFHSHLEPMKFLRPDTIQDQITFSYSHYSKDEMN
Bobtail_Squid_Fng                 GYIIGHILKKQLTIVDSFHSHLEALWLLRPFDLDRQVTFSYSKY-GEKMN
Sea_Squirt_Fng                    GFIITNRLKIDLTLTDLLHSTRESLRDLNPGTLHEQVSIGQGPGNTVNLD
Zebrafish_Lfng                    GYIIESVLGVSLTRSSLFHSHLENLQQVSKSEVHKQITLSYGMF-ENKRN
Zebrafish_Mfng                    GFIVERRLGISMIHSNMFHSHLENLLLLSPSDIPKQVTLSYGWF-ESKMN
Zebrafish_Rfng                    GYIIEALLEVPLTHTGLFHSHLENLQRLPAENILRQVTLSYGGF-ENRRN
Eastern_Newt_Rfng                 GYIIEGLLGVKMHHTPLFHSHLENLQRLPLQSVLKQVTLSYGGP-DNKRN
African_Clawed_Frog_Lfng          GYIIESVLGVKLIRSNLFHSHLENLHQVPQSEIHNQVTLSYGMF-ENKRN
African_Clawed_Frog_Rfng          GYIIEGMLDVKMQHSNLFHSHLEHLQRLPTESLLKQVTLSYGGP-DNKWN
Western_Clawed_Frog_Lfng          GYIIESVLGVKLIRSNLFHSHLENLHQVPQSEIHNQVTLSYGMF-ENKRN
Chinese_Softshell_Turtle_Lfng     GYIIESVLGVKLIRSNLFHSHLENLHQVPKTEIHKQVTLSYGMF-ENKRN
Chinese_Softshell_Turtle_Mfng     GYIIECKLGGRLLPNALFHSHLENLQLISRPQLPEQVTLSYGVF-EKKLN
Chinese_Softshell_Turtle_Rfng     GYIIEGLLEVKMLHSSLFHSHLENLQRLRGELVLQQVTLSYGGP-ENKHN
                                  *:*:   :   :    :**  * :  :    .  *:
```

Figure 13. ClustalW Sequence Alignment of All of the Fringe and Fringe-like Enzymes

The first and last residue of the loops that were not observed in the Mfng structure (Jinek, *et al.*, 2006) are indicated by diamonds above the sequence. Sequence conservation is indicated below the sequence. An asterisk indicates identical residues, two dots strongly type-conserved, and a single dot weakly type-conserved residues. The numbers above the alignment indicate the mouse Lfng sequence position.

level of conservation. Figure 14 shows the sequence conservation mapped onto the Lfng model surface in the vicinity of the active site.

Our mutagenesis strategy focused on four areas. First, we introduced a mutation (F187L) associated with the human genetic disorder SCD (Sparrow, *et al.*, 2006). Secondly, we focused on conserved residues in the two disordered loops flanking the active site for which there was no electron density in the Mfng structure (Jinek, *et al.*, 2006). Thirdly, residues in the vicinity of the two clusters of Fuc on either side of the putative catalytic aspartate 289 were mutated. Finally, we mutated residues in the vicinity of the UDP-GlcNAc binding site. We chose residues which might confer specificity for this bulkier nucleotide sugar donor compared to UDP-Glc, the substrate for the thrombospondin modifying β3GlcT (Kozma, *et al.*, 2006, Lee, *et al.*, 2006). All together we made 35 mutants encompassing 22 separate residues (Table 3 and Figures 18 and 22). Five of these (R173A, D289N, D289S, C290A, and S312Q) failed to produce assayable protein, three of which could not be visualized in a western blot of cell lysates or media (Figure 15).

The Spondylocostal Dysostosis Mutant

F187L Lfng secretion into the media was sporadic and hard to reproduce. We were able to observe protein expression approximately 25% of the time. In approximately half of these examples, expression was limited to the lysate. When protein is secreted to the media (Figure 15) we can measure significant activity rivaling the activity of wild-type enzyme in media. The F187L mutant is equivalent to the F188L mutant in human Lfng, which has been implicated as a genetic cause for SCD. Previously, Sparrow *et al.* (Sparrow, *et al.*, 2006) reported that this mutant was catalytically inactive. The authors attempted to assay protein from cell lysates which may or may not have been mature. We have found that protein that is presumably folded, mature, and secreted into the media is indeed catalytically active. Our results and those of Sparrow and coworkers would indicate that there is a protein expression, folding, trafficking or localization defect with this mutant. The enzymatic activity of folded enzyme does not appear to be affected by the mutation however.

Catalytically Inactive Mutants

Of the mutants that produced protein which could be purified and assayed, eight (L176A, S228L, S228Y, F251S, T253A, D288A, D289E, and S312T) are essentially catalytically inactive with the UDP-GlcNAc donor, as measured in our assay (Table 3). The S312T mutant is unique among this group in that despite a lack of Fng β1,3GlcNAc transferase enzyme activity, this mutant retains some residual activity with a UDP-glucose (UDP-Glc) donor which will be discussed later. Interestingly, the inactive mutants include the D288A variant, next to the putative catalytic aspartate 289. Although some residual catalytic activity remains, it is measurable only at saturating acceptor substrate concentrations, and is barely above background. The Mfng structure (Jinek, *et al.*, 2006) and Lfng model show that Aspartate 288 forms a hydrogen bond with serine 177 at the base of the short loop (Figure 16). If the short loop becomes ordered during or subsequent to substrate binding, the position of serine 177 may be altered, which could in turn affect aspartate 288 and the back-bone conformation of the putative catalytic aspartate 289. Thus, ordering of the short loop may move Asp289 into a catalytically

Figure 14

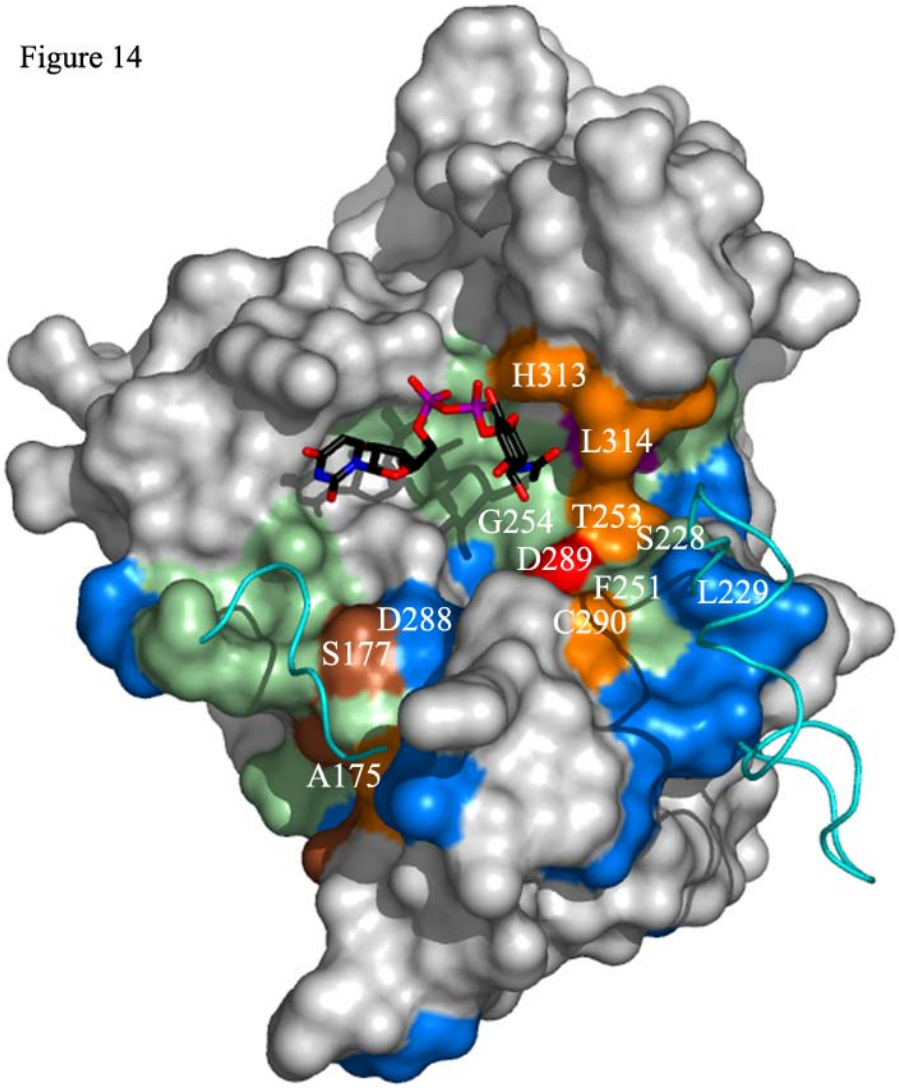


Figure 14. Sequence Conservation in Lunatic Fringe
Sequence conservation of Fng enzymes has been mapped on the Lfng surface (white) as follows: Identical residues are colored green, strongly conserved residues are colored blue, and weakly conserved residues are colored brown based on the alignment in Figure 13. The orange residues are not identically conserved in the sequence alignment due to cysteine 290 being valine in insects, threonine 253 being histidine in the sea-squirt histidine 313 being threonine in sea-squirt and leucine 314 being arginine in sea-squirt. Otherwise they would be completely conserved. The strictly conserved serine 312 is colored purple to help identify its position occluded by leucine 314 in the picture. The UDP-GlcNAc is displayed in black stick representation with all-bonds coloring. The catalytic aspartate is colored red. The modeled backbone of the short and long loops is shown in cyan ribbon.

Figure 15

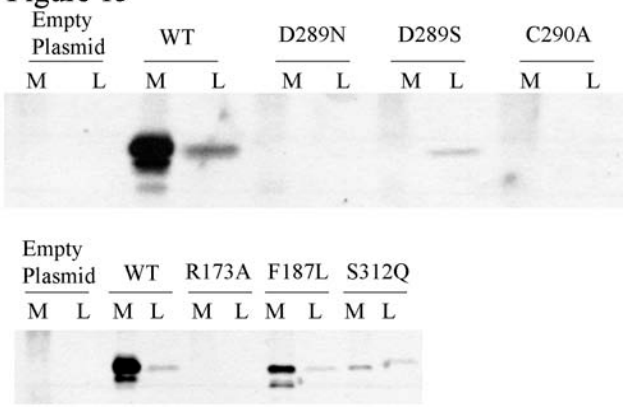


Figure 15. Western-Immuno Blots of Non-Expressing and Purification Recalcitrant Lunatic Fringe Mutants

M denotes samples from the media and L those from the lysate.

Figure 16

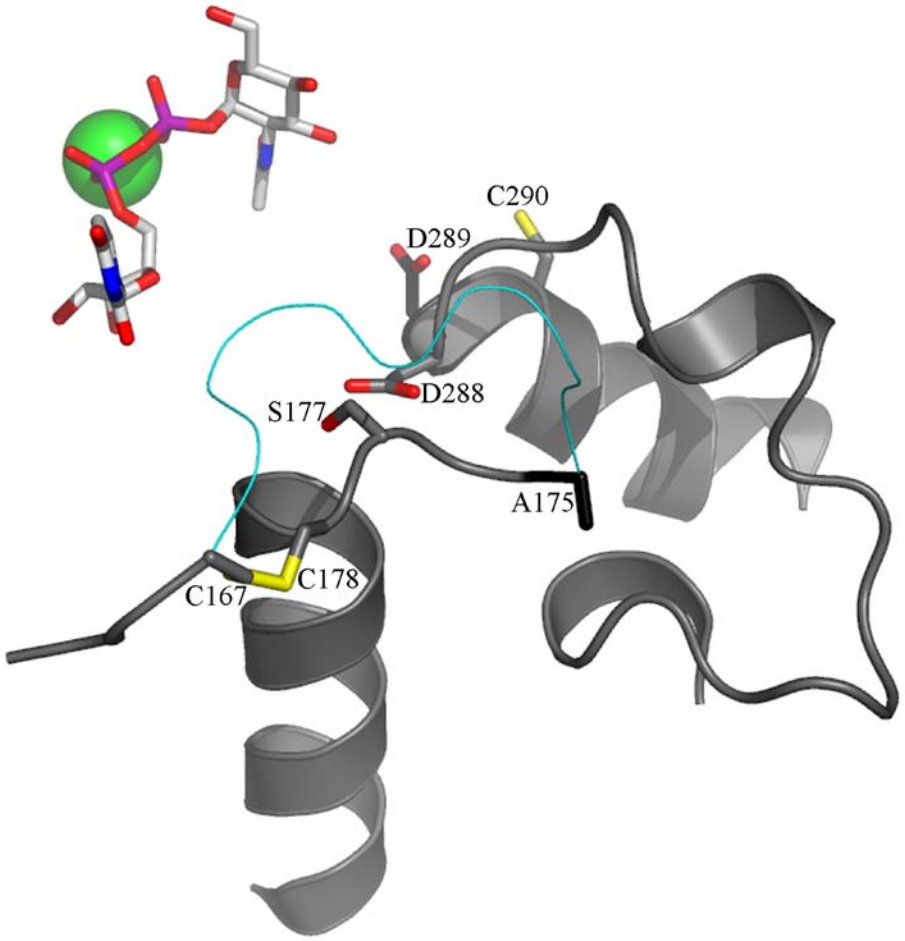


Figure 16. Model of the Short Loop and the Serine 177-Aspartate 288 Hydrogen Bond

Lfng is represented by a grey cartoon with C167, S177, C178, D288, and C290 in grey sticks with all-bonds coloring. A175 and D289 are shown in black sticks with all-bonds coloring. UDP-GlcNAc is shown with white sticks and all bonds coloring. The Mn^{2+} ion is shown as a green sphere. The modeled backbone of the short loop is shown in cyan ribbon.

optimal configuration (Figure 16). We were unable to express either the D289S or D289N mutants, despite robust expression and secretion of mutants affecting the adjacent residues on either side of this aspartate, as illustrated by the D288A and C290S mutants (Table 3). The D289E variant was the only mutant of the catalytic aspartate that produced secreted protein which we could purify. Assays demonstrate that this mutant is catalytically inactive (Table 3).

Effects of Mutations in the Loops and Residues Flanking Aspartate 289

Jinek and coworkers referred to the region between the long loop and the catalytic site as a putative fucose binding pocket (Jinek, *et al.*, 2006). This is approximately the area encompassed by T253, S228, and F251 in Figure 14. Although disordered, the long loop has to be adjacent to one of the Fuc clusters we observed in our docked solutions for the EGF-*O*-fucose (Figure 12D). When we tabulate V_{max} and K_M values for our loop mutants (Table 3) we find a striking pattern where most of the K_M defects localize to mutants in or near the long loop and the putative Fuc binding pocket, while the more significant V_{max} defects localize to the short loop region or near the alternative Fuc cluster we observed in our docking experiment (Figure 17). This suggests that the Fuc is indeed localized near the putative Fuc binding pocket on the long loop side of the active site, and not on the short loop side. The V_{max} defects indicate that the short loop closes one side of the active site pocket and does not make significant contributions to the affinity of the acceptor for the active site. These short loop mutants do not show substantial K_M defects for UDP-GlcNAc (Table 3) suggesting minimal contact is made with the donor substrate.

Perhaps the most striking defect occurred with the C290S mutant (Figure 18A). This free cysteine is a direct neighbor of the putative catalytic aspartate 289. Despite the conservative nature of the mutation, it was one of four mutants with a large K_M defect, while no V_{max} defect was present (Table 3). This strongly suggests that the Fuc binds near the long loop, and that the sulfur of C290 makes an important contact with the acceptor. Phenylalanine 251 when mutated to tyrosine (Figure 18B) exhibits a striking K_M defect while the serine mutant of this residue is catalytically inactive (Table 3). This phenylalanine forms part of the floor of the active site to the long-loop side of aspartate 289 and would appear from these results to provide needed hydrophobic bulk at the base of the long-loop (Figure 14). The K_M defect from the tyrosine mutation implies a steric clash with the pNP-fucose reinforcing the conclusion that the substrate is on this side of the catalytic pocket. Mutating the highly conserved residue serine 228 to leucine or tyrosine produced inactive enzyme (Table 3, Figure 18C,D). Making the less drastic mutant S228A produced a significant K_M defect (Table 3) suggesting that the hydroxyl of the serine makes contact with the pNP-fucose substrate, further reinforcing the probability that the acceptor is positioned in this area of the active site (Figure 14). The S228T mutant produced a mixed K_M and V_{max} defect (Table 3) possibly indicating that any additional bulk in this position is poorly tolerated. This is not surprising considering the broad conservation of serine at this position in Fng and Fng-like enzymes (Figure 13). On the long-loop side of aspartate 289, threonine 253 forms part of the floor of the active site, near the edge of the donor-substrate binding site (Figure 14). Mutation of threonine 253 to alanine (Figure 18E), a relatively conservative substitution, produces inactive enzyme (Table 3), suggesting that the bulk of the threonine is important for the

Figure 17

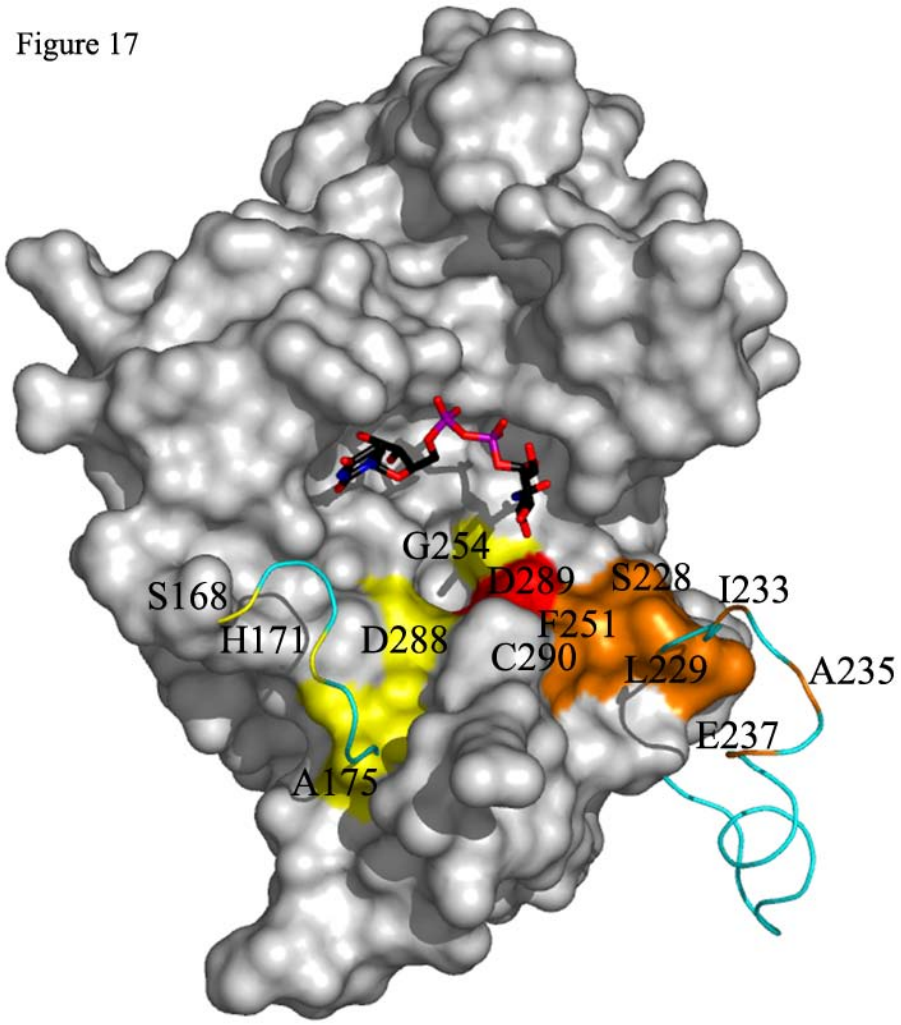


Figure 17. Clustering of V_{max} and K_M Defects to Either Side of the Catalytic Aspartate

Lfng is shown in white surface representation, UDP-GlcNAc in black sticks and all-bonds coloring, mutants with V_{max} defects in orange, and those with K_M defects in yellow and the catalytic aspartate in red. A model of the backbone of the two loops is shown in cyan ribbon.

Figure 18

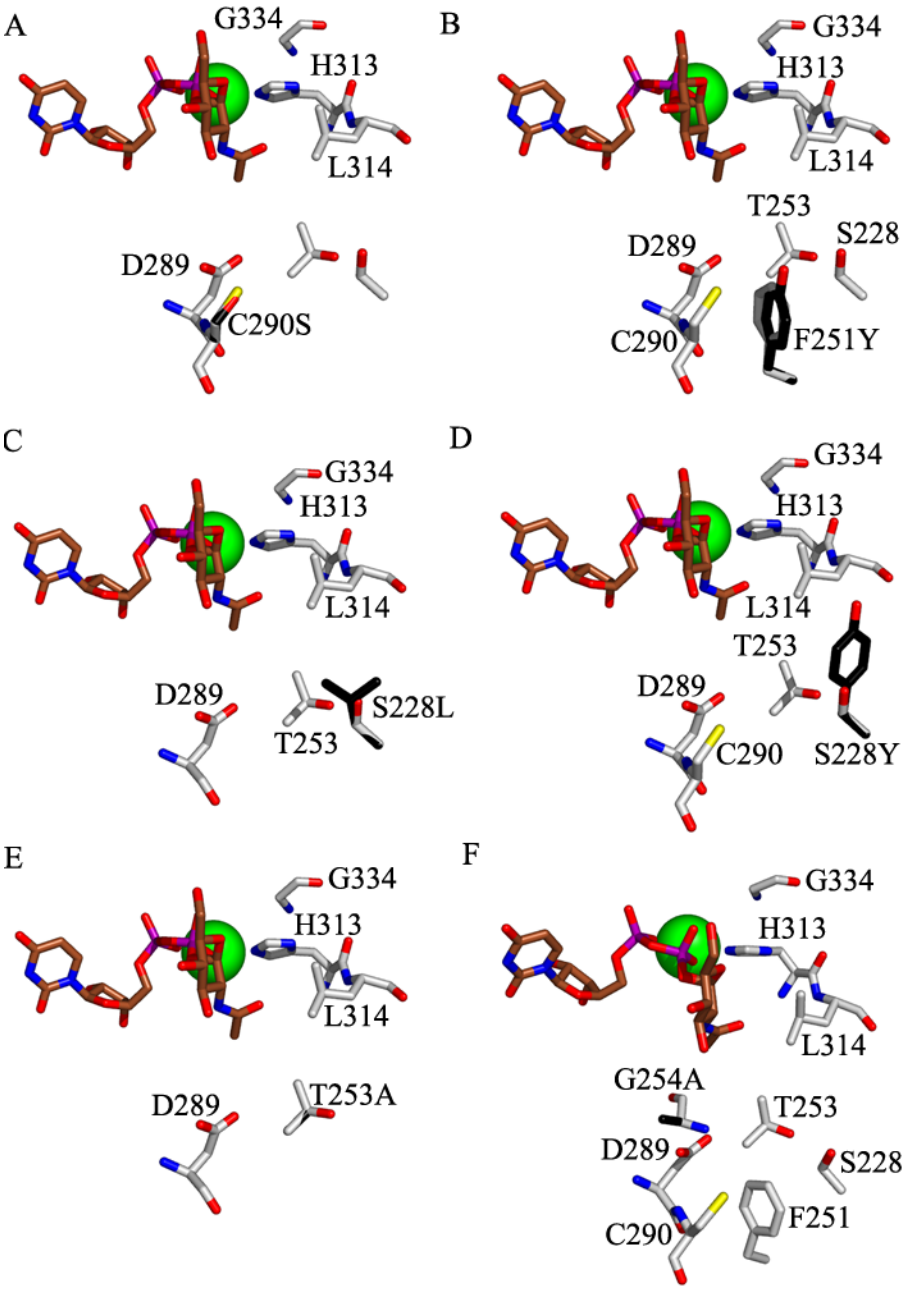


Figure 18. Lunatic Fringe Active Site Mutants
Amino acid side chains are displayed as white sticks with all-bonds coloring and mutant side-chains are displayed as black sticks with all-bonds coloring, UDP-GlcNAc is displayed as brown sticks with all-bonds coloring with the Mn^{2+} as a green sphere. **A)** C290S, **B)** F251Y, **C)** S228L, **D)** S228Y **E)** T253A and **F)** G254A.

conformation of the pocket since there are no obvious hydrogen bonds involving the hydroxyl group of the threonine. Proximity of this residue to leucine 314 and serine 312 suggests that the catalytic defect could be manifested in perturbation of the position of the GlcNAc (Figure 14, 18E). Glycine 254 is positioned near the mouth of a small cavity on the short-loop side of the catalytic aspartate 289 (Figure 14). The mutation of glycine 254 to alanine produces a V_{max} defect presumably through steric affects on the positioning of aspartate 289 (Table 3, Figure 18F).

The possibility existed that the mutants exhibiting V_{max} defects were in fact unstable, which would decrease the overall catalytic efficiency of a given aliquot of enzyme. We eliminated this possibility by repeating the UDP-hexanolamine agarose experiment with the V_{max} defective mutant enzymes, and observed behavior identical to wild-type enzyme (Figure 19). This indicates that these mutations are affecting catalysis and the V_{max} deficits are not due to misfolding, or catalytically inactive contaminants in the enzyme aliquots.

Effects of Mutants on the Utilization of the EGF-O-fucose Acceptor

As described above, we were unable to generate sufficient amounts of the EGF-*O*-fucose acceptor substrate to saturate the enzyme (Figure 10). Because of this, only the linear portion of the Michaelis-Menten curve was accessible in our assays. Due to the limited amount of this substrate available, we decided to measure the effect of the mutants on Lfng utilization of EGF-*O*-fucose by looking for a change in the slope of the linear portion of the Michaelis-Menten curve. Since the V_{max} and K_M defects clustered with the pNP-fucose acceptor substrate, we looked for evidence of clustering with the EGF-*O*-fucose as well. The K_M cluster of mutants (Figure 20A) show very little difference from wild-type Lfng in the linear portion of the Michaelis-Menten curve, whereas the V_{max} cluster of mutants show obvious changes in slope compared to wild-type Lfng (Figure 20B). While we can draw no kinetic or mechanistic conclusions from these data, the appearance that these two groups of mutations continue to produce separate and distinct effects with the EGF-*O*-fucose acceptor substrate reinforces our proposal that the effects we see with the pNP-fucose substrate are due to distinct mechanistic properties of the enzyme, and that we are drawing reasonable conclusions from our kinetic data with the small molecule acceptor substrate.

Specificity for UDP-N-acetylglucosamine versus UDP-Glucose

Recently a Fng homologue was identified that transfers glucose from UDP-glucose to *O*-fucose on TSRs forming a Glc-β1,3-Fuc-α-*O*-Ser/Thr linkage (Kozma, *et al.*, 2006, Lee, *et al.*, 2006). While the degree of sequence homology between Mfng and the β3GlcT is somewhat low (20.7% identity, see Figure 21), we were able to create a homology model of the β3GlcT threaded onto the Mfng structure (pdb 2J0B) using the ESyPred3D server (Lambert, *et al.*, 2002). We compared the models of Lfng and the β3GlcT to help ascertain which residues might contribute to the ability of these enzymes to discriminate between UDP-GlcNAc and UDP-Glc. Comparisons of these two structures are shown in Figure 22. Figure 22A shows four side chains from the β3GlcT homology model that protrude from the Lfng surface, and may affect the shape of the UDP-sugar binding pocket. Figure 22B shows these side-chains from Lfng and β3GlcT overlapped, while Figure 22C shows the wild-type residues and 8D those residues

Figure 19

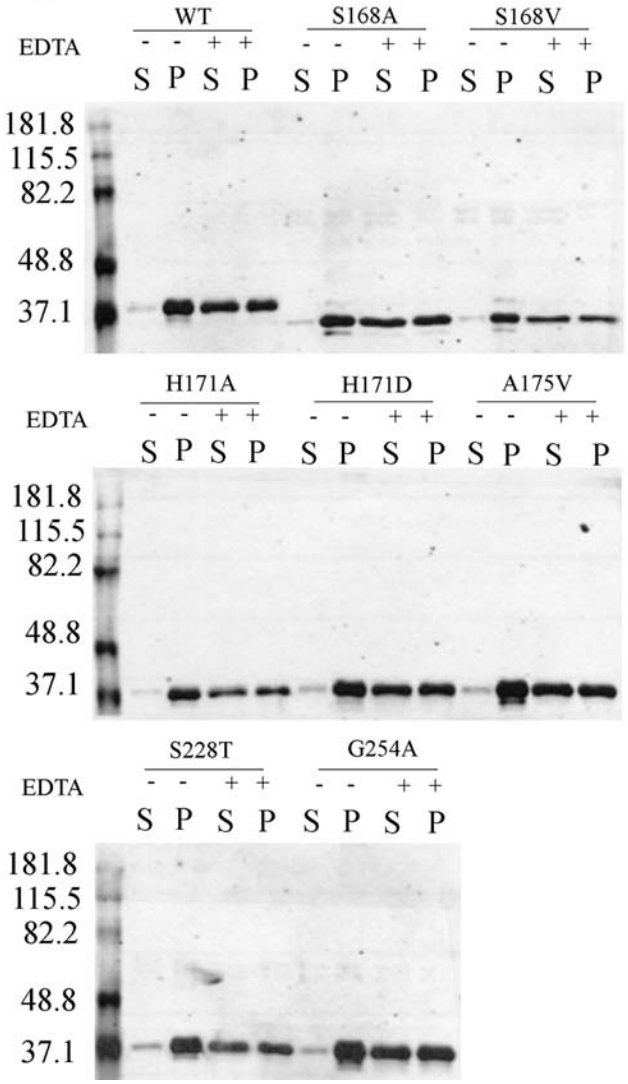


Figure 19. UDP-Hexanolamine Incubations for Wild-type and V_{max} Mutants of Lunatic Fringe

The incubations were performed in the presence or absence of 100 mM EDTA which will inhibit the ability of the UDP-hexanolamine to bind to the enzyme. All assays were performed at the highest concentration of UDP-hexanolamine agarose in Figure 11. All mutants show identical behavior with the UDP-hexanolamine agarose when compared to the WT control indicating that the V_{max} mutants are not exhibiting V_{max} defects due to a misfolded or unstable enzyme.

Figure 20

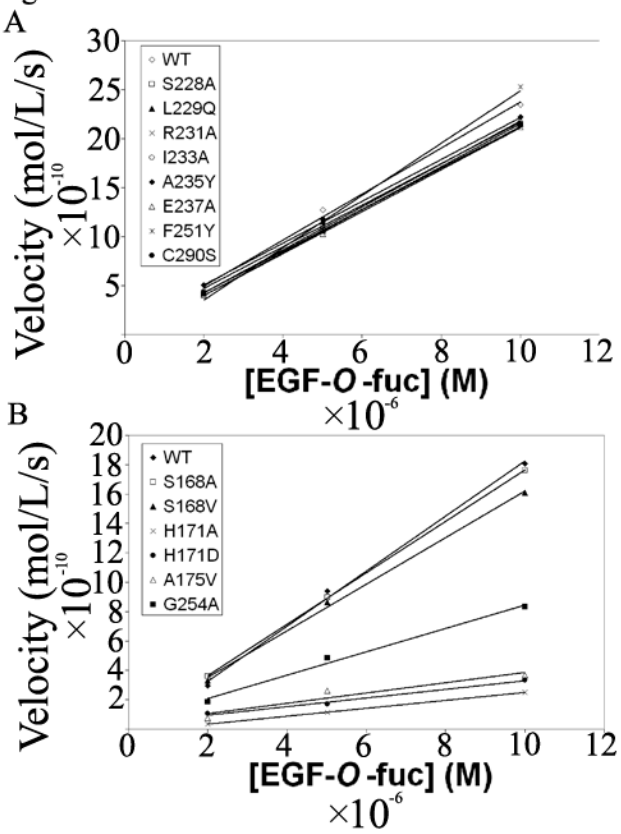


Figure 20. **Linear Portion of the Michaelis-Menten Curve for Lunatic Fringe Mutants With the EGF-*O*-fucose Acceptor Substrate**

A) The K_M cluster of mutants show little difference from wild-type Lfng with the EGF-*O*-fucose substrate. **B)** The V_{max} cluster of mutants show obvious changes in slope when compared with wild-type Lfng.

Figure 21

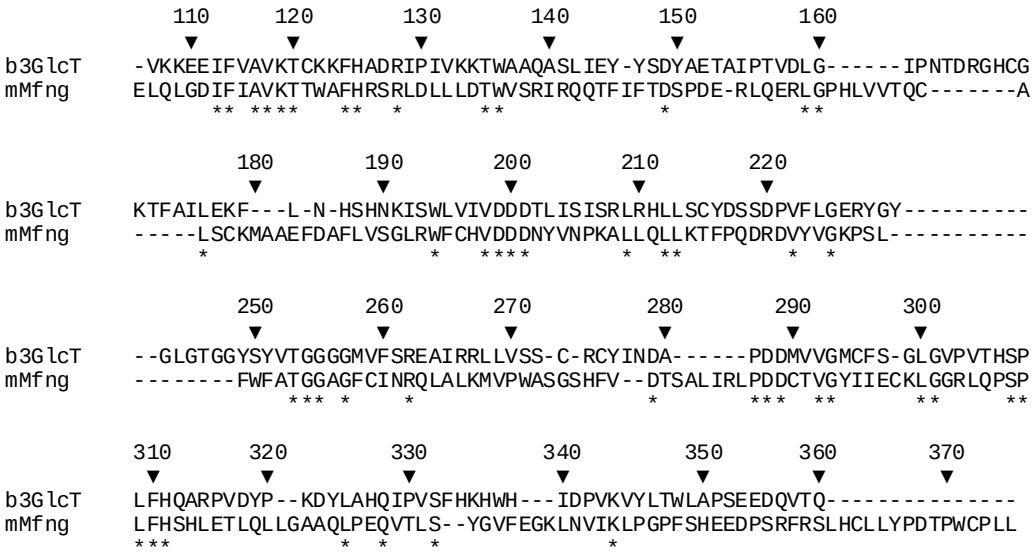


Figure 21. Alignment of β1,3Glucosyltransferase and Manic Fringe Returned by the ESyPred3D Server

Identical residues are marked below the sequence with an asterisk. The corresponding mouse Lfng numbering from Figures 5 and 13 is shown above the alignment.

Figure 22

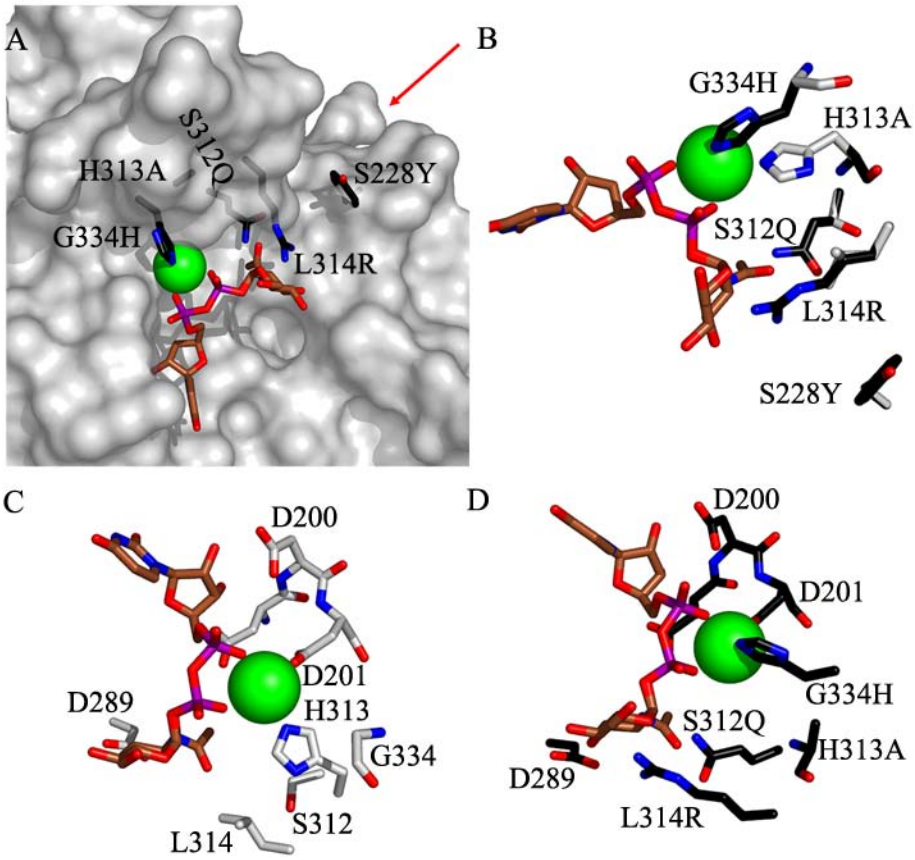


Figure 22. Lunatic Fringe Donor Specificity Mutants
A) Lfng is displayed as white surface, β3GlcT residues from the ESyPred3D model are displayed as black sticks with all-bonds coloring, UDP-GlcNAc is displayed as brown sticks with all-bonds coloring, and the Mn^{2+} is displayed as a green sphere. This panel shows G334H, S312Q, L314R and S228Y. H313A is buried under the Lfng surface. The red arrow shows the approximate viewpoint of the following panels. **B)** Nucleotide-sugar binding pocket showing all mutants. Lfng is displayed as white sticks with all-bonds coloring. Mutated residues are displayed as black sticks with all-bonds coloring. Shown here are S228Y, S312Q, H313A, L314R, and G334H. **C)** Lfng is displayed as white sticks with all bonds coloring. This panel shows the wild-type Lfng active site pocket. **D)** The active site pocket from the β3GlcT model displayed as black sticks with all-bonds coloring. Shown are the DDD motif, the catalytic aspartate D289, S312Q, H313A, L314R, and G334H. The numbering of residues refers to the corresponding mouse Lfng residues.

mutated to the β3GlcT sequence. Both the Lfng and β3GlcT models show a histidine chelating the Mn^{2+} ion. In Lfng this is histidine 313 (alanine in ß3GlcT) while in the β3GlcT it is histidine 334 (glycine in Lfng) (Figure 22 B, C, D). Additionally, the side chain of leucine 314 in Lfng corresponds to an arginine in β3GlcT (Figure 22B), while the highly conserved serine 312 in Lfng is in the same position as a glutamine in β3GlcT (Figure 22B, C, D).

To examine whether any of these residues affected the ability of Lfng to utilize UDP-glucose versus UDP-GlcNAc, we generated a series of mutants converting the corresponding residues to those of the β3GlcT. Wild-type Lfng transfers glucose from UDP-glucose to pNP-fucose at approximately ten-fold reduced catalytic efficiency when compared to transfer of GlcNAc from UDP-GlcNAc. Thus, we measured mutant catalytic efficiency toward each donor separately as a percentage of the wild type activity with that donor. We took the activity for the mutants utilizing UDP-Glc, and divided by the activity of wild-type Lfng with UDP-Glc to compute the percentage of wild-type activity retained. We then repeated this process with UDP-GlcNAc as the donor. The H313A mutant is all but dead when analyzed for UDP-GlcNAc transfer, whereas it retains 75% of the wild-type Glc transfer activity (Table 3, Figure 23). Thus, the loss of this histidine is considerably more permissive toward UDP-Glc utilization in a productive complex, and seems to prevent UDP-GlcNAc utilization. Replacement of G334 with histidine (Figure 22A-D) does not seem to have a dramatic effect on catalytic efficiency with either donor (Figure 23). However, this highly non-conservative mutation is strikingly permissive to the utilization of either of the donor nucleotide sugars tested, with roughly 78% of activity retained with UDP-Glc and 94% of activity retained with UDP-GlcNAc (Figure 23). Not surprisingly, the double mutant of H313A/G334H is all but dead when utilizing UDP-GlcNAc, which is likely due mainly to the H313A mutation, while there is a significant (34%) retention of UDP-Glc utilization (Table 3, Figure 22D). Leucine 314 is pointing toward the GlcNAc of the donor substrate in Lfng (Figure 22A-C) and mutation of this residue to arginine disrupts GlcNAc transfer considerably, but is highly permissive toward Glc transfer, exhibiting an even higher than wild-type catalytic efficiency when utilizing the UDP-Glc donor (Figure 23). Jinek and coworkers mutated this leucine in Mfng to phenylalanine and saw no effect on their *in vivo* ligand binding assay (Jinek, *et al.*, 2006). This suggests that although L314 is a conserved residue in the enzyme, a significant alteration and introduction of side chain bulk by phenylalanine can be tolerated well, in strong contrast to the considerable effect we see when L314 is replaced with a positively charged arginine side-chain. We also noticed that while S312T was enzymatically dead when utilizing the UDP-GlcNAc donor, it retained 18% of the wild-type Glc transfer activity (Table 3, Figure 23). Correia *et al*. reported a lethal phenotype for the equivalent serine to threonine mutation in a genetic screen of *Drosophila* Fng (Correia, *et al.*, 2003). The S312Q mutant (Figure 22A,B, and D) seemed a promising candidate to exclude UDP-GlcNAc and permit UDP-Glc utilization but it failed to express well or purify. It is curious, however, that the additional methyl group of a threonine at this highly conserved position would be considerably less disruptive to UDP-Glc utilization by the enzyme, compared with UDP-GlcNAc suggesting that this residue may indeed be part of a selective mechanism for the donor.

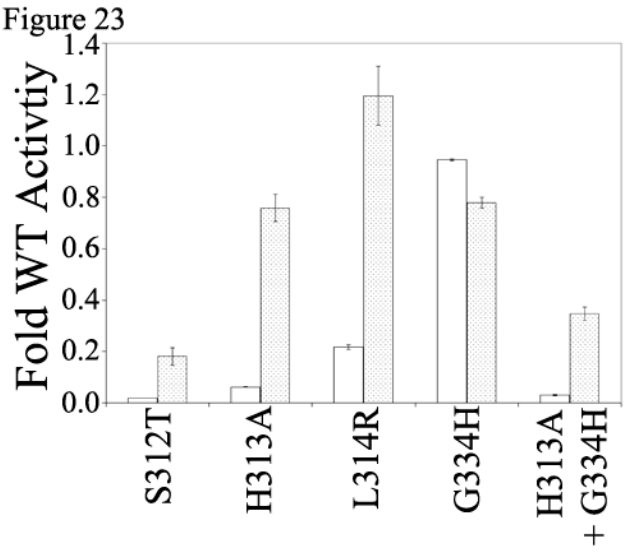


Figure 23. Relative Activity of Lunatic Fringe Mutants With UDP-N-acetylglucosamine and UDP-Glucose Donor Substrates

The percentage of wild-type Lfng activity with each donor is shown for several mutants. The white bars indicate UDP-GlcNAc utilization and the shaded bars indicate UDP-Glc utilization. The error bars show the standard deviation of the measurements. These data are tabulated in Table 3 under the heading Donor Specificity.

Acceptor Substrate Specificity

The use of an easy-to-obtain small molecule acceptor substrate like pNP-fucose allowed us to analyze a large number of mutant forms of the enzyme. Additionally, since the Fuc is a significant portion of the small molecule acceptor, we might observe K_M defects associated with the Fuc that could be obscured by the higher affinity of the EGF-repeat protein mass for Lfng. We have taken advantage of the availability of a number of readily available *p*-nitrophenyl-sugar compounds to investigate the specificity of Lfng for the Fuc portion of the acceptor. Our assays of Lfng activity with these alternative acceptor substrates were carried out at the 100 mM saturating concentration we determined for pNP-fucose. The most obvious source of specificity for Fuc would be the methyl substituent at C6 which in all other known sugars in animals is hydroxylated. We assayed the activity of Lfng with pNP-β-L-arabinopyranoside, which differs from pNP-α-L-fucose only in the complete absence of a C6 substituent (Figure 24B). The arabinopyranoside did not serve as an acceptor in our assay, and as such, the importance of a C6 substituent on the acceptor sugar is absolute (Figure 24J). The pNP-α-L-rhamnose acceptor contains a C6 methyl, and is identical in conformation at C1 and C3, differing only in the orientation of the C2 and C4 hydroxyls (Figure 24C). Despite our anticipation that this acceptor would most successfully mimic pNP-fucose, it showed only minimal activity at approximately 0.2% of pNP-α-L-fucose activity (Figure 24J). This suggested that the conformation at C2 and C4 influences specificity for Fuc. While L-fucose is related in structure to both galactose (6 deoxy-L-galactose) and mannose, we were surprised to see that both of these sugars served as better alternative substrates than rhamnose since galactose and mannose are both D-sugars (Figure 24J). If we consider β-D-mannose, which is the best mimic among the alternative substrates at approximately 4% of pNP-α-L-fucose activity, we see that the C2 substituent is flipped from equatorial in the Fuc to axial above the ring (Figure 24F). If we compare this to α-L-rhamnose, which is a twenty-fold worse substrate, we see that the major difference is that the rhamnose has an axial C2 substituent below the ring (Figure 24C). Thus it seems reasonable to conclude that an axial substituent below the ring further impedes catalysis with the rhamnose substrate compared to β-D-mannose. In L sugars the C6 substituent is below the ring plane, and equatorial in orientation. In Fuc, the O3 substituent is axial and below the ring. The catalytic defect from an axial O2 substituent below the ring may indicate that interactions with the methyl group orient one face of the Fuc toward aspartate 289 and that an axial substituent below the ring at O2 represents a steric hindrance to orienting the Fuc ring properly.

We notice that with the α-D-gal and both α and β-D-mannose substrates that rotating the ring 180° about the ring oxygen allows the sugars to superimpose much better (Figure 24G-I). In the case of β-D-mannose, which was the best mimic, all of the substituents now match. The corollary of this of course is that the *p*-nitrophenyl portion of the acceptor would be mimicking the methyl group present on Fuc and the C6 substituent of galactose or mannose would be in the place of the pNP portion of the pNP-fucose (Figure 24G-I). Perhaps this is the reason that even the best mimic shows only 4% wild-type activity (Figure 24J). This would further reinforce the conclusion that there is an absolute necessity for a C6 substituent, as even a bulky *p*-nitrophenyl would be a better substitute at this position than the complete absence of a C6 substituent as seen with pNP-β-L-arabinopyranoside (Figure 24J). Finally, we were surprised to find that the

Figure 24

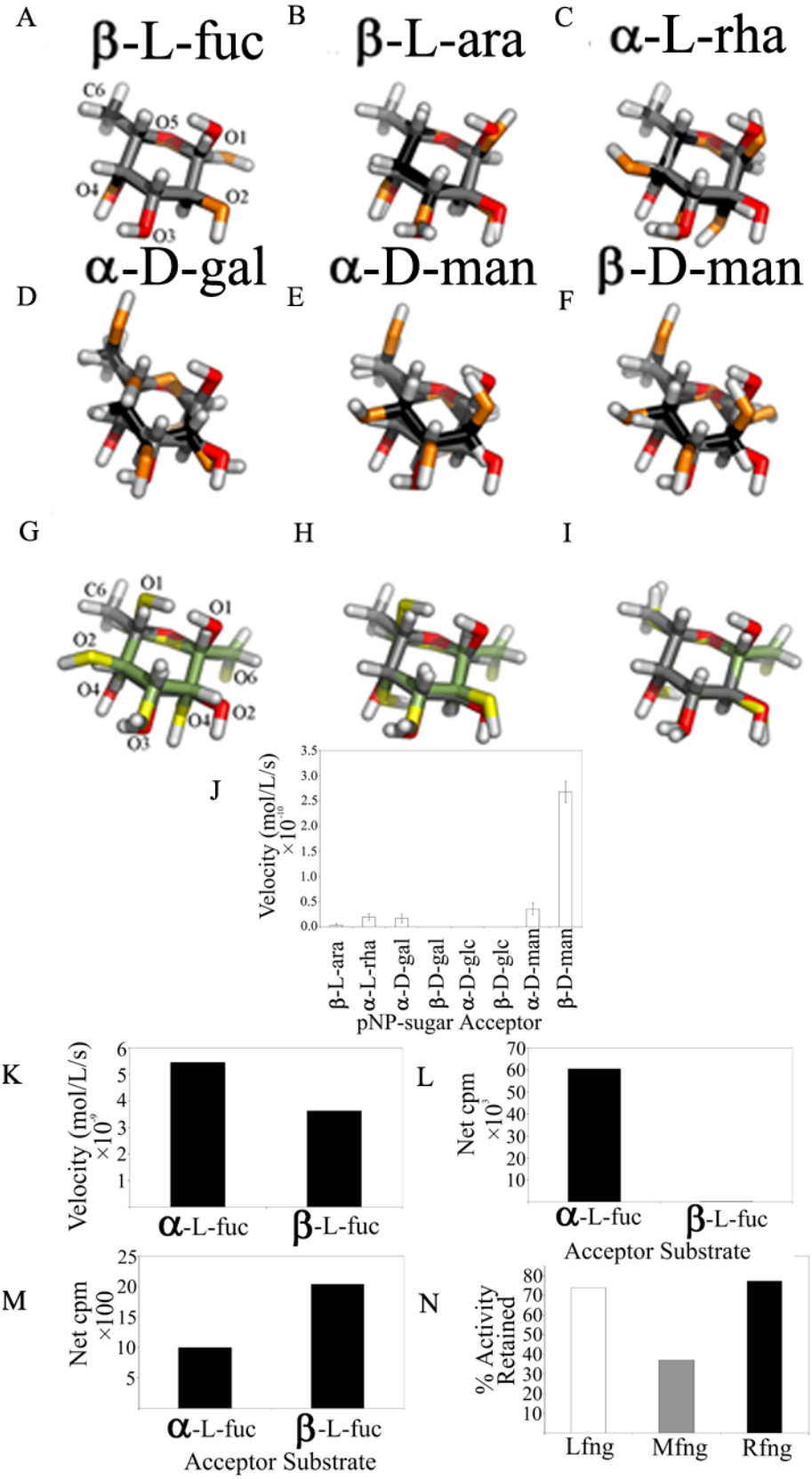

A
β-L-fuc
C6
O5
O1
O4
O2
O3
B
β-L-ara
C
α-L-rha
D
α-D-gal
E
α-D-man
F
β-D-man
G
C6
O1
O1
O2
O6
O4
O4
O2
O3
H
I
J
Velocity (mol/L/s)
×10⁻¹⁰
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
β-L-ara
α-L-rha
α-D-gal
β-D-gal
α-D-glc
β-D-glc
α-D-man
β-D-man
pNP-sugar Acceptor
K
Velocity (mol/L/s)
×10⁻⁹
6
5
4
3
2
1
α-L-fuc
β-L-fuc
L
Net cpm
×10³
70
60
50
40
30
20
10
α-L-fuc
β-L-fuc
Acceptor Substrate
M
Net cpm
×100
25
20
15
10
5
α-L-fuc
β-L-fuc
Acceptor Substrate
N
% Activity Retained
80
70
60
50
40
30
20
10
Lfng
Mfng
Rfng

Figure 24. Comparison of Alternative Sugar Acceptors With α-L-Fucose
In each panel all hydrogens are colored white, the α-L-fucose carbons are colored grey, the α-L-fucose oxygens are colored red. In panels A-F the carbons of the sugar being superimposed on α-L-fucose are colored black, and the oxygens of the sugar being superimposed are colored orange. In panels G-I the carbons of the sugar being superimposed on α-L-fucose are colored green, the oxygens of the sugar being superimposed are colored yellow. Panels G to I show the superimposition of sugars that have first been rotated 180° around the ring oxygen. The numbering of the constituents is shown in panels A and G. **A)** β-L-fucose, **B)** β-L-arabinose, **C)** α-L-rhamnose, **D)** α-D-galactose, **E)** α-D-mannose, **F)** β-D-mannose, **G)** α-D-galactose flipped, **H)** α-D-mannose flipped and **I)** β-D-mannose flipped. **J).** Comparison of Lfng activity with various non fucose pNP-sugar acceptor substrates. **K)** Comparison of Lfng activity with pNP-α-L-fucose and pNP-β-L-fucose. **L)** β3GlcT activity toward pNP-α-L-fucose and pNP-β-L-fucose. **M)** Lfng glucosyltransferase activity toward pNP-α-L-fucose and pNP-β-L-fucose. **N)** Percent of Fng activity toward pNP-α-L-fucose retained with the pNP-β-L-fucose acceptor for Lfng, Mfng and Rfng. Refer to Experimental Procedures for method of calculating percent activity.

pNP-β-L-fucose acceptor substrate retained 65% of the pNP-α-L-fucose activity (Figure 24K). We were concerned that this indicated pNP-fucose may be a poor mimic of EGF-*O*-fucose. However, if that were the case, then we would expect less drastic decreases in activity with the pNP-sugar acceptor substrates as previously mentioned. Additionally, β3GlcT shows an absolute preference for pNP-α-L-fucose and no activity with the beta form of the acceptor (Figure 24L) as we reported previously (Moloney and Haltiwanger, 1999). Strangely, when we tested the ability of Lfng to transfer Glc to the pNP-β-L-fucose, we found that it was twice as efficient as with the pNP-α-L-fucose (Figure 24M). The experiment in Figure 24N shows a comparison of GlcNAc-transferase activity toward the alpha and beta forms of the pNP-fucose acceptor for each of the three mammalian Fng enzymes. Mfng shows a significantly decreased ability to modify the beta form of the acceptor while Lfng and Rfng show indistinguishable behavior in this regard.

Lfng produced in Pro5 CHO cells will contain complex or hybrid type *N*-linked glycans containing Fuc, whereas Lec1 CHO cells produce only high-mannose type *N*-linked glycans without Fuc (Stanley and Chaney, 1985). If Lfng from Pro5 cells is incubated under assay conditions with only itself as acceptor substrate, a significant modification can be observed (data not shown, Raajit Rampal, Li Shao). Addition of EDTA to the assay eliminates the modification indicating it is a function of Lfng catalytic activity. Lfng from Lec1 CHO cells does not show the self-modification activity (data not shown, Raajit Rampal, Li Shao), presumably due to the lack of Fuc on the *N*-linked glycans produced in these cells (Stanley and Chaney, 1985). Together, these results are highly suggestive of promiscuous acceptor substrate specificity for these enzymes.

Discussion

Until now, no kinetic characterization of Fng enzymes has been published with saturating concentrations of acceptor substrate. More importantly, the ability to saturate has allowed us to utilize assays to assess the relative effect of various mutations in the Lfng active site. We conclude that the Fuc is situated on the long loop side of aspartate 289 near the putative Fuc binding pocket proposed by Jinek and coworkers (Jinek, *et al.*, 2006). The fact that mutations in the short loop result in V_{max} but not K_M defects led us to the proposal that the short loop becomes ordered coincident with, or subsequent to acceptor substrate binding. We propose that ordering of the short loop may affect a hydrogen bond between serine 177 and aspartate 288 and as such could affect the conformation of the catalytic aspartate 289. We have reported a surprisingly high, estimated K_M of 2 mM for the EGF-*O*-fucose substrate, and the appearance of a slow-on component to the enzyme mechanism at low enzyme/substrate concentrations. Furthermore, we conclude that several residues in the vicinity of the bound UDP-GlcNAc are specifically permissive toward utilization of this larger donor compared with UDP-Glc. We have also report that the F187L mutant of mouse Lfng analogous to the human F188L mutant associated with SCD is indeed catalytically active.

Our kinetic characterization of Lfng suggests that UDP likely forms dead-end complexes with the acceptor substrate. Nevertheless, this dead-end complex should have little effect on the *in vivo* activity of Lfng, since UDP is rapidly degraded to UMP in the Golgi (Novikoff and Goldfischer, 1961), and UMP has a much higher K_i for the enzyme.

We have observed an inability to saturate the enzyme with relatively high concentrations of an EGF-*O*-fucose acceptor substrate (Figure 10A). We estimate the K_M for this substrate to be surprisingly high at 2 mM. The Hanes-Woolf transformation of the data at high EGF-*O*-fucose concentrations shows the expected linear relationship with a positive slope. At low concentrations of EGF-*O*-fucose substrate and low concentrations of enzyme, the Hanes-Woolf transformation of the data is clearly curved rather than linear, with a negative slope (Figure 10B). These distortions from the expected linear relationship suggest a slow-on/slow-off component to the mechanism (Bieth, 1995). The slow-on component is second order, and as such, the linear relationship can be restored by increasing the concentration of the enzyme (Figure 10B). It is interesting to speculate that this sort of mechanism could play a role in the oscillatory nature of the segmentation clock (Cole, *et al.*, 2002, Dale, *et al.*, 2003, Forsberg, *et al.*, 1998, Jouve, *et al.*, 2002, McGrew, *et al.*, 1998, Morales, *et al.*, 2002). As the concentration of the Lfng enzyme increases over time it reaches a level where the kinetics of Lfng modification of the Notch EGF-repeats is linear and has increased efficiency compared to the case where the total concentration of enzyme and substrate are at lower levels. Once a surge of Notch signaling arises from this threshold being overcome, the system would fall back to a case where the concentration of Lfng is too low for the higher efficiency of modification until the enzyme levels again recover.

All our data (Table 3 and Figure 17) indicate that the Fuc cluster on the opposite side of the catalytic aspartate represents the correct site of interaction (Figure 12D). Furthermore the authors of the Mfng structure refer to this area as the putative Fuc binding pocket based on a comparison with human glucuronyltransferase I (Jinek, *et al.*, 2006, Pedersen, *et al.*, 2000). The dramatic K_M defect from mutating cysteine 290 to serine is highly suggestive that the sulfur of the cysteine makes contact with the Fuc and that the Fuc is on the long loop side of the active site. An alternative explanation would be that the cysteine makes a critical contact with the long loop as it becomes ordered, and as such, affects acceptor substrate binding indirectly, a possible, albeit not very likely explanation. The conservative nature of the mutation and the dramatic effect it has on K_M can most easily be explained by the decreased size of the oxygen of the serine relative to the sulfur of the cysteine. If this were affecting the ordering of the long loop, one might expect a far smaller K_M defect as additional contacts between the ordered loop and the enzyme substrate complex would contribute to the overall stability of the complex. Indeed, we see more modest K_M defects when we introduce non-conservative mutations for identically conserved residues in the long loop such as with I233A, A235Y, and E237A. It seems unlikely that a single contact, altered in such a conservative manner as the C290S mutation could so dramatically affect the large protein mass of the loop, whereas a dramatic effect from a single contact with the small molecule acceptor is more easily conceivable. At first glance the cluster of docking solutions on the short-loop side of the active site seemed reasonable, whereby an activated Fuc could flip around and collide with the UDP-GlcNAc near the anomeric carbon. However, access to the anomeric carbon from this side would appear to require the GlcNAc to rotate approximately 180°, which would be blocked by the significant protein mass composed of residues such as histidine 313, and leucine 314 (Figure 6, 14). Additionally, Jinek and coworkers(Jinek, *et al.*, 2006) modeled the UDP-GlcNAc into the active-site based on the

rabbit GlcNAc-transferase I structure (Unligil, *et al.*, 2000), with the GlcNAc folded over to expose the anomeric carbon toward the long-loop side of the enzyme.

It is possible that the N-acetyl group could be oriented up and out of the pocket, rather than in the position shown in Figure 14. There appears to be room for such an arrangement in our docked solutions (Figure 12C,D). While there is no reason this could not be the case, it seems more likely that the N-acetyl group would be probed by the binding pocket to discern the identity of the donor. With the N-acetyl moiety of the GlcNAc oriented toward Leu314, we can see an explanation for the severely decreased utilization of UDP-GlcNAc by the L314R mutant (Figure 22A,B, and D), while UDP-Glc utilization remains at or above wild-type levels (Figure 23). Perhaps following activation there is a rearrangement allowing the activated Fuc to access the anomeric carbon of the GlcNAc sugar. A requirement for this sort of rearrangement may contribute to a slow-on/slow-off mechanism suggested by our kinetic data with the EGF-*O*-fucose acceptor substrate. Since both the slow-on and slow-off component of such a mechanism is affected by substrate concentration, it is interesting to speculate that the enzyme may favor the locally higher substrate concentration created by strings of EGF-repeats which is the situation with Notch substrates.

Both clusters of docked EGF-*O*-fucose solutions show that a significant area of the short loop interacts with the EGF-*O*-fucose (Figure 12A). However, we have been unable to measure any significant K_M defect from mutations of residues in this loop in assays using the pNP-fucose acceptor. This suggests that the short loop does not interact with pNP-fucose. We considered the possibility that this loop interacts predominantly with the UDP-GlcNAc; however, we observed little to no effect on the K_M for the donor with the short loop mutants (Table 3). This suggests there is at best, minimal contact with the nucleotide portion of the donor.

The V_{max} defects associated with these short loop mutants could be attributed to an altered local environment in the catalytic pocket if the loop is unable to completely close off one side of the active site from the solvent. However, there seems to be little access to the active site from this side in our docked structures (Figure 12A). Such dramatic catalytic effects from the short loop mutants in this case are not so easy to rationalize. Another possibility is that the closing of the loop upon substrate binding alters the positioning of the catalytic aspartate. The A175V mutant is at the base of the loop closest to the catalytic aspartate. This residue, along with L176, appears to anchor one side of the loop, at a point that is behind the catalytic aspartate, and movement of the anchor point, or disruption of the anchoring residues by mutation could propagate to the backbone of the catalytic residue through the hydrogen bond between serine 177 and aspartate 188 (Figure 16). While the Richardson lab has found that altering the backbone for most residues may change the frequency of the preferred side-chain rotamer, but not its position (Lovell, *et al.*, 2000), the situation is different for aspartate and asparagine, where altering the backbone also changes the position (Lovell, *et al.*, 2000). The possibility exists that closing the short loop could affect the positioning of the side-chain of the catalytic residue in the pocket. If these V_{max} defects were due merely to a backbone alteration due to mutation of a conserved residue, we might expect that mutations in the loop itself, and mutation of the termini farthest from the catalytic residue would be less effective at altering catalytic efficiency compared to the mutants at the termini closest to and behind aspartate 289 (Figure 16). However, when we compare the

effects of mutants at each terminus, and in the middle of the loop, we see similar effects (Table 3). The simplest explanation of these results is that disrupting the folding of this loop affects catalytic efficiency. The mutation of aspartate 288 to alanine produces enzyme that is all but completely inactive. While some residual activity with saturating pNP-fucose concentrations can be measured, it is barely above background, clearly indicating that this mutant produces a most dramatic V_{max} deficit. It is hard to imagine that an alanine at this position could produce such a dramatic affect without larger structural implications. This suggests at the very least that the hydrogen bond between aspartate 288 and serine 177 is critical for enzymatic activity. Aspartate 288 is at the beginning of an α-helix, and as such, it is reasonable to postulate from the effects of the D288A mutation that the active site terminus of this helix is less stable in the absence of the D288-S177 hydrogen bond. As such, we can conclude that perturbation of this hydrogen bond by an altered short loop conformation could propagate a conformational change to the neighboring aspartate 289 residue. The question remains whether the presence of this hydrogen bond is critical for the proper orientation of aspartate 289 in a static fashion, or if a conformational change occurs. Our data showing catalytic defects from mutation of short loop residues is suggestive of the latter interpretation. The fact that the mutations associated with the V_{max} and K_M clusters continue to show differences in behavior with the EGF-*O*-fucose substrate (Figure 20), despite our inability to saturate in these assays serves to reinforce our conclusions from the pNP-fucose data. Our inability to saturate with EGF-*O*-fucose when performing assays with this acceptor substrate prevents us from concluding whether the change in slope associated with the V_{max} cluster of mutants is due to a catalytic defect or an altered affinity for the EGF-*O*-fucose. These data show that we are able to observe affects due to the fucose with the pNP-fucose substrate that would not be evident with the larger EGF-*O*-fucose substrate. I also propose that the V_{max} defects evident for small loop mutants with the pNP-fucose substrate are mechanistic details that might be masked by larger affinity components when using the EGF-*O*-fucose substrate. Thus, our use of the small pNP-fucose acceptor substrate may be a useful probe of structural and mechanistic details that might otherwise be inaccessible with other substrates. Interestingly, Correia and coworkers in a genetic screen of Fng in *Drosophila* found a lethal phenotype when the residue equivalent to leucine 176 at the base of one of the short-loop termini was mutated to phenylalanine (Correia, *et al.*, 2003) suggesting that disruption of this loop can have similar dramatic effects *in vivo*. An aspartate or glutamate at the neighboring 288 position is conserved across a large number of glycosyltransferases (Correia, *et al.*, 2003). The acidic residue at this position is involved in hydrogen bonds in at least two other enzyme structures (Unligil, *et al.*, 2000, Yuan, *et al.*, 2007), and mutation of the corresponding residue in peptidoglycan glycosyltransferase (Yuan, *et al.*, 2007) results in severely defective enzyme. This mirrors our result with this residue. Another question that occurs is whether cysteine at position 177 would function similarly to serine since insect Fng enzymes have cysteine in this position, as do all of the Rfng enzymes (Figure 13).

Mutations of residues surrounding the donor substrate as suggested by the β3GlcT homology model were surprisingly effective at curtailing or eliminating utilization of UDP-GlcNAc as a donor while showing markedly less, or no effect on the ability of the enzyme to utilize UDP-Glc. Mutation of histidine 313 to alanine all but abolished activity with UDP-GlcNAc, while retaining 75% of wild-type UDP-Glc associated

activity (Table 3, Figure 23) suggesting that this residue is necessary to position the UDP-GlcNAc donor in a catalytically productive orientation (Figure 22A,B, and D). Mutation of glycine 334 to histidine, has little effect on activity with either donor (Table 3, Figure 23) and is surprisingly permissive toward both donors for such a non-conservative mutation. The fact that this glycine to histidine mutation can be so easily tolerated suggests the possibility that positioning of the Mn^{2+} ion, and thus the positioning of the phosphates of the nucleotide-sugar donor does not affect donor specificity dramatically. This suggests that the bulk of the interactions that affect donor specificity involve the sugar portion of the nucleotide and not the manner in which the sugar is displayed to the enzyme by the bound nucleotide moiety. Leucine 314 appears to be permissive toward UDP-GlcNAc utilization, since arginine in this position severely curtails utilization of the larger donor. Activity with the smaller UDP-Glc donor is higher than wild-type, while utilization of the UDP-GlcNAc donor drops more than five-fold with the L314R mutation (Table 3, Figure 22A,B,D). Interestingly, this leucine is conserved in all of the sequences we included in the alignment, save for the Fng-like protein from the sea squirt *Ciona intestinalis*, which has an arginine at this position (Figure 13). This sea squirt protein also contains a threonine at the position of histidine 313 (Figure 13), which suggests, considering what we have learned about histidine 313 and leucine 314 in Lfng that this Fng-like protein is not a Fng enzyme. Finally, S312T, while catalytically dead toward utilization of UDP-GlcNAc (Table 3, Figure 23), retains some activity (18%) with the UDP-Glc donor-substrate. Correia and coworkers reported a lethal phenotype for the corresponding mutant in *Drosophila* Fng (Correia, *et al.*, 2003). Since elongation past the disaccharide does not appear to occur in *Drosophila* cells (Xu, *et al.*, 2007), one wonders if overexpression of this particular mutant could at least partially rescue the lethal phenotype caused by the same mutation through addition of Glc rather than GlcNAc? This would probably entail overexpressing a construct with an ER localization sequence at the C-terminus to provide the enzyme with the necessary UDP-Glc donor-substrate.

The ability of Lfng to modify the beta form of a Fuc acceptor substrate was unexpected. Comparison of GlcNAc-transferase and glucosyltransferase activity of Lfng toward the beta form of the Fuc acceptor shows that Lfng actually prefers the beta form when transferring Glc. While there is no reason to believe that any of these activities are present *in vivo*, it is suggestive of promiscuous acceptor substrate specificity. The possibility that this is due merely to the use of a small molecule acceptor can be reasonably ruled out by the absence of any glucosyltransferase activity toward the beta form of pNP-fucose by the β3GlcT enzyme. Arguments that the β3GlcT enzyme may simply be too unrelated to Lfng to provide an appropriate comparison are hard to sustain in that the β3GlcT also modifies an *O*-fucosylated cysteine-rich motif (TSRs) and our modeling of the β3GlcT sequence onto the Mfng structure was successful in identifying several residues responsible for the difference in donor specificity between these two enzymes.

We previously reported that the three Fng enzymes modify the same substrates, although with different catalytic efficiency (Rampal, *et al.*, 2005b). The ability of Lfng to self-modify is proof that this enzyme can catalyze the addition of GlcNAc to two very different Fuc substrates as the presence of Fuc on the *N*-linked glycan of the enzyme appears to be necessary for it to serve as an acceptor (data not shown, Raajit Rampal, Li

Shao). Whether there are instances of Fuc in other contexts, which might be modified by Fng enzymes and whether this self-modification has any *in vivo* consequence is as yet unknown. These differences in catalytic efficiency, taken together with the current results are highly suggestive that the three Fng enzymes have differences in specificity. The differences in catalytic efficiency toward the same substrates may be a consequence of substrate promiscuity, as the specificity of each enzyme may be optimized for a different substrate. Another possibility would be that one form of Fng enzyme may be more promiscuous than another, leading to an overall decrease in catalytic efficiency for the more promiscuous enzyme regardless of the substrate.

Several of our attempts to mutate the catalytic residue aspartate 289 failed. The D289N mutant failed to express altogether, while the D289S mutant was present only in cell lysates (Figure 15). However, the D289E mutant was purified and assayed and showed no catalytic function. Expression and secretion of this mutant was somewhat erratic and hard to reproduce, suggesting a possible folding or trafficking defect. It would seem that some care needs to be exercised to confirm protein production and localization if mutants of a Fng catalytic residue are to be used to draw conclusions about function *in vivo*. Well expressed catalytically inactive mutants such as S228L and T253A may be better choices.

We have observed that Lfng migrates as a doublet during SDS-PAGE (Figure 15, 19) and that PNGase F digestion of *N*-linked glycans eliminates this doublet (data not shown). Mutation of serine 168 to alanine or valine in the sequence NCS eliminates this doublet suggesting that asparagine 166 carries an *N*-linked glycan (Figure 19). This coincides with asparagine 109 in mouse Mfng which was mutated to glutamine by the authors of the Mfng structure to eliminate *N*-linked glycosylation at this site (Jinek, *et al.*, 2006). Unlike Mfng, mouse Lfng does not contain a second *N*-linked glycosylation site located in the long loop. In mouse Mfng, asparagine 185 in the sequence NRT is likely glycosylated, while the corresponding position in mouse Lfng is histidine 242 (Figure 5). Human Lfng however retains the asparagine, but the sequence is NRV, which should not be glycosylated. Since Mfng exhibits considerably lower catalytic efficiency toward EGF-*O*-fucose *in vitro* in comparison to Lfng (Rampal, *et al.*, 2005b), it is possible that this glycan contributes at least in part, to this lower efficiency by modulating the interaction of the long loop with the acceptor substrate. Considering the wealth of evidence that a glycan will restrict unfolded proteins to conformational space more closely resembling folded conformations (Arnold, *et al.*, 1999, Choi, *et al.*, 2008, Corzana, *et al.*, 2006, Imperiali and Rickert, 1995, Wyss, *et al.*, 1995), it is not hard to imagine that a large *N*-linked glycan might have a significant effect on the behavior of the long loop in solution.

The kinetic data from the Lfng mutants suggests that the Fuc is most likely positioned near what has previously been referred to as the putative Fuc binding pocket (Jinek, *et al.*, 2006). Whether this is indeed the case, and whether there is a rearrangement step in the mechanism will require further studies, including structural data with bound substrates and products. The V_{max} defects of the short loop mutants, coupled with a lack of K_M deficit on that side of the active site suggests only marginal contacts between the short loop and either UDP-GlcNAc or pNP-fucose. Another possibility is that the loop changes conformation subsequent to substrate binding, affecting the hydrogen bond between serine 177 and aspartate 288. The altered

conformation would then affect the position of the catalytic residue to achieve maximum catalytic efficiency. Perhaps most importantly, the F187L mutant of Lfng (F188L in human) may be catalytically active, contrary to previous reports (Sparrow, *et al.*, 2006) and as such, a simple enzymatic defect may not be the genetic cause of SCD.

Chapter Two

Characterization of Glycans

Summary

We used high pH anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) to identify monosaccharides, and determine disaccharide glycan structures released by β-elimination from modified protein substrates. The elution time of unknown glycans was compared with the elution time of standards of known structure to make the identification. We analyzed the enzymatic product of Rfng, providing definitive evidence that Rfng is a β3GlcNAcT. We performed a similar analysis with the β3GlcT showing conclusively that this is the enzyme elongating *O*-fucose glycans on TSRs. Finally, we analyzed the *O*-glycans from the TSRs of ADAMTS13. ADAMTS13 cleaves von Willebrand factor (vWf) into smaller fragments which are less thrombogenic (Schneppenheim, *et al.*, 2003). Decreased ADAMTS13 activity leads to the disease thrombotic thrombocytopenic purpura (TTP) (for a review see (Loirat, *et al.*, 2006)). In the case of a genetic deficiency of ADAMTS13 (also known as (Upshaw-Schulman Syndrome) (Loirat, *et al.*, 2006), treatment typically involves complete plasma transfusion every two to three weeks (Born, *et al.*, 2008). Recently Majerus and coworkers showed that loss of *O*-glycans from the TSRs of ADAMTS13 significantly decreased the secretion of this important enzyme (Ricketts, *et al.*, 2007).

Methods

Characterization of the Radical Fringe Product

The Rfng elongated 6[^{3}H]GlcNAc-β1,3-Fuc-*O*-EGF-repeat 26 of mouse N1 (prepared by Raajit Rampal) was dried in a speed-vac to approximately 50 µL volume. To this was added 250 µL of Buffer A (0.1% tri-fluoro-acetic acid (TFA) in MQ H_2O). The sample was centrifuged in a microfuge at maximum setting, and the supernatant removed to a clean tube multiple times until no pelleted debris was present. Injected sample onto a high-performance liquid chromatography (HPLC) system with a C18

column. Fractions were collected from a 30 minute gradient from 0% to 75% Buffer B (0.1% TFA, 80% acetonitrile in MQ H_2O), and the fractions containing radioactivity were consolidated and dried in a speed-vac. The *O*-glycans were released by β-elimination as described previously (Moloney, *et al.*, 1997) producing sugar-alditol products. Briefly, the sample was resuspended in 500 µL of 1 M $NaBH_4$ in 100 mM NaOH and heated at 55° C overnight. The sample was then pH neutralized drop-wise on ice with 4 M acetic acid. The sample was desalted over a 10-fold molar excess of Dowex 50W-X8 (BioRad). The flowthrough and 15 mL of MQ H_2O wash was collected and lyophilized overnight. The sample was redissolved in 100% methanol (MeOH) and dried on the speed-vac until no residual salt was evident. The sample was redissolved in MQ H_2O and samples were injected over the HPLC system using a superdex peptide column and fractions were collected. Fractions containing radioactivity were consolidated and dried in the speed-vac. The sample was redissolved in 100 mL of MQ H_2O and a fraction was injected over the Dionex DX300 HPAEC system (Dionex Corp.) using a CarboPac™ MA-1 column (Dionex Corp.). Fractions were collected using a 26 mM isocratic gradient of NaOH. After adding 100 µL of 1 M acetic acid to prevent chemi-luminescence, the samples were mixed with 5 mL of Scintiverse II (Fisher) scintillation fluid and scintillation counted on a 1209 Rackbeta scintillation counter (LKB Wallac). The elution time of the radioactive sample was then compared to the elution time of appropriate disaccharide standards to identify the structure.

Characterization of the β1,3-Glucosyltransferase Product

Characterization of the 6[^{3}H]Glc-β1,3-Fuc-*O*-TSR product of β3GlcT (prepared by Hofsteenge and coworkers) was performed as just described, with the following changes. Subsequent to desalting over Dowex 50W-X8, the sample was pushed through a Sep-pak light C18 cartridge (Waters), and the flowthrough was collected for the next step. Additionally, due to unexplained lability of the radioactive product upon drying in the speed-vac, samples were dried under nitrogen stream or by lyophilization. The gradient used for the HPAEC step was as described previously (Moloney, *et al.*, 1997).

Characterization of the ADAMTS13 O-*Glycans*

Characterization of the disaccharides released from ADAMTS13 was performed as described for the β3GlcT product characterization with the following changes. The modified ADAMTS13 samples (prepared by Majerus and coworkers) were delivered on proteinA sepharose beads. The beads were resuspended in PNGase digestion buffer A (1% sodium dodecyl-sulfate (SDS), 1% β-mercaptoethanol, vortexed, and heated at 100° C for 5 minutes. The samples were cooled to room temperature and 450 µL of PNGase digest buffer B (50 mM Tris-HCl pH 8.6, 10 mM ethylene-diamine-tetra-acetic acid (EDTA), 0.7% w/v nonidet P-40 (NP-40), and 0.02% sodium azide (NaN_3)) was added. The buffer B had been freshly prepared with 1 tablet of complete-Mini, EDTA-free protease inhibitor (Roche) dissolved in 10 mL. The PNGase enzyme was added, and the samples were incubated at room temperature overnight. The next day, 100 mL of 10% SDS was added, the samples heated to 100° C for 5 minutes, and then cooled on ice. The samples were chromatographed over a 30 cm by 1 cm G50 column in G50 buffer (50 mM ammonium formate, 0.1% SDS, 0.02% NaN_3). Fractions were collected and the radiation containing fractions were consolidated and lyophilized overnight. The lyophilized

samples were resuspended in 1 mL of MQ H_2O and then acetone precipitated with 8 volumes of ice cold acetone at -20° C for 4 hours. The samples were pelleted in a microfuge, and the pellet retained. The *O*-glycans were then β-eliminated from the pellets as previously described. The speed vac was used for drying the samples rather than a nitrogen stream. The oligosaccharide and potential monosaccharide peaks observed after the superdex sizing column were consolidated separately and dried in the speed vac. The samples were acid hydrolyzed in 2 M TFA at 100° C for 2 hours, dried in the speed-vac, resuspended in MQ H_2O and chromatographed over the HPAEC system.

Radical Fringe is a β1,3-N-acetylglucosaminyltransferase

Results

Purified, overexpressed Radical fringe was incubated with UDP-[^{3}H]GlcNAc and an *O*-fucosylated EGF repeat (Rampal, *et al.*, 2005b). To confirm that the appropriate product was formed, the product was characterized chromatographically. The radiolabeled EGF repeat was purified by reverse-phase HPLC and the tubes containing the radioactivity were consolidated (Figure 25A). The *O*-glycans were released from the HPLC purified product by alkali-induced β-elimination. The released glycans were size fractionated using a Superdex gel filtration column and compared to the elution time of sugar standards of increasing size (Figure 25B). These results indicate that all of the radioactivity was present in a disaccharide as expected. Finally the samples from the tubes containing the radioactivity from the sizing column were subjected to HPAEC separation and compared to standards of known structure (Figure 25C and (Rampal, *et al.*, 2005b)). This shows that the glycan product was exclusively GlcNAc-β1,3-Fucitol as expected and that Rfng is a true Fng enzyme (Figure 25 C and (Rampal, *et al.*, 2005b)).

Discussion

We have shown conclusively that Rfng is a fucose β1,3-N-acetylglucosaminyl-transferase Fng enzyme. While the presence of radioactivity on EGF-*O*-fucose substrate after incubation with Rfng and UDP-[6-^{3}H]GlcNAc, and the sequence similarity with the other Fng enzymes (Figure 5) were highly suggestive of the enzymatic function of Rfng, only by isolating and characterizing the radioactive product of the enzyme could the identity of the Rfng enzyme activity be conclusively determined.

The β1,3-Glucosyltransferase Product

Results

Purified, overexpressed β3GlcT was incubated with UDP-[6-^{3}H]glucose and an *O*-fucosylated TSR (Kozma, *et al.*, 2006). To confirm that the enzyme was making the predicted structure, we characterized the product chromatographically. We purified the β3GlcT radioactively modified rat F-Spondin TSR-4 substrate (prepared by Hofsteenge and coworkers (Kozma, *et al.*, 2006)) over a C18 column on an HPLC (Figure 26 A). The radioactive peak was subjected to β-elimination to release the *O*-linked glycans. The released glycans were run over a Superdex peptide sizing column and the elution time of the radioactive peak was compared to the elution time of sugar standards of increasing size (Figure 26B). This indicated that all of the radioactivity was present in a disaccharide as expected. Finally the samples containing the radioactivity from the sizing column were analyzed by HPAEC and compared to standards of known structure (Figure

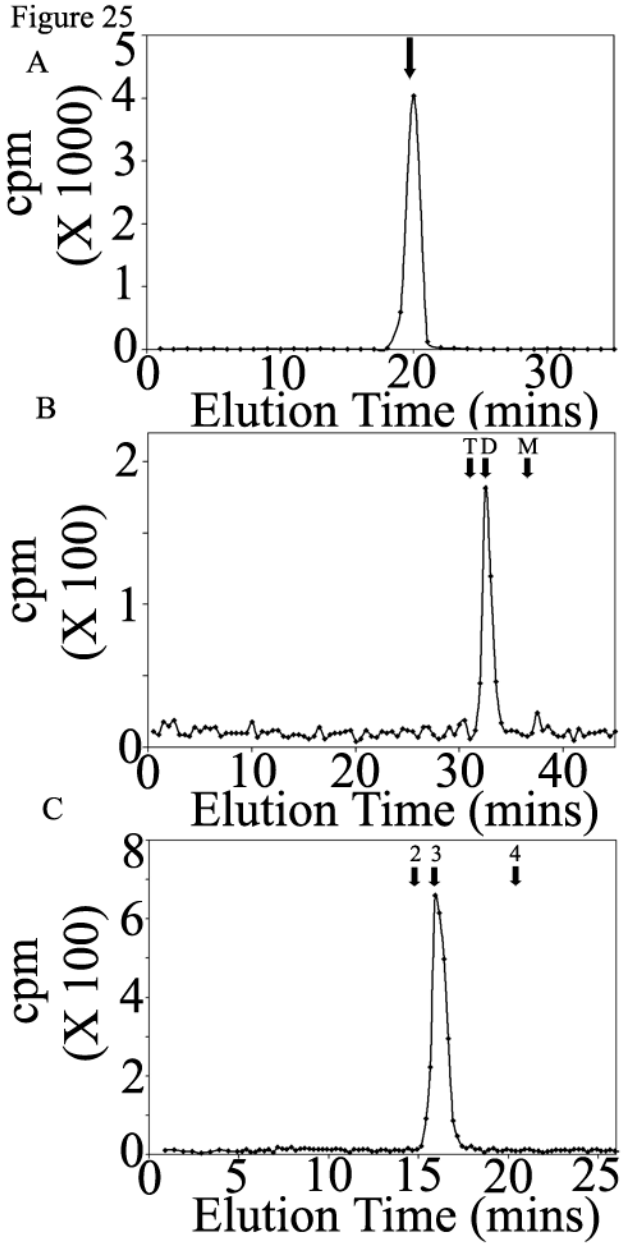


Figure 25. Analysis of *O*-Glycans From Radical Fringe Modified EGF-like repeat-26

A) HPLC analysis of Rfng modified EGF-repeat 26. The arrow indicates the elution point of EGF. **B)** Superdex analysis of β-eliminated *O*-glycans from Rfng modified EGF-repeat 26. The elution time of trisaccharide (T), disaccharide (D), and monosaccharide (M) are indicated. **C)** HPAEC analysis of the disaccharide peak from panel C. The elution points of GlcNAc-β1,2-fucitol (2), GlcNAc-β1,3-fucitol (3), and GlcNAc-β1,4-fucitol (4) standards are indicated.

Figure 26

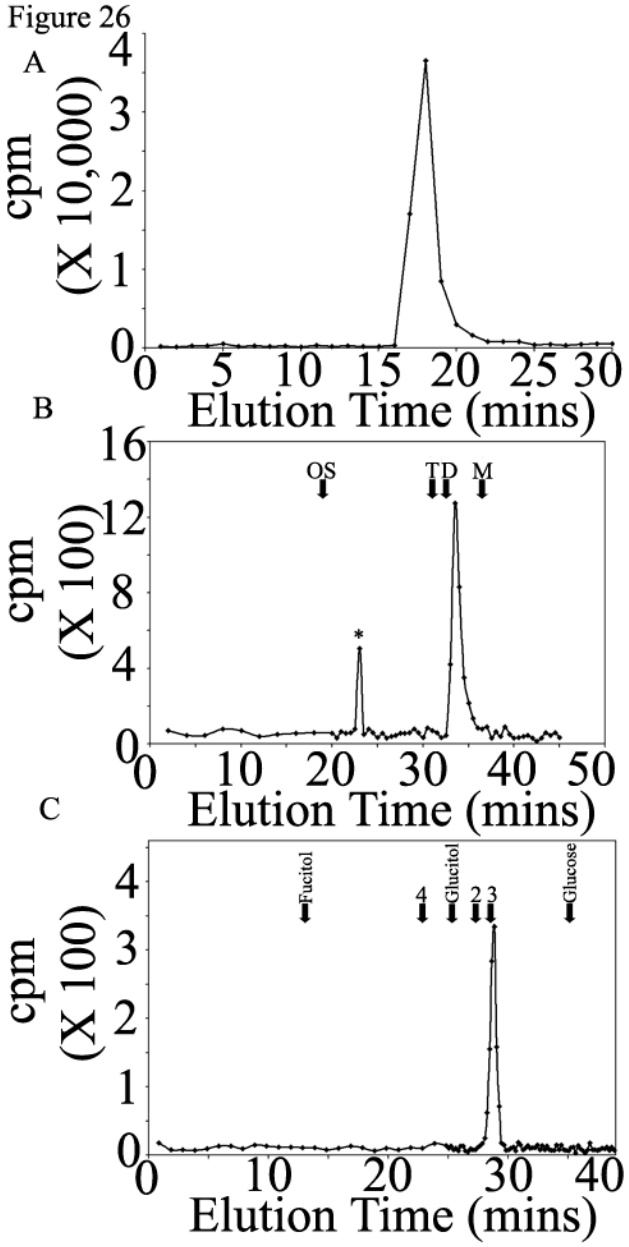


Figure 26. Characterization of *O*-Glycans From β1,3-Glucosyltransferase Modified Rat F-Spondin TSR4

A) HPLC profile of rat F-spondin TSR4. **B)** Superdex profile of *O*-glycans β-eliminated from TSR4. Arrows indicate elution position of mono (M), di (D), tri (T), and oligosaccharide (OS) standards. The asterisk indicates a likely phantom peak. C) HPAEC profile of the disaccharide peak from panel B. Arrows indicate elution position of Glc-β1,2-fucitol (2), Glc-β1,3-fucitol (3), Glc-β1,4-fucitol (4), Fucitol, Glucitol, and Glucose standards.

26C and (Kozma, *et al.*, 2006)). This shows that the glycan product was exclusively Glc-β1,3-Fucitol as expected (Figure 26 C and (Kozma, *et al.*, 2006)).

Discussion

We showed conclusively that all radioactivity incorporated into rat F-spondin TSR4 by the β3GlcT enzyme was present as a disaccharide of the form Glc-β1,3-Fucitol after β-elimination. and that this is the TSR modifying β1,3-Glucosyltransferase enzyme (Figure 26 C and (Kozma, *et al.*, 2006)).

Glycans on Thrombospondin Repeats of ADAMTS13

Results

To analyze the structure of the [^{3}H]fucose-labeled structures on ADAMTS-13, we performed the following analysis. Majerus and coworkers grew HEK293 TREx cells stably expressing V5 epitope tagged human ADAMTS13 in serum-free media overnight with 50 μCi/mL of L-[6-^{3}H]fucose. They then harvested the media and lysed the cells in the presence of protease inhibitors and precipitated the protein with protein-G Sepharose. The samples were then subjected to PNGase digestion (Ricketts, *et al.*, 2007). We removed low molecular weight contaminants from the radioactively labeled human ADAMTS13 samples (Ricketts, *et al.*, 2007)) over a G50 column (Figure 27A). We then released the *O*-glycans from the human ADAMTS-13 by alkali-induced β-elimination and analyzed the released glycans using a Superdex peptide column (Figure 27B and (Ricketts, *et al.*, 2007)). This revealed that the radioactive Fuc was present in both a disaccharide, and an unexpected oligosaccharide *O*-linked glycan structure. The radioactivity for each peak was consolidated and analyzed chromatographically. The disaccharide peaks were analyzed by HPAEC and compared to standards of known structure (Figure 27C and (Ricketts, *et al.*, 2007)).

This shows that the disaccharide is exclusively Glc-β1,3-Fucitol, the expected form of *O*-fucose on the TSRs. In addition this data suggests that ADAMTS-13 TSRs carry predominantly the disaccharide form of the glycan, with very little monosaccharide (Figure 27 C and (Ricketts, *et al.*, 2007)). To determine whether the oligosaccharide peaks were derived from *O*-fucose or other types of *O*-glycans with terminal fucose residues, the samples were acid hydrolyzed into their monosaccharide constituents and then run over the HPAEC system to identify fucitol (derived from *O*-fucose glycans) or Fuc (derived from other types of *O*-glycans) (Figure 27D). The presence of fucitol indicates that the radioactive oligosaccharide peaks contained the β-elimination products of *O*-fucose glycans. There was also radioactive Fuc present suggesting that there may be a mixture of structures in the peaks from panel B. The structure(s) contained in this oligosaccharide peak have not been further characterized.

Discussion

We have shown that the TSRs of ADAMTS-13 appear to contain exclusively elongated (or otherwise modified) *O*-fucose glycans. These glycans are predominantly of the expected disaccharide form Glc-β1,3-Fuc-*O* (Figure 27 C and (Ricketts, *et al.*, 2007)). Additionally, we have observed the presence of an as yet uncharacterized *O*-fucose glycan structure on these TSRs (Figure 27D). There can be a number of reasons why this structure would elute from the sizing column at approximately near the void

Figure 27

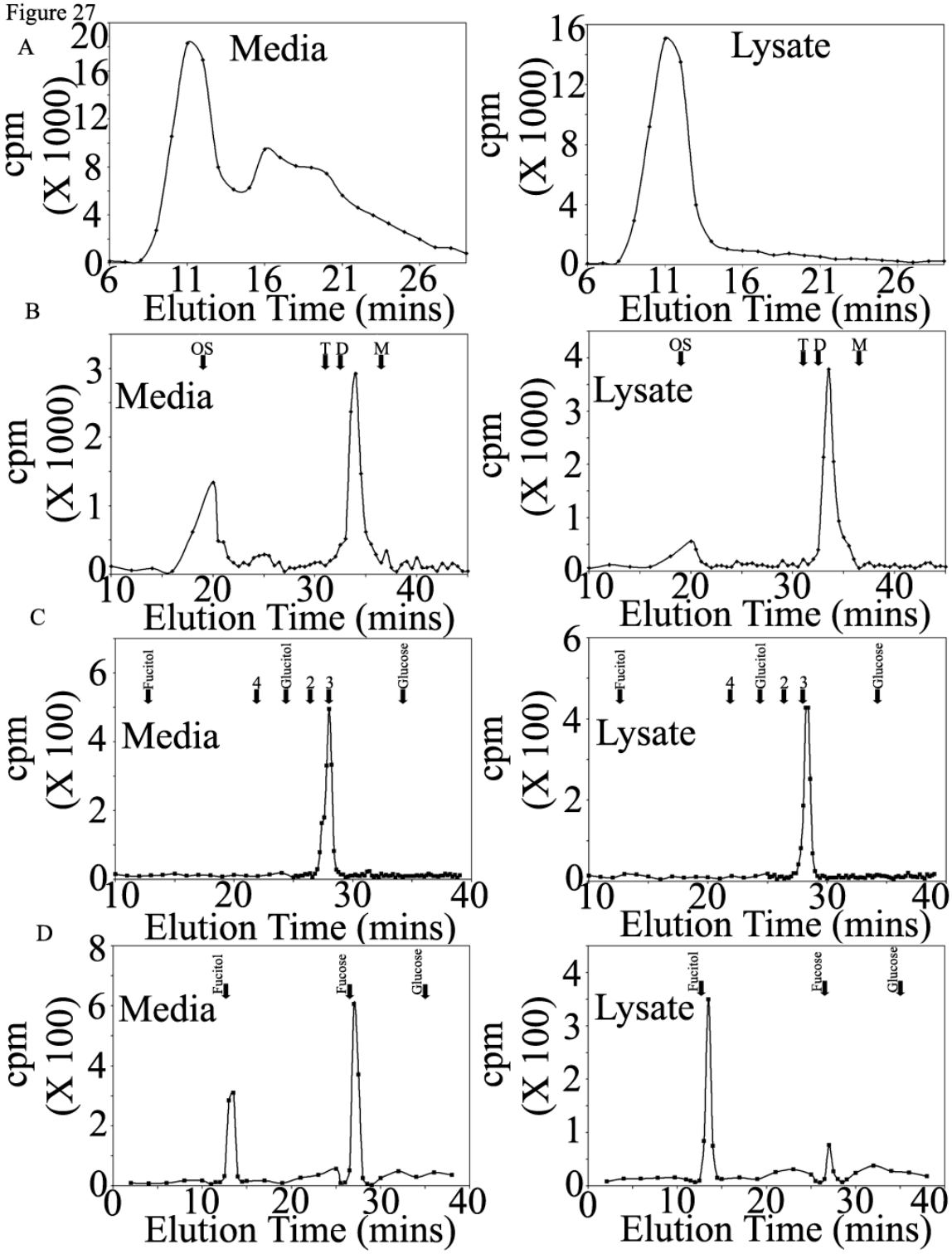

A
Media
Lysate
cpm (X 1000)
Elution Time (mins)
B
OS
T
P
M
C
Fucitol
4
Glucitol
2
3
Glucose
cpm (X 100)
D
Fucose

Figure 27. Analysis of *O*-Glycans From ADAMTS-13
A) Sephadex G50 cleanup of ADAMTS13 **B)** Superdex analysis of β-eliminated *O*-glycans from ADAMTS-13. The elution position of monosaccharide (M), disaccharide (D), trisaccharide (T), and oligosaccharide (OS) standards are indicated by arrows. **C)** HPAEC analysis of disaccharide peaks from panel B. Elution position of Glc-β1,2-fucitol (2), Glc-β1,3-fucitol (3), Glc-β1,4-fucitol (4), Fucitol, Glucitol, and Glucose standards are indicated by arrows. **D)** HPAEC analysis of the acid hydrolyzed oligosaccharide peaks from panel B.

volume. First, it may be a large oligosaccharide. Second, it could contain a charged group, as they run aberrantly large in this system (Moloney, *et al.*, 1997). An intriguing possibility is suggested by an as yet unpublished observation that in *Drosophila*, *O*-fucose glycans on EGF-repeats may be modified by glucuronate (GlcA), presumably as a branched structure in addition to GlcNAc from Fng modification (Mike Tiemeyer, personal communication). Alternatively it could be modified with Sia, phosphate or sulfate. The presence of radioactive Fuc in addition to fucitol after acid hydrolysis (Figure 27 D) suggests that the *O*-fucose glycan may be further fucosylated, that the oligosaccharide peak contains a heterogenous mixture of glycans, some of which contain Fuc, and/or that the peak contains some residual *N*-linked glycans despite the use of PNGase F to eliminate them prior our analysis.

Chapter Three

Database Searches For O*-Glycan Consensus Bearing Sequences*

Summary

We have attempted to identify all of the known mammalian and marsupial proteins potentially modified by *O*-fucose and all known proteins potentially modified with *O*-glucose glycans by searching the SwissProt database with the published consensus sequences for glycan addition. We identified 111 proteins potentially modified with *O*-fucose (110 mammalian, 1 metatherian). We identified 45 proteins potentially modified with *O*-glucose, 5 of which are not predicted to also contain *O*-fucose glycans.

Methods

Identification of proteins containing EGF-repeats that may be modified by *O*-fucose was achieved by using the Motif search capability located on the web at: http://motif.genome.jp/MOTIF2.html. We utilized the PROSITE (Hofmann, *et al.*, 1999) format for pattern searching, utilizing the search patterns shown in Table 4. We repeated the process for potential *O*-glucose modified proteins using the appropriate PROSITE (Hofmann, *et al.*, 1999) formatted search string shown in Table 4. In each case, the search results were manually culled for redundant hits and the presence of a consensus sequence for glycosylation was manually confirmed for each hit.

EGF-like Repeats With *O*-fucose Consensus Sequences

Results

We began our search for proteins potentially modified by *O*-fucose by first refining the search string annotated for EGF-repeats found at PROSITE (Hofmann, *et al.*, 1999). The original search strings obtained from PROSITE, and our modified versions are shown in Table 4. We then inserted the published consensus sequence for *O*-fucosylation (Shao, *et al.*, 2003) to obtain our final search string (Table 4). The final result of our search is shown in Table 5.

Discussion

We identified a total of 111 proteins potentially modified by *O*-fucose (Rampal, *et al.*, 2007). We had to limit ourselves to mammalian proteins due to the overwhelming amount of data obtained (one hit for every EGF-repeat potentially modified), and the large number of proteins from various invertebrates and plants that were identified. We do not know if plants indeed possess *O*-fucosyltransferases that might modify such sequences. We discovered early on that the original EGF-repeat consensus sequences from PROSITE that we had begun with were producing large

Table 4. PROSITE and Regular Expression Search Syntax For *O*-Glycosylated EGF-like Repeats in the SwissProt Database

For PROSITE: a: aromatic amino acids, x: any amino acid, (4,5): four to five copies of the previous term, {c}: any amino acid except cysteine, [ST]: serine or threonine, - is a required spacer between terms and a period is required at the end of the search string. For general regular expressions: [^C]: any amino acid except cysteine, {4,5}: four to five copies of the previous term, [ST]: serine or threonine.

Search Strings For EGF-like Repeats Containing *O*-Fucose Consensus Sites	
PROSITE EGF-Repeat sequence	x(4)-C-x(0,48)-C-x(3,12)-C-x(1,70)-C-x(1,6)-C-x(2)-G-a-x(0,21)-G-x(2)-C-x. x(4)-C-x(0,48)-C-x(3,12)-C-x(1,70)-C-x(1,6)-C-x(2)-G-a-x(0,21)-G-x(2)-C-x.
Altered EGF-repeat search sequence	C-{C}(0,48)-C-{C}(3,12)-C-{C}(1,70)-C-{C}(1,6)-C-{C}(2)-G-[FYW]-{C}(0,24)-C-x. C-{C}(0,48)-C-{C}(3,12)-C-{C}(1,70)-C-{C}(1,6)-C-{C}(3)-[FYW]-{C}(0,21)-G-{C}(2)-C-x.
Plus *O*-fucose Consensus	C-{C}(0,48)-C-{C}(4,5)-[ST]-C-{C}(1,70)-C-{C}(1,6)-C-{C}(2)-G-[FYW]-{C}(0,24)-C-x. C-{C}(0,48)-C-{C}(4,5)-[ST]-C-{C}(1,70)-C-{C}(1,6)-C-{C}(3)-[FYW]-{C}(0,21)-G-{C}(2)-C-x.
General Regular Expression Syntax	C[^C]{0,48}C[^C]{4,5}[ST]C[^C]{1,70}C[^C]{1,6}C[^C]{2}G[FYW][^C]{0,24}Cx C[^C]{0,48}C[^C]{4,5}[ST]C[^C]{1,70}C[^C]{1,6}C[^C]{3}[FYW][^C]{0,21}G[^C]{2}Cx
Search Strings For EGF-like Repeats Containing *O*-Fucose Consensus Sites	
EGF-Repeat With *O*-glucose Consensus	C-{C}-[ST]-{C}-P-C-{C}(3,12)-C-{C}(1,70)-C-{C}(1,6)-C-{C}(2)-G-[FYW]-{C}(0,24)-C-x. C-{C}-[ST]-{C}-P-C-{C}(3,12)-C-{C}(1,70)-C-{C}(1,6)-C-{C}(3)-[FYW]-{C}(0,21)-G-{C}(2)-C-x.
General Regular Expression Syntax	C[^C][ST][^C]PC[^C]{3,12}C[^C]{1,70}C[^C]{1,6}C[^C]{2}G[FYW][^C]{0,24}Cx C[^C][ST][^C]PC[^C]{3,12}C[^C]{1,70}C[^C]{1,6}C[^C]{3}[FYW][^C]{0,21}G[^C]{2}Cx

Table 5. Proteins Identified in the SwissProt Database Containing EGF-like Repeats with *O*-fucose Consensus Sequences

The table is organized alphabetically, first by species, then by protein name.

Protein Name	Species	Accession Number
Insulin receptor-related protein	*Cavia porcellus*	P14617
megalin	*Didelphis virginiana*	Q6E0K3
CHz-cadherin	*Gallus gallus*	Q6W4W6
Teneurin 1	*Gallus gallus*	Q9W6V6
Agrin	*Homo sapiens*	O00468
AMACO	*Homo sapiens*	Q70UZ8
Basement membrane-specific heparan sulfate proteoglycan core protein	*Homo sapiens*	P98160
Brevican core protein	*Homo sapiens*	Q96GW7
Cadherin EGF LAG seven-pass G-type receptor 1	*Homo sapiens*	Q9NYQ6
Cadherin EGF LAG seven-pass G-type receptor 2	*Homo sapiens*	Q9HCU4
Cadherin EGF LAG seven-pass G-type receptor 3	*Homo sapiens*	Q9NYQ7
Cadherin-related tumor suppressor homolog	*Homo sapiens*	Q14517
CD97 antigen	*Homo sapiens*	P48960
Cell surface receptor (PRV1) (Polycythemia rubra vera 1)	*Homo sapiens*	Q9HDA5
Coagulation factor IX	*Homo sapiens*	P00740
Coagulation factor VII	*Homo sapiens*	P08709
Complement component C1q receptor	*Homo sapiens*	Q9NPY3
Cripto-1 growth factor	*Homo sapiens*	P13385
Cripto-3 growth factor	*Homo sapiens*	P51864
Crumbs protein homolog 1	*Homo sapiens*	P82279
Crumbs-like protein 2	*Homo sapiens*	Q5IJ48
Cryptic	*Homo sapiens*	Q9GZR3
Cubilin	*Homo sapiens*	Q5VTA6
Delta-like protein 1	*Homo sapiens*	O00548
Delta-like protein 3	*Homo sapiens*	Q9NYJ7
Delta-like protein 4	*Homo sapiens*	Q9NR61
DNER protein	*Homo sapiens*	Q8TB42
EGF-like module containing mucin-like hormone receptor-like 1	*Homo sapiens*	Q14246
EGF-like module containing mucin-like hormone receptor-like 2	*Homo sapiens*	Q9UHX3
EGF-like module containing mucin-like hormone receptor-like 3	*Homo sapiens*	Q9BY15
EGF-like module containing mucin-like hormone receptor-like 4	*Homo sapiens*	Q86SQ3
EGF-like repeats and discoidin I-like domains protein 3	*Homo sapiens*	O43854

FAT3 protein	*Homo sapiens*	Q96AU6
Fat-like cadherin FATJ protein	*Homo sapiens*	Q6V0I7
Fibrillin 1	*Homo sapiens*	Q75N87
Fibrillin 2	*Homo sapiens*	P35556
Fibrillin 3	*Homo sapiens*	Q75N90
Fibulin-1	*Homo sapiens*	P23142
Fibulin-2	*Homo sapiens*	P98095
Hepatocyte Growth Factor Activator	*Homo sapiens*	Q04756
HSPC013	*Homo sapiens*	Q9Y2R7
Jagged 1	*Homo sapiens*	P78504
Jagged 2	*Homo sapiens*	Q9Y219
KRAB zinc finger protein	*Homo sapiens*	Q86UQ0
latent transforming growth factor beta binding protein 2	*Homo sapiens*	Q14767
low density lipoprotein receptor related protein 1	*Homo sapiens*	Q07954
MEGF10 protein	*Homo sapiens*	Q96KG7
MEGF11 protein	*Homo sapiens*	Q96KG6
multimerin 1	*Homo sapiens*	Q13201
Multiple EGF-like-domain protein 3	*Homo sapiens*	O75095
Multiple EGF-like-domain protein 7	*Homo sapiens*	Q9UHF1
Multiple EGF-like-domain protein 9	*Homo sapiens*	Q6UY11
Netrin G1	*Homo sapiens*	Q9Y2I2
Netrin G2	*Homo sapiens*	Q96CW9
Neurexin 1-alpha	*Homo sapiens*	Q9ULB1
Neurexin 2-alpha	*Homo sapiens*	Q9P2S2
Neurexin 3-alpha	*Homo sapiens*	Q9Y4C0
Neurocan core protein	*Homo sapiens*	O14594
Neurogenic locus notch homolog protein 1	*Homo sapiens*	P46531
Neurogenic locus notch homolog protein 2	*Homo sapiens*	Q04721
Neurogenic locus notch homolog protein 3	*Homo sapiens*	Q9UM47
Neurogenic locus notch homolog protein 4	*Homo sapiens*	Q99466
Nidogen-2	*Homo sapiens*	Q14112
Notch homolog 2 (*Drosophila*) N-terminal like	*Homo sapiens*	Q5VTG9
P18 protein	*Homo sapiens*	Q96RT2
PLAT protein	*Homo sapiens*	Q9BU99
Protein kinase C-binding protein NELL1	*Homo sapiens*	Q92832
Protocadherin Fat 2	*Homo sapiens*	Q9NYQ8
Reelin	*Homo sapiens*	P78509
Slit homolog 1 protein	*Homo sapiens*	O75093
Slit homolog 2 protein	*Homo sapiens*	O94813
Slit homolog 3 protein	*Homo sapiens*	O75094
SNED 1 protein	*Homo sapiens*	Q8N369
Stabilin 1	*Homo sapiens*	Q9NY15
Stabilin 2	*Homo sapiens*	Q8WWQ8
Tenascin	*Homo sapiens*	P24821

Thrombospondin 3	*Homo sapiens*	P49746
Tissue-type plasminogen activator	*Homo sapiens*	P00750
Tumor necrosis factor receptor superfamily member 11B	*Homo sapiens*	O00300
Tumor necrosis factor receptor superfamily member 5	*Homo sapiens*	P25942
Tumor necrosis factor receptor superfamily member 5	*Homo sapiens*	Q86YK5
Tyrosine-protein kinase receptor Tie-1	*Homo sapiens*	P35590
Urokinase-type plasminogen activator	*Homo sapiens*	P00749
Uromodulin	*Homo sapiens*	P07911
Vasorin	*Homo sapiens*	Q6EMK4
Versican core protein	*Homo sapiens*	P13611
Wnt inhibitory factor 1	*Homo sapiens*	Q9Y5W5
Cell surface antigen 114/A10	*Mus musculus*	P19467
Crumbs-like protein 1	*Mus musculus*	Q8VHS2
DOC4	*Mus musculus*	O70465
EGF repeat transmembrane protein	*Mus musculus*	Q61204
F12 protein	*Mus musculus*	Q80YC5
Jedi protein	*Mus musculus*	Q8VHL7
Lactadherin	*Mus musculus*	P21956
MAFA	*Mus musculus*	O88713
Novel protein, ortholog of human scavenger receptor class F, member 1 SCARF1	*Mus musculus*	Q5ND28
ODZ3	*Mus musculus*	Q9JLC1
Polydom protein	*Mus musculus*	Q9ES77
Proz	*Mus musculus*	Q8CI01
Secreted nidogen domain protein	*Mus musculus*	Q70E20
Secreted protein SST3	*Mus musculus*	Q810H2
Similar to Fta3 protein	*Mus musculus*	Q80VA2
Slit-like 2 protein	*Mus musculus*	Q8R2G5
Tdgf1 protein	*Mus musculus*	Q7TQ06
Ten m4	*Mus musculus*	Q9WTS7
Teneurin-2 protein	*Mus musculus*	Q9QYZ1
Ten-m1	*Mus musculus*	Q9WTS4
Coagulation factor X	*Oryctolagus cuniculus*	O19045
mucin 13	*Rattus norvegicus*	P97881
Multiple EGF-like-domain protein 7	*Rattus norvegicus*	Q6AZ60
FXII	*Sus scrofa*	O97507

numbers of false positive hits and that many of these hits involved sequences from the C-terminal end of one EGF-repeat and the N-terminal end of a neighboring EGF-repeat in various proteins. We then determined that placing an X in the search string to indicate any amino acid, in between the conserved cysteines was inappropriate, and that the appropriate syntax was {C} indicating any amino acid except cysteine (Table 4). This greatly improved the accuracy of the search. Performing the search in this fashion using publicly available resources accessible on the World Wide Web is useful, and has provided us with a wealth of previously unidentified potential targets for the Pofut1 enzyme. However, there is one major shortcoming. Searching with a consensus sequence for *O*-fucosylation that has been arrived at by analysis of only a few proteins, with EGF-repeat sequences potentially related to each other could bias the search. The best search would identify all EGF-repeats possessing a serine or threonine anywhere in the sequence between the second and third conserved cysteines. Unfortunately, performing this search using the publicly available resources previously mentioned is not a trivial task. The PROSITE syntax for motif searches appears to be a form of regular expression used for performing string searches by computer. As best as we can determine, the generally accepted regular expression syntax is not used in the PROSITE search strings, and the PROSITE form of the strings unfortunately uses certain syntax structures that mean something entirely different in the general regular expression syntax. As one example, the statement {C}(4-5) is PROSITE syntax for four to five amino acids of any identity except cysteine. In the general syntax for regular expressions this would be [^C]{4-5}. Note specifically that the curly braces indicate two very different things in the general and PROSITE syntax. The general regular expression syntax equivalent of the PROSITE search strings we used at the Motif search web site are included in Table 4 for comparison.

EGF-like Repeats With *O*-glucose Consensus Sequences

Results

We have attempted to identify all known proteins in the SwissProt database that contain the previously published consensus sequence for *O*-glucosylation (Moloney, *et al.*, 2000b). The search strings used are shown in Table 4. We have identified 45 proteins that are potentially modified by *O*-glucose, shown in Table 6. There are only five proteins that contain *O*-glucose on EGF-repeats, which do not also contain EGF-repeats with *O*-fucose. Three of these are thrombospondins which do contain *O*-fucose on TSRs (Table 6). The remaining proteins containing only *O*-glucose glycans are hyaluronan-binding protein 2 in mice, and 63 kDa sperm flagellar membrane protein in the purple sea urchin.

Discussion

There is considerably less data accumulated from Motif searches of SwissProt for *O*-glucosylation compared to *O*-fucosylation so we have not been forced to restrict ourselves to mammalian and marsupial proteins. The few examples where *O*-glucose is present on EGF-repeats without EGF-repeat *O*-fucosylation also being present is interesting to note. The implications of this, if any, are unknown.

Table 6. Proteins Identified in the SwissProt Database Containing EGF-like Repeats with *O*-glucose Consensus Sequences

Grey cells with white type indicate EGF-repeat containing proteins predicted to contain *O*-glucose glycans without the presence of *O*-fucose glycans. The table is organized alphabetically first by species, then by protein name.

Protein Name	Species	Accession Number
Aggrecan Core Protein	*Bos taurus*	P13608
Cadherin 4	*Caenorhabditis elegans*	Q19319
Protocadherin-like wing polarity protein stan	*Drosophila melanogaster*	Q9V5N8
Cadherin EGF LAG seven-pass G-type receptor 1	*Homo sapiens*	Q9NYQ6
Cadherin EGF LAG seven-pass G-type receptor 2	*Homo sapiens*	Q9HCU4
Cadherin EGF LAG seven-pass G-type receptor 3	*Homo sapiens*	Q9NYQ7
Coagulation Factor VII	*Homo sapiens*	P08709
Coagulation Factor IX	*Homo sapiens*	P00740
Complement component C1q receptor	*Homo sapiens*	Q9NPY3
Crumbs homolog 1	*Homo sapiens*	P82279
Crumbs homolog 2	*Homo sapiens*	Q5IJ48
Cubulin	*Homo sapiens*	O60494
Delta and Notch-like epidermal growth factor-related receptor	*Homo sapiens*	Q8NFT8
Delta-like 1	*Homo sapiens*	O00548
Delta-like 4	*Homo sapiens*	Q9NR61
Delta-like protein	*Homo sapiens*	P80370
Delta-like protein 2	*Homo sapiens*	Q6UY11
Fibulin 1	*Homo sapiens*	P23142
Fibrillin 2	*Homo sapiens*	P35556
Fibrillin 3	*Homo sapiens*	Q75N90
Hepatocyte growth factor activator	*Homo sapiens*	Q04756
Jagged 1	*Homo sapiens*	P78504
Jagged 2	*Homo sapiens*	Q9Y219
Notch 1	*Homo sapiens*	P46531
Notch 2	*Homo sapiens*	Q04721
Notch 3	*Homo sapiens*	Q9UM47
Notch 4	*Homo sapiens*	Q99466
Protocadherin Fat 1	*Homo sapiens*	Q14517
Slit homolog 1	*Homo sapiens*	O75093
Slit homolog 2	*Homo sapiens*	O94813
Slit homolog 3	*Homo sapiens*	O75094
Thrombospondin 1	*Homo sapiens*	P07996
Thrombospondin 2	*Homo sapiens*	P35442

Thrombospondin 4	*Homo sapiens*	P35443
Versican core protein	*Homo sapiens*	P13611
Vitamin K-dependent protein Z	*Homo sapiens*	P22891
Delta-like 3	*Mus musculus*	O88516
Hyaluronan-binding protein 2	*Mus musculus*	Q8K0D2
Adhesive plaque matrix protein 2	*Mytilus galloprovincialis*	Q25464
Mucin 13	*Rattus norvegicus*	P97881
Protocadherin Fat 2	*Rattus norvegicus*	O88277
63 kDa sperm flagellar membrane protein	*Strongylocentrotus purpuratus*	Q07929
Fibropellin 1	*Strongylocentrotus purpuratus*	P10079
Fibropellin 3	*Strongylocentrotus purpuratus*	P49013
Sperm receptor for egg jelly	*Strongylocentrotus purpuratus*	Q26627

Future Directions

The mutagenesis study we conducted on Lfng has given us some insight into the behavior of the two disordered loops in this enzyme. It appears that the short loop does not play a role in substrate binding since there are no K_M affects from mutations in this region. Some researchers have proposed that disordered loops could be responsible for ejecting substrates (specifically nucleotides) remaining after catalysis in order to prevent dead-end complexes from forming (Unligil and Rini, 2000). We judge this unlikely in that we see very small effects on the K_M for UDP-GlcNAc when we mutate this loop. If this is the function of the short loop, we might expect a less dramatic V_{max} effect using a small molecule acceptor substrate like pNP-fucose which presumably does not have strictly ordered binding with the enzyme like the larger EGF-*O*-fucose acceptor substrate. Determining whether this loop has some effect on the positioning of the catalytic residue would likely necessitate co-crystallization of Mfng with UDP and EGF-*O*-fucose, and with UDP and a disaccharide modified EGF-repeat as well, if possible. Considering how difficult it is to produce large amounts of pure EGF-*O*-fucose this could be a daunting task. However, with current technologies, including crystallization robots allowing the researcher to screen large numbers of conditions with very small amounts of protein this goal would seem to be achievable. Additionally, Chigira *et al.* have recently published a paper showing that during attempts to engineer human-like glycosylation in yeast, they have managed to produce fucosylated EGF-repeats (Chigira, *et al.*, 2008). This may be a cheap and efficient means toward producing large amounts of this valuable reagent. In the meantime, attempts at soaking small molecule acceptor substrates into Mfng crystals might provide some insight into substrate binding. This may require screening for co-crystallization conditions despite the small size of the pNP-fucose acceptor substrate due to the very limited solubility of the compound in aqueous solution.

Our work has shown that the Fng assay is a very sensitive readout for mutagenesis screening of the Fng enzymes. We were able to see dramatic effects for the most conservative of mutations (namely C290S) and relatively modest effects from some non-conservative mutations (such L229Q, I233A and A235Y). We have not identified any amino acid sidechains in Lfng that we can say affect the affinity of the enzyme for EGF-*O*-fucose mainly because of a lack of fucosylated EGF-repeat at our disposal. Anticipating the possibility that making large amounts of EGF-*O*-fucose may have recently become economically and technically feasible (Chigira, *et al.*, 2008), the next obvious step would be to determine if the enzyme can be saturated with the EGF-*O*-fucose substrate and proceed to analyze the Lfng mutants we have made with the EGF-*O*-fucose acceptor. We know that some EGF-repeats are easily modified by Fng (Factor IX, mouse N1 EGF-repeat 26), while others are not (Factor VII, mouse N1 EGF-repeat 24) (Rampal, *et al.*, 2005b). It may be that part of the specificity determinant lies with the enzyme, and the ability to make large amounts of these reagents coupled with the exquisite sensitivity possible with the Fng assay might allow the researcher to determine

a Fng consensus sequence. This would be enormously valuable considering how many potentially *O*-fucosylated EGF-repeats are found in mammalian proteins alone. The goal of teasing out the details of how Fng fine tunes Notch signaling would be advanced enormously by identification of a Fng consensus sequence. We also have tantalizing evidence that Lfng may have a slow-on component to its mechanism with the EGF-*O*-fucose substrate and investigating this possibility with EGF-*O*-fucose substrate, and possibly larger fragments of the Notch extracellular domain would be of interest.

Perhaps, most significantly, utilization of the Fng assay to characterize the transition state of the enzyme-substrate complex may be possible. Heavy isotope substitution of each position of the UDP-GlcNAc donor and pNP-fucose acceptor substrate and measurement of the kinetic isotope effect in the assay may allow modeling of the transition state, and subsequent development of inhibitor compounds for the enzyme (Taylor Ringia, *et al.*, 2006). While the ultimate goal would be the production of pharmaceutical compounds, the enormous benefit to Notch and Fng related biological research from a Fng inhibitor would justify this avenue of research even in the absence of any pharmacologically useful results. Specifically, Schramm and coworkers have reported that even in homologous enzymes from different species, where there is no structural difference in the active site between the two enzymes, the transition state is different (Taylor Ringia, *et al.*, 2006). They report that from these differences they are able to produce inhibitors specific to the enzyme of each species (Taylor Ringia, *et al.*, 2006). Clearly, the ability to produce Lfng, Mfng, and Rfng specific inhibitors would have the potential to open a window on a very confusing area of inquiry.

References

Acar, M, Jafar-Nejad, H, Takeuchi, H, Rajan, A, Ibrani, D, Rana, NA, Pan, H, Haltiwanger, RS and Bellen, HJ (2008) *Cell* **132** (2): 247-58

Adams, JC and Lawler, J (2004) *Int J Biochem Cell Biol* **36** (6): 961-8

Adams, JC and Tucker, RP (2000) *Develop. Dynamics* **218** (2): 280-99.

Andrade, RP, Palmeirim, I and Bajanca, F (2007) *Birth Defects Res C Embryo Today* **81** (2): 65-83

Appella, E, Weber, IT and Blasi, F (1988) *FEBS Lett* **231** (1): 1 -4

Arboleda-Velasquez, JF, Rampal, R, Fung, E, Darland, DC, Liu, M, Martinez, MC, Donahue, C P, N avarro-Gonzalez, MF, Libby, P, D'Amore, PA, Aikawa, M, Haltiwanger, RS and Kosik, KS (2005) *Hum Mol Genet* **14** (12): 1631-9

Arnold, U, Schierhorn, A and Ulbrich-hofmann, R (1999) *Eur J Biochem* **259** (1-2): 470-5

Bamdad, M, Volle, D, Dastugue, B and Meiniel, A (2004) *Cell Tissue Res* **315** (1): 15-25

Bieth, JG (1995) *Methods Enzymol* **248:** 59-84

Born, H, Peters, A and Ettinger, R (2008) *Pediatr Blood Cancer* **50** (5): 956-7

Bray, SJ (2006) *Nat Rev Mol Cell Biol* **7** (9): 678-89

Breton, C, Snajdrova, L, Jeanneau, C, Koca, J and Imberty, A (2006) *Glycobiology* **16** (2): 29R-37R

Bruckner, K, Perez, L, Clausen, H and Cohen, S (2000) *Nature* **406** (6794): 411-5

Bulman, MP, Kusumi, K, Frayling, TM, McKeown, C, Garrett, C, Lander, ES, Krumlauf, R, Hattersley, AT, Ellard, S and Turnpenny, PD (2000) *Nat Genet* **24** (4): 438-41

Chen, J, Kang, L and Zhang, N (2005) *Genesis* **43** (4): 196-204

Chen, J, Lu, L, Shi, S and Stanley, P (2006) *Gene Expr Patterns* **6** (4): 376-82

Chen, J, Moloney, DJ and Stanley, P (2001) *Proc.Natl.Acad.Sci. USA* **98** (24): 13716-21

Chigira, Y, Oka, T, Okajima, T and Jigami, Y (2008) *Glycobiology* **18** (4): 303-14

Choi, Y, Lee, JH, Hwang, S, Kim, JK, Jeong, K and Jung, S (2008) *Biopolymers* **89** (2): 114-23

Cole, SE, Levorse, JM, Tilghman, SM and Vogt, TF (2002) *Dev Cell* **3** (1): 75-84

Correia, T, Papayannopoulos, V, Panin, V, Woronoff, P, Jiang, J, Vogt, TF and Irvine, KD (2003) *Proc Natl Acad Sci U S A* **100** (11): 6404-9

Corzana, F, Busto, JH, Engelsen, SB, Jimenez-Barbero, J, Asensio, JL, Peregrina, JM and Avenoza, A (2006) *Chemistry* **12** (30): 7864-71

Coutinho, PM, Deleury, E, Davies, GJ and Henrissat, B (2003) *J Mol Biol* **328** (2): 307-17

Dale, JK, Maroto, M, Dequeant, ML, Malapert, P, McGrew, M and Pourquie, O (2003) *Nature* **421** (6920): 275-8

Davies, GJ, Gloster, TM and Henrissat, B (2005) *Curr Opin Struct Biol* **15** (6): 637-45

Davis, CG (1990) *New Biol* **2** (5): 410-9

De Celis, JF and Bray, SJ (2000) *Development* **127** (6): 1291-302

De Strooper, B, Annaert, W, Cupers, P, Saftig, P, Craessaerts, K, Mumm, JS, Schroeter, EH, Schrijvers, V, Wolfe, MS, Ray, WJ, Goate, A and Kopan, R (1999) *Nature* **398** (6727): 518-22
Doucey, MA, Hess, D, Blommers, MJ and Hofsteenge, J (1999) *Glycobiology* **9** (5): 435-41
Doucey, MA, Hess, D, Cacan, R and Hofsteenge, J (1998) *Molecular Biology of the Cell* **9** (2): 291-300
Eldadah, ZA, Hamosh, A, Biery, NJ, Montgomery, RA, Duke, M, Elkins, R and Dietz, HC (2001) *Hum Mol Genet* **10** (2): 163-9.
Evrard, YA, Lun, Y, Aulehla, A, Gan, L and Johnson, RL (1998) *Nature* **394** (6691): 377-81
Fleming, RJ, Gu, Y and Hukriede, NA (1997) *Development* **124** (15): 2973-81
Forsberg, H, Crozet, F and Brown, NA (1998) *Curr Biol* **8** (18): 1027-30
Freeze, HH and Aebi, M (2005) *Curr Opin Struct Biol* **15** (5): 490-8
Furmanek, A, Hess, D, Rogniaux, H and Hofsteenge, J (2003) *Biochemistry* **42** (28): 8452-8
Gastinel, LN, Cambillau, C and Bourne, Y (1999) *EMBO J* **18** (13): 3546-57
Ge, C and Stanley, P (2008) *Proc Natl Acad Sci U S A* **105** (5): 1539-44
Gordon, WR, Vardar-Ulu, D, Histen, G, Sanchez-Irizarry, C, Aster, JC and Blacklow, SC (2007) *Nat Struct Mol Biol* **14** (4): 295-300
Hambleton, S, Valeyev, NV, Muranyi, A, Knott, V, Werner, JM, McMichael, AJ, Handford, PA and Downing, AK (2004) *Structure (Camb)* **12** (12): 2173-83
Hase, S, Nishimura, H, Kawabata, S, Iwanaga, S and Ikenaka, T (1990) *J.Biol.Chem.* **265** (4): 1858-61
Hess, D, Keusch, JJ, Oberstein, SA, Hennekam, RC and Hofsteenge, J (2008) *J Biol Chem* **283** (12): 7354-60
Hicks, C, Johnston, SH, DiSibio, G, Collazo, A, Vogt, TF and Weinmaster, G (2000) *Nature Cell Biology* **2** (8): 515-20
Hilton, DJ, Watowich, SS, Katz, L and Lodish, HF (1996) *J Biol Chem* **271** (9): 4699-708
Hofmann, K, Bucher, P, Falquet, L and Bairoch, A (1999) *Nucleic Acids Res* **27** (1): 215-9
Hofsteenge, J, Huwiler, KG, Macek, B, Hess, D, Lawler, J, Mosher, DF and Peter-Katalinic, J (2001) *J.Biol.Chem.* **276** (9): 6485-98
Hofsteenge, J, Müller, DR, De Beer, T, Löffler, A, Richter, WJ and Vliegenthart, JFG (1994) *Biochemistry* **33** (46): 13524-30
Imperiali, B and Rickert, KW (1995) *Proc.Natl.Acad.Sci.USA* **92** (1): 97-101
Ioffe, E and Stanley, P (1994) *Proc.Natl.Acad.Sci.USA.* **91** (2): 728-32
Iruela-Arispe, ML, Lombardo, M, Krutzsch, HC, Lawler, J and Roberts, DD (1999) *Circulation* **100** (13): 1423-31
Irvine, KD (1999) *Curr.Opin.Genet.Dev.* **9** (4): 434-41
Irvine, KD and Rauskolb, C (2001) *Annu Rev Cell Dev Biol* **17**: 189-214
Irvine, KD and Wieschaus, E (1994) *Cell* **79** (4): 595-606
Jinek, M, Chen, YW, Clausen, H, Cohen, SM and Conti, E (2006) *Nat Struct Mol Biol* **13** (10): 945-6

John, GR, Shankar, SL, Shafit-Zagardo, B, Massimi, A, Lee, SC, Raine, CS and Brosnan, CF (2002) *Nat Med* **8** (10): 1115-21
Joutel, A, Corpechot, C, Ducros, A, Vahedi, K, Chabriat, H, Mouton, P, Alamowitch, S, Domenga, V, Cecillion, M, Marechal, E, Maciazek, J, Vayssiere, C, Cruaud, C, Cabanis, E-A, Ruchoux, MM, Weissenbach, J, Bach, JF, Bousser, MG and Tournier-Lasserve, E (1996) *Nature* **383** (6602): 707-10
Jouve, C, Iimura, T and Pourquie, O (2002) *Development* **129** (5): 1107-17
Kageyama, R, Masamizu, Y and Niwa, Y (2007) *Dev Dyn* **236** (6): 1403-9
Kao, Y-H, Lee, GF, Wang, Y, Starovasnik, MA, Kelley, RF, Spellman, MW and Lerner, L (1999) *Biochemistry* **38** (22): 7097-110
Kidd, S and Lieber, T (2002) *Mech Dev* **115** (1-2): 41-51
Kikuchi, N, Kwon, YD, Gotoh, M and Narimatsu, H (2003) *Biochem Biophys Res Commun* **310** (2): 574-9
Kozma, K, Keusch, JJ, Hegemann, B, Luther, KB, Klein, D, Hess, D, Haltiwanger, RS and Hofsteenge, J (2006) *J Biol Chem* **281** (48): 36742-51
Krantz, ID, Smith, R, Colliton, RP, Tinkel, H, Zackai, EH, Piccoli, DA, Goldmuntz, E and Spinner, NB (1999) *Am J Med Genet* **84** (1): 56-60.
Kusumi, K, Sun, ES, Kerrebrock, AW, Bronson, RT, Chi, DC, Bulotsky, MS, Spencer, JB, Birren, BW, Frankel, WN and Lander, ES (1998) *Nat Genet* **19** (3): 274-8
Ladi, E, Nichols, JT, Ge, W, Miyamoto, A, Yao, C, Yang, LT, Boulter, J, Sun, YE, Kintner, C and Weinmaster, G (2005) *J Cell Biol* **170** (6): 983-92
Lambert, C, Leonard, N, De Bolle, X and Depiereux, E (2002) *Bioinformatics* **18** (9): 1250-6
Lee, NV, Sato, M, Annis, DS, Loo, JA, Wu, L, Mosher, DF and Iruela-Arispe, ML (2006) *EMBO J* **25** (22): 5270-83
Lei, L, Xu, A, Panin, VM and Irvine, KD (2003) *Development* **130** (26): 6411-21
Lesnik Oberstein, SA, Kriek, M, White, SJ, Kalf, ME, Szuhai, K, den Dunnen, JT, Breuning, MH and Hennekam, RC (2006) *Am J Hum Genet* **79** (3): 562-6
Levenberg, K (1944) *The Quarterly Review of Mathematics* **2** (2): 164-8
Li, L, Krantz, ID, Deng, Y, Genin, A, Banta, AB, Collins, CC, Qi, M, Trask, BJ, Kuo, WL, Cochran, J, Costa, T, Pierpont, MEM, Rand, EB, Piccoli, DA, Hood, L and Spinner, NB (1997) *Nature Genet.* **16** (3): 243-51
Loirat, C, Veyradier, A, Girma, JP, Ribba, AS and Meyer, D (2006) *Semin Thromb Hemost* **32** (2): 90-7
Lovell, SC, Word, JM, Richardson, JS and Richardson, DC (2000) *Proteins* **40** (3): 389-408
Luo, Y and Haltiwanger, RS (2005) *J Biol Chem* **280** (12): 11289-94
Luo, Y, Koles, K, Vorndam, W, Haltiwanger, RS and Panin, VM (2006a) *J Biol Chem* **281** (14): 9393-9
Luo, Y, Luther, K.B., Haltiwanger, R.S. (2007) *O*-Fucosylation of Glycoproteins in *Comprehensive Glycoscience* (Kamerling, JP ed.) Elsevier, Amsterdam
Luo, Y, Nita-Lazar, A and Haltiwanger, RS (2006b) *J Biol Chem* **281** (14): 9385-92
Maillette de Buy Wenniger-Prick, LJ and Hennekam, RC (2002) *Ann Genet* **45** (2): 97-103
Marquardt, DW (1963) *Journal of the Society for Industrial and Applied Mathematics* **11** (2): 431-41

McGrew, MJ, Dale, JK, Fraboulet, S and Pourquie, O (1998) *Curr Biol* **8** (17): 979-82
Meiniel, A, Meiniel, R, Goncalves-Mendes, N, Creveaux, I, Didier, R and Dastugue, B (2003) *Int Rev Cytol* **230**: 1-39
Mer, G, Hietter, H and Lefevre, JF (1996) *Nat Struct Biol* **3** (1): 45 -53
Minamida, S, Aoki, K, Natsuka, S, Omichi, K, Fukase, K, Kusumoto, S and Hase, S (1996) *J.Biochem.(Tokyo)* **120** (5): 1002-6
Moloney, DJ and Haltiwanger, RS (1999) *Glycobiology* **9** (7): 679-87
Moloney, DJ, Lin, AI and Haltiwanger, RS (1997) *J.Biol.Chem.* **272** (30): 19046-50
Moloney, DJ, Panin, VM, Johnston, SH, Chen, J, Shao, L, Wilson, R, Wang, Y, Stanley, P, Irvine, KD, Haltiwanger, RS and Vogt, TF (2000a) *Nature* **406** (6794): 369-75
Moloney, DJ, Shair, L, Lu, FM, Xia, J, Locke, R, Matta, KL and Haltiwanger, RS (2000b) *J.Biol.Chem.* **275** (13): 9604-11
Morales, AV, Yasuda, Y and Ish-Horowicz, D (2002) *Dev Cell* **3** (1): 63 -74
Morera, S, Imberty, A, Aschke-Sonnenborn, U, Ruger, W and Freemont, PS (1999) *J Mol Biol* **292** (3): 717-30
Morgan, TH (1917) *Am. Nat.* **51** (609): 513-44
Morimoto, M, Takahashi, Y, Endo, M and Saga, Y (2005) *Nature* **435** (7040): 354-9
Mumm, JS, Schroeter, EH, Saxena, MT, Griesemer, A, Tian, X, Pan, DJ, Ray, WJ and Kopan, R (2000) *Mol. Cell* **5** (2): 197-206
Munte, CE, Gade, G, Domogalla, B, Kremer, W, Kellner, R and Kalbitzer, HR (2008) *FEBS J* **275** (6): 1163-73
Muroi, E, Manabe, S, Ikezaki, M, Urata, Y, Sato, S, Kondo, T, Ito, Y and Ihara, Y (2007) *Glycobiology* **17** (9): 1015-28
Nita-Lazar, A and Haltiwanger, RS (2006a) *Methods Enzymol* **417**: 93-111
Nita-Lazar, A and Haltiwanger, RS (2006b) *Methods Mol Biol* **347**: 57-68
Novikoff, AB and Goldfischer, S (1961) *Proc Natl Acad Sci U S A* **47** (6): 802-10
Okajima, T and Irvine, KD (2002) *Cell* **111** (6): 893-904
Okajima, T, Reddy, B, Matsuda, T and Irvine, KD (2008) *BMC Biol* **6**: 1
Okajima, T, Xu, A, Lei, L and Irvine, KD (2005) *Science* **307** (5715): 1599-603
Omichi, K, Aoki, K, Minamida, S and Hase, S (1997) *Eur.J.Biochem.* **245** (1): 143-6
Panin, VM, Papayannopoulos, V, Wilson, R and Irvine, KD (1997) *Nature* **387** (6636): 908-12
Parks, AL, Klueg, KM, Stout, JR and Muskavitch, MA (2000) *Development* **127** (7): 1373-85.
Pedersen, LC, Tsuchida, K, Kitagawa, H, Sugahara, K, Darden, TA and Negishi, M (2000) *J Biol Chem* **275** (44): 34580-5
Pei, Z and Baker, NE (2008) *BMC Dev Biol* **8**: 4
Perez, L, Milan, M, Bray, S and Cohen, SM (2005) *Mech Dev* **122** (4): 479-86
Qasba, PK, Ramakrishnan, B and Boeggeman, E (2005) *Trends Biochem Sci* **30** (1): 53-62
Radtke, F, Wilson, A and MacDonald, HR (2004) *Curr Opin Immunol* **16** (2): 174-9
Ramakrishnan, B, Boeggeman, E, Ramasamy, V and Qasba, PK (2004) *Curr Opin Struct Biol* **14** (5): 593-600
Rampal, R, Arboleda-Velasquez, JF, Nita-Lazar, A, Kosik, KS and Haltiwanger, RS (2005a) *J Biol Chem* **280** (37): 32133-40

Rampal, R, Li, AS, Moloney, DJ, Georgiou, SA, Luther, KB, Nita-Lazar, A and Haltiwanger, RS (2005b) *J Biol Chem* **280** (51): 42454-63
Rampal, R, Luther, KB and Haltiwanger, RS (2007) *Curr Mol Med* **7** (4): 427-45
Rao, Z, Handford, P, Mayhew, M, Knott, V, Brownlee, GG and Stuart, D (1995) *Cell* **82** (1): 131-41
Ricketts, LM, Dlugosz, M, Luther, KB, Haltiwanger, RS and Majerus, EM (2007) *J Biol Chem* **282** (23): 17014-23
Ripka, J, Adamany, A and Stanley, P (1986) *Arch.Biochem.Biophys.* **249** (2): 533-45
Ritchie, DW (2003) *Proteins* **52** (1): 98 -106
Sasaki, N, Sasamura, T, Ishikawa, HO, Kanai, M, Ueda, R, Saigo, K and Matsuno, K (2007) *Genes Cells* **12** (1): 89-103
Sasamura, T, Ishikawa, HO, Sasaki, N, Higashi, S, Kanai, M, Nakao, S, Ayukawa, T, Aigaki, T, Noda, K, Miyoshi, E, Taniguchi, N and Matsuno, K (2007) *Development* **134** (7): 1347-56
Sasamura, T, Sasaki, N, Miyashita, F, Nakao, S, Ishikawa, HO, Ito, M, Kitagawa, M, Harigaya, K, Spana, E, Bilder, D, Perrimon, N and Matsuno, K (2003) *Development* **130** (20): 4785-95
Sato, T, Sato, M, Kiyohara, K, Sogabe, M, Shikanai, T, Kikuchi, N, Togayachi, A, Ishida, H, Ito, H, Kameyama, A, Gotoh, M and Narimatsu, H (2006) *Glycobiology* **16** (12): 1194-206
Schneppenheim, R, Budde, U, Oyen, F, Angerhaus, D, Aumann, V, Drewke, E, Hassenpflug, W, Haberle, J, Kentouche, K, Kohne, E, Kurnik, K, Mueller-Wiefel, D, Obser, T, Santer, R and Sykora, KW (2003) *Blood* **101** (5): 1845-50
Schwede, T, Kopp, J, Guex, N and Peitsch, MC (2003) *Nucleic Acids Res* **31** (13): 3381-5
Shao, L, Moloney, DJ and Haltiwanger, R (2003) *J Biol Chem* **278** (10): 7775-82
Shi, S and Stanley, P (2003) *Proc.Natl.Acad.Sci.USA* **100** (9): 5234-9
Shi, S and Stanley, P (2006) *Cell Cycle* **5** (3): 274-8
Shimizu, K, Chiba, S, Hosoya, N, Kumano, K, Saito, T, Kurokawa, M, Kanda, Y, Hamada, Y and Hirai, H (2000a) *Mol Cell Biol* **20** (18): 6913-22
Shimizu, K, Chiba, S, Kumano, K, Hosoya, N, Takahashi, T, Kanda, Y, Hamada, Y, Yazaki, Y and Hirai, H (1999) *J.Biol.Chem.* **274** (46): 32961-9
Shimizu, K, Chiba, S, Saito, T, Kumano, K and Hirai, H (2000b) *Biochem Biophys Res Commun* **276** (1): 385-9
Shimizu, K, Chiba, S, Saito, T, Kumano, K, Takahashi, T and Hirai, H (2001) *J Biol Chem* **276** (28): 25753-8.
Silverstein, RL and Febbraio, M (2007) *Curr Pharm Des* **13** (35): 3559-67
Sinnott, MJ (1990) *Chemical Reviews* **90** (7): 1171-202
Sparrow, DB, Chapman, G, Turnpenny, PD and Dunwoodie, SL (2007) *Birth Defects Res C Embryo Today* **81** (2): 93-110
Sparrow, DB, Chapman, G, Wouters, MA, Whittock, NV, Ellard, S, Fatkin, D, Turnpenny, PD, Kusumi, K, Sillence, D and Dunwoodie, SL (2006) *Am J Hum Genet* **78** (1): 28-37
Stahl, M, Uemura, K, Ge, C, Shi, S, Tashima, Y and Stanley, P (2008a) *J Biol Chem* **283** (20): 13638-51
Stahl, M, Uemura, K, Ge, C, Shi, S, Tashima, Y and Stanley, P (2008b) *J Biol Chem*

Stanley, P (2007) *Curr Opin Struct Biol* **17** (5): 530-5
Stanley, P and Chaney, W (1985) *Mol Cell Biol* **5** (6): 1204-11
Tan, K, Duquette, M, Liu, JH, Dong, Y, Zhang, R, Joachimiak, A, Lawler, J and Wang, JH (2002) *J Cell Biol* **159** (2): 373-82
Taylor Ringia, EA, Tyler, PC, Evans, GB, Furneaux, RH, Murkin, AS and Schramm, VL (2006) *J Am Chem Soc* **128** (22): 7126-7
Thompson, JD, Higgins, DG and Gibson, TJ (1994) *Nucleic Acids Res* **22** (22): 4673-80
Tolsma, SS, Volpert, OV, Good, DJ, Frazier, WA, Polverini, PJ and Bouck, N (1993) *J Cell Biol* **122** (2): 497-511
Tossavainen, H, Pihlajamaa, T, Huttunen, TK, Raulo, E, Rauvala, H, Permi, P and Kilpelainen, I (2006) *Protein Sci* **15** (7): 1760-8
Tucker, RP (2004) *Int J Biochem Cell Biol* **36** (6): 969-74
Unligil, UM and Rini, JM (2000) *Curr Opin Struct Biol* **10** (5): 510-7
Unligil, UM, Zhou, S, Yuwaraj, S, Sarkar, M, Schachter, H and Rini, JM (2000) *EMBO J* **19** (20): 5269-80
Varki, A (1993) *Glycobiology* **3** (2): 97-130
Vrielink, A, Ruger, W, Driessen, HP and Freemont, PS (1994) *EMBO J* **13** (15): 3413-22
Wacker, M, Linton, D, Hitchen, PG, Nita-Lazar, M, Haslam, SM, North, SJ, Panico, M, Morris, HR, Dell, A, Wren, BW and Aebi, M (2002) *Science* **298** (5599): 1790-3
Wang, LW, Dlugosz, M, Somerville, RP, Raed, M, Haltiwanger, RS and Apte, SS (2007) *J Biol Chem* **282** (23): 17024-31
Wang, Y, Shao, L, Shi, S, Harris, RJ, Spellman, MW, Stanley, P and Haltiwanger, RS (2001) *J.Biol.Chem.* **276** (43): 40338-45
Wharton, KA, Johansen, KM, Xu, T and Artavanis-Tsakonas, S (1985) *Cell* **43** (3 Pt. 2): 567-81
Whittock, NV, Sparrow, DB, Wouters, MA, Sillence, D, Ellard, S, Dunwoodie, SL and Turnpenny, PD (2004) *Am J Hum Genet* **74** (6): 1249-54
Wiggins, CA and Munro, S (1998) *Proc Natl Acad Sci U S A* **95** (14): 7945-50
Wyatt, R, Kwong, PD, Desjardins, E, Sweet, RW, Robinson, J, Hendrickson, WA and Sodroski, JG (1998) *Nature* **393** (6686): 705-11
Wyss, DF, Choi, JS, Li, J, Knoppers, MH, Willis, KJ, Arulanandam, AR, Smolyar, A, Reinherz, EL and Wagner, G (1995) *Science* **269** (5228): 1273-8
Xu, A, Haines, N, Dlugosz, M, Rana, NA, Takeuchi, H, Haltiwanger, RS and Irvine, KD (2007) *J Biol Chem* **282** (48): 35153-62
Xu, A, Lei, L and Irvine, KD (2005) *J Biol Chem* **280** (34): 30158-65
Yang, LT, Nichols, JT, Yao, C, Manilay, JO, Robey, EA and Weinmaster, G (2005) *Mol Biol Cell* **16** (2): 927-42
Young, NM, Brisson, JR, Kelly, J, Watson, DC, Tessier, L, Lanthier, PH, Jarrell, HC, Cadotte, N, St Michael, F, Aberg, E and Szymanski, CM (2002) *J Biol Chem* **277** (45): 42530-9
Yuan, Y, Barrett, D, Zhang, Y, Kahne, D, Sliz, P and Walker, S (2007) *Proc Natl Acad Sci U S A* **104** (13): 5348-53
Zhang, N and Gridley, T (1998) *Nature* **394** (6691): 374-7
Zhang, N, Norton, CR and Gridley, T (2002) *Genesis* **33** (1): 21-8
Zhu, H and Dhar, PK (2006) *IEEE Trans Nanobioscience* **5** (3): 193-203