O-GLYCOSYLATION OF THROMBOSPONDIN TYPE 1 REPEATS

Identification and Characterisation of β1,3-Glucosyltransferase

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

von

KRISZTINA KEUSCH

aus Dunaújváros, Ungarn

Dissertationsleiter: Prof. Dr. Jan Hofsteenge

Friedrich Miescher Institute for Biomedical Research

BASEL, 2008

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr Jan Hofsteenge

PD. Dr. Jack Rohrer

Prof. Dr. Fred Meins

Prof. Dr. Ruth Chiquet-Ehrismann

Basel, den 24.06.2008

Prof. Dr. Hans-Peter Hauri

Dekan/Dekanin

“A good idea is about 10 percent implementation and hard work, and luck is 90 percent.”

Guy Kawasaki

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr Jan Hofsteenge

PD. Dr. Jack Rohrer

Prof. Dr. Fred Meins

Prof. Dr. Ruth Chiquet-Ehrismann

Basel, den 24.06.2008

Prof. Dr. Hans-Peter Hauri

Dekan/Dekanin

1. ACKNOWLEDGEMENTS

First and foremost I would like to express my thanks to my supervisor, Dr. Jan Hofsteenge, who gave me the opportunity to discover the fascinating field of glycobiology. He supported me throughout my PhD and his door was always open for discussion of any subject.

My appreciation goes to my thesis committee members, Prof. Frederick Meins and Dr. Jack Rohrer for their advice during my studies, the time they invested in my thesis committee meetings and still need to invest in evaluating this thesis.

I would like to thank the former and current members of the Hofsteenge lab, Björn, Pascal, Jasmin, Alexandra, Guillaume, Florence, Carsten, Chun-I, Claudia, Matthias, Dominique, Stefano, Constanze, Nomio, Ivana and Anne. I enjoyed working and discussing with them all. I especially enjoyed sharing their home-made cakes from the unofficial “Hofsteenge Lab Protocols Book” and walking off the excess calories on our Swiss lab hikes. Warm thanks go to Marianne, our former technician for her great help and guidance. I really appreciated her big morning smiles, positive thinking and enthusiasm which made an excellent working atmosphere in lab 310. Thanks go the members of the Protein Analysis Facility, Ragna, Reto and Daniel for showing me the tricks of how to work with different mass analysing instruments.

A very big thank from my heart goes to my husband and colleague, Dr. Jeremy Keusch, who showed me how to be passionate about science with his constant enthusiasm and unbreakable curiosity. Being a wonderful and generous person he has supported and guided me with all his scientific skills and knowledge during my PhD. I also would like to thank him for his suggestions during writing up and the critical reading of this thesis.

My deepest love goes to Zoe, my daughter, the person who has improved my multitasking skills immensely.

Last but not least, my thanks go to my parents for their heartfelt encouragement during my studies and because they have been supporting me throughout my life. They always assured me that I made the right choice even though this decision took me more than 1000 km away from them. Köszönöm Anya és Apa, hogy mellettem álltatok, támogattatok és bátorítottatok nemcsak a PhD munkám során, de minden döntésemben, még akkor is, ha ezt több mint ezer kilométer távolságból kellett megtennetek.

2. ABBREVIATIONS

ADAM	a disintegrin and metalloproteinase
ADAMTS	a disintegrin and metalloproteinase with thrombospondin type 1 repeats
AspN	endoproteinase from *Pseudomonas fragi* mutant
B3GALTL	*β*1,3-galactosyltransferase-like
CADASIL	cerebrovascular dementia
CAZy	Carbohydrate-Active enZYmes database
CDG	Congenital Disorders of Glycosylation
Cer	ceramide
CRD	carbohydrate-recognition domains
CS	chondroitin sulphate
CSP	circum-sporozoite protein
ECL	enhanced chemiluminescence
ECM	extracellular matrix
EGF	epidermal growth factor-like module
Endo H_f	endoglycosidase H_f
ER	endoplasmic reticulum
ERAD	ER-associated degradation
ERGIC	ER-Golgi intermediate compartment
ERT	enzyme replacement therapy
FGF-2	fibroblast growth factor-2
FGFR	fibroblast growth factor receptor
FPLC	fast protein liquid chromatography
FRET	fluorescence energy resonance transfer
GAG	glycosaminoglycan
GHs	glycosidases
GlcNAc-P	*N*-acetylglucosamine-phosphate
GnT-V	GlcNAc-transferase-V
GPI	glycosylphosphatidylinositol
GTG-DEEP	Global Trace Graph
GTs	glycosyltransferases
HA	hyaluronan
HEK293T	human embryonic kidney cells with SV40 large T-antigen
hLys	hydroxylysine
HPLC	high performance liquid chromatography
IEF	isoelectic focusing
IP	immunoprecipitation
k_{cat}	turnover number
KS	keratan sulphate
LADII	leukocyte adhesion deficiency II
Lfng	lunatic Fringe
LLO	lipid-linked oligosaccharide
LSD	lysosomal storage diseases
LysC	endoproteinase from *Achromobacter lyticus*
m/z	mass over charge
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
Mfng	manic Fringe
MPI	phosphomannose-isomerase
MS	mass spectrometry
OFUT1	*O*-fucosyltransferase 1
OFUT2	*O*-fucosyltransferase 2
OGT	*O*-GlcNAc transferase
OST	oligosaccharyltransferase
PI	phosphatidylinositol
PIC	protease inhibitor cocktail
PNGase F	peptide-*N*-glycosidase F
POFUT1	protein *O*-fucosyltransferase 1
POFUT2	protein *O*-fucosyltransferase 2
ppGalNAcT	polypeptide *N*-acetlygalactosaminyltransferase
PPS	Peters Plus syndrome
Rfng	radical Fringe
Rspo	R-spondin
Tf	transferrin
TGN	trans-Golgi network
TRAP	thrombospondin-related anonymous protein
TRSP	thrombospondin-related sporozoites protein
TSP	thrombospondin
TSP-1	thrombospondin-1
TSR	thrombospondin type 1 repeat
TSR4	4th TSR from F-spondin
uPA	urinary-type plasminogen activator
*β*3Glc-T	*β*1,3-glucosyltransferase
*β*4GalT	*β*1,4-galactosyltransferse

Standard abbreviations of amino acids

Three letter code	One letter code	Amino acid
Ala	A	alanine
Arg	R	arginine
Asn	N	asparagine
Asp	D	aspartate
Cys	C	cysteine
Gln	Q	glutamine
Glu	E	glutamate
Gly	G	glycine
His	H	histidine
Ile	I	isoleucine
Leu	L	leucine
Lys	K	lysine
Met	M	methionine
Phe	F	phenylalanine
Pro	P	proline
Ser	S	serine
Thr	T	threonine
Trp	W	tryptophan
Tyr	Y	tyrosine
Val	V	valine

Standard abbreviations of sugars

Fuc	Fucose
Glc	Glucose
GlcNAc	*N*-acetylglucosamine
Gal	Galactose
GalNAc	*N*-acetylgalactosamine
Xyl	Xylose
Man	Mannose
Dol-P-Man	Dolichyl Phosphomannose
Dol-P-Glc	Dolichyl Phosphoglucose
Dol-P-Glc	Dolichyl Phosphoglucose

3. ABSTRACT

Rare types of protein glycosylation often occur in a domain-specific manner required for important biological functions (Okajima et al. 2008). *O*-linked fucosylation, where the fucose is linked directly to the hydroxyl groups of serine or threonine residues, has so far only been reported on three distinct protein modules: epidermal growth factor (EGF)-like domains; thrombospondin type 1 repeats (TSRs); and on the protease inhibitor PMP-C. The function of *O*-fucosylation is well studied on EGFs, where changes in the *O*-fucose glycan on Notch EGFs alter Notch signalling (reviewed in Okajima et al. 2008).

TSRs are modified by the unusual Glc-*β*1,3-Fuc-*O*- disaccharide (Hofsteenge et al. 2001; Gonzalez de Peredo et al. 2002). Protein O-fucosyltransferase 2 (POFUT2) initiates the fucosylation of properly folded TSRs. Deletion of *POFUT2* in *C. elegans* causes an early dorsal migration phenotype of the anterior distal tip cell (Canevascini et. al, manuscript in preparation). Recently it was shown that *Pofut2* null embryos die by mid-gastrulation (Du et al. 2007). Further, *O*-fucosylation of TSRs in ADAMTS13 (Ricketts et al. 2007) and ADAMTS-like 1/punctin (Wang et al. 2007) is required for their secretion in mammalian cells. Since these experiments abolished the fucosylated glycan, Glc-*β*1,3-Fuc-*O*-, it remains to be determined what role the glucose plays. Therefore, the *β1,*3-glucosyltransferase (*β*3Glc-T) that catalyses the last step in the formation of this disaccharide, needed to be identified and characterised.

In order to identify the *β*3Glc-T a relevant enzyme substrate in the form of properly folded TSR-fucose needed to be produced. I describe a reproducible purification strategy for the production of a large amount of highly pure, correctly folded TSR and fucosylated TSR domain (Chapter 6.1). Using this TSR-fucose substrate in a specific radiochemical assay, I was able to identify the *β*3Glc-T as a member of the glycosyltransferase family GT31, which includes Fringe, the enzyme that modifies *O*-fucose in EGF repeats. The cloned *β*3Glc-T specifically catalyses the transfer of glucose from UDP-glucose to TSR-fucose to yield Glc-*β*1,3-Fuc-*O*- (Chapter 6.2). There is no reactivity towards non-fucosylated substrates or fucosylated EGF. The *β*3Glc-T protein consists of an N- and C-terminal domain of approximately equal size that share internal sequence homology. Within the C-terminal domain putative catalytic centre, several key residues were identified (Asp-349, Asp-351, Asp-421) for *β*3Glc-T activity (Chapters 6.2 and 6.3). In addition, the removal of either of the two *N*-glycosylation sites in *β*3Glc-T resulted in diminished activity (Chapter 6.3). The *β*3Glc-T N-terminal domain does not possess intrinsic *β*3Glc-T activity but affects its activity. Expression of the *β*3Glc-T C-terminal domain requires the co-expression of the *β*3Glc-T N-terminal domain. I suggest that the *β*3Glc-T N-terminal domain may function as a stabiliser or internal chaperone (Chapter 6.4). Moreover, it was recently shown that mutations in the gene encoding *β*3Glc-T cause the severe developmental disease, Peters Plus syndrome (PPS) (Lesnik Oberstein et al. 2006). Constructs encoding these mutated *β*3Glc-T sequences express truncated proteins lacking the catalytic domain (Chapter 6.4). This raises the possibility that the truncated *β*3Glc-T proteins may play a role in PPS.

The work presented in this thesis provides the basis for further studies on the role of *β*3Glc-T in regulating TSR biological function, through glycosylation, in health and disease.

4. RESEARCH OBJECTIVE

(1) We would like to understand the biological function of the Glc-*β*1,3-Fuc-*O*- disaccharide structure in TSRs. For this my aim was to identify the TSR-fucose-specific *β*1,3-glucosyltransferase (*β*3Glc-T) and further characterise it at the molecular level.

(2) Subsequently, we also planned to study the *β*3Glc-T by genetic manipulation in the multicellular model organism *C. elegans*. This strategy altered with the finding that mutations in the human *β*3Glc-T cause a new congenital disorder of glycosylation, Peters Plus syndrome. I then focused on understanding the function of the *β*3Glc-T N- and C-terminal domains with reference to the truncating mutations found in PPS.

5. General Introduction

5.1. The Concept of Glycobiology

Glycosylation consists of the attachment of sugars to other sugars or aglycans such as proteins and lipids. There are nine different sugars in mammals that may be linked in several positions and stereoisomeric configurations, creating a huge variety of glycans. In addition, there are four main classes of protein glycosylation based upon the core linkage of the carbohydrate to the protein. Greater than 3% of the human genome is involved in glycosylation processes resulting in more than 50% of proteins expressed being glycosylated (Apweiler et al. 1999; Davies et al. 2005). Clearly this modification generates a very complex proteome.

Glycobiology is the study of the structure, biosynthesis and biology of carbohydrates. Glycosylation is conserved from bacteria to man. Glycobiology is one of the most rapidly developing fields in life sciences. The three main molecular players of glycobiology: glycans, lectins, and carbohydrate active enzymes (glycosyltransferases and glycosidases), are functionally closely connected with each other (Figure 1). Their interactions are time- and location dependent.

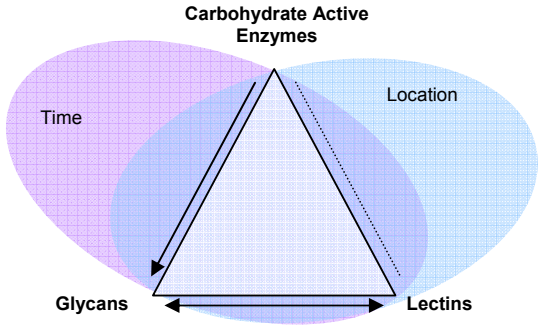


Figure 1. The players of glycobiology

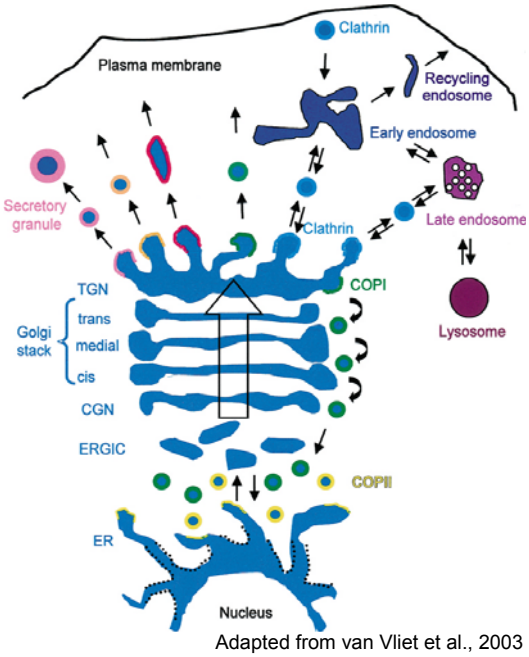
Glycans are synthesised by large families of carbohydrate active enzymes. Glycosyltransferases are involved in the covalent attachment of sugars to other sugars, proteins or lipids, whilst the glycosidases interact at specific points in this assembly process to trim off sugar residues and remodel the developing glycan structure. The lectins, which are glycan-binding proteins, recognise particular glycan patterns and help decode the glycan structure into a biological effect.

The primary structure of the first glycan was determined in 1958 (Meyer 1958), however the interest in glycobiology only started to grow rapidly from the late 1980s. Today it is evident that carbohydrates have an important biological role in processes including cell-cell interaction, protein trafficking, adhesion, signalling, fertilisation and immunological recognition. Several human diseases, including haematological disorders, some inflammatory skin diseases, diabetes, several oncogenic transformations and inherited glycosylation disorders are known to involve changes in glycosylation or glycan recognition (Varki et al. 1999).

In contrast to genomics and proteomics, where many of the research tools have been available for the scientists, the development of analytical and biochemical technologies to investigate the glycan structure-function relationship has presented great challenges. The reasons are the following: first, in contrast to protein synthesis, the assembly of glycan structures is not a template-driven process, but involves consecutive enzymatic reactions. Moreover, some of these carbohydrate active enzymes have tissue-specific isoforms. Second, since no proofreading machinery is present and there is competition for the glycosylated substrate, heterogeneity and great diversity of different glycan structures (glycoforms) appear. Thus, the presence of the same glycoform in different proteins and different glycoforms on the same protein, even at the same site, make it difficult to establish structure-function relationships. Finally, understanding the biochemical basis of the interaction between glycans and lectins is also a demanding task (Raman et al. 2005).

5.2. The Three Players of Glycobiology

In eukaryotic cells many important cellular processes occur in the different compartments of the secretory pathway. Newly synthesised proteins destined for the endoplasmic reticulum (ER), Golgi and lysosome, as well as the secretory and plasma membrane proteins are first targeted to the rough ER. The nascent polypeptides contain an amino-terminal hydrophobic signal sequence which initiates their transport into the ER lumen via the translocon. Within the ER, the folding of the proteins is checked and the properly folded ones are transported to the Golgi apparatus by vesicular trafficking. The secretory, lysosomal and plasma membrane proteins are transported through the Golgi apparatus and sorted according to their final destination at the trans-Golgi network (Figure 2). Several steps within the secretory pathway are modulated by the *N*-glycans on the transported glycoproteins.



Adapted from van Vliet et al., 2003

Figure 2. Schematic representation of the secretory pathway

The ER is the entry point of the secretory pathway. Here nascent proteins are correctly folded and exit the ER in COPII-coated transport vesicles. Fusion of these vesicles forms the ER-Golgi intermediate compartment (ERGIC). Anterograde cargo moves through the Golgi stack and the sorting takes place at the trans-Golgi network (TGN). ER proteins which leaked out of the ER to the ERGIC and Golgi stacks are recycled by COPI coated vesicles.

In the following sections I will discuss some of the ways glycans control these and other events. In addition, I will outline how the carbohydrate active enzymes and lectins are involved in different biological processes.

Glycans

The glycome, the entire repertoire of glycans in humans, is much greater in size than the number of proteins encoded by the genome (Freeze 2006). Glycans are formed on protein, lipid or saccharide acceptors by glycosyltransferases and glycosidases as they pass through the secretory pathway (Figure 2). The glycans can vary from simple monosaccharides to highly branched and complex glycan structures.

In mammals eleven activated sugar donors and several amino acids, using two anomeric configurations of glycosidic bonds, participate in different combinations to build up a large variety of glycoconjugates (Table 1).

		Sugar Donors										
		GDP-fucose	UDP-galactose	UDP-GalNAc	UDP-Glucose	UDP-GlcNAc	UDP-Glucuronic acid	GDP-mannose	CMP-sialic acid	UDP-xylose	Dol-P-mannose	Dol-P-glucose
Protein and Lipid Acceptors	Ser/Thr	α1	-	α1	β1	β1	-	-	-	β1	α1	β1
	Asn	-	-	-	β1	$	-	-	-	-	-	-
	hLys	-	β1	-	-	-	-	-	-	-	-	-
	Trp	-	-	-	-	-	-	-	-	-	α1	-
	Tyr	-	-	-	α1	-	-	-	-	-	-	-
	Ceramide	-	β1	-	β1	-	-	-	-	-	-	-
	PI	-	-	-	-	α1	-	-	-	-	-	-
Saccharide Acceptors	Fucose	-	-	-	β1-3	β1-3	-	-	-	-	-	-
	Galactose	α1-2	α1-3 α1-4 β1-3	α1-3 β1-3 β1-4	α1-2	β1-3 β1-6	β1-3 β1-4	-	α2-3 α2-6	-	-	-
	GalNAc	-	β1-3	α1-3 α1-6	-	β1-6	β1-3	-	α2-6	-	-	-
	Glucose	-	β1-4	-	-	-	-	-	-	α1-3	-	α1-2 α1-3
	GlcNAc	α1-3 α1-4 α1-6	β1-3 β1-4	β1-4	-	α1-6 β1-4	β1-3 β1-4	β1-4	-	-	α1-4	-
	Glucuronic acid	-	-	β1-4	-	α1-4 β1-6	-	-	-	-	-	-
	Mannose	-	-	-	-	β1-2	-	α1-2 α1-3 α1-6	-	-	α1-2 α1-3 α1-6	α1-3
	Sialic acid	-	-	-	-	-	-	-	α2-8	-	-	-
	Xylose	-	β1-4	-	-	-	-	-	-	α1-3	-	

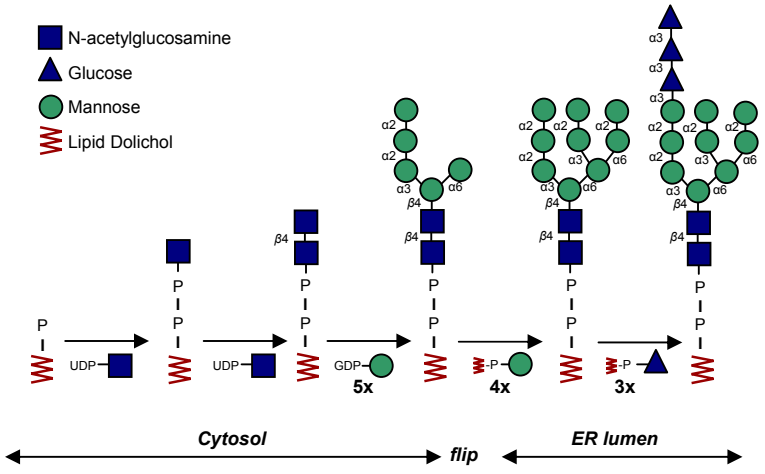
Table 1. Mammalian glycans

In mammals eleven activated sugar donors and multiple protein and lipid acceptors exist, which produce 15 glycans in *α* or *β* configurations linked at a number of positions of the donor sugar ring (coloured in yellow). The attached monosaccharide often becomes a saccharide acceptor in 1 of 52 other glycosyltransferase reactions (depicted in blue). This results in glycosidic bonds with *α* or *β* configurations of the donor saccharide linked through position 1 or 2 to position 2, 3, 4, 6 or 8 of the saccharide acceptor. The *en bloc* attachment of the *N*-linked glycans is marked with a "$" sign. Dol-P, dolichyl phosphate; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; Ser/Thr, serine/threonine; Asn, asparagine; hLys, hydroxylysine; Trp, tryptophan; Tyr, tyrosine; PI, phosphatidylinositol. The table is modified from Ohtsubo et al, 2006.

Due to these permutations the information content of the glycans is enormous (Spiro 2002; Ohtsubo et al. 2006). There are four main classes of protein glycosylation according to the attachment site of the carbohydrate chain: *N*-glycosylation, *O*-glycosylation, glycosyl phosphatidylinositol-anchored glycosylation and the less common *C*-glycosylation.

N-glycans

The biosynthesis of *N*-glycans is the most extensive and complicated type of glycan assembly. It can be divided into three consecutive phases: the step-wise assembly of the 14-sugar containing lipid-linked oligosaccharide (LLO) precursor (Figure 3), the *en bloc* transfer of the precursor to the amide group of the asparagine in the consensus sequence Asn-X-Ser/Thr (where X can be any amino acid except proline) of the nascent protein, and the processing of the *N*-linked glycan (Kornfeld et al. 1985). This complex processing is initiated in the ER, continues in the *cis* , *medial* and *trans* Golgi, and through into the *trans*-Golgi network (Figure 2) (Kornfeld et al. 1985).



Modified from Varki et al., 1999

Figure 3. Biosynthesis of the LLO precursor

The assembly of the LLO precursor occurs in two locations (Figure 3). On the cytoplasmic side of the ER, two GlcNAc residues are attached to the lipid, dolichol, via a pyrophosphate linkage. Subsequently, five mannose residues are added by mannosyltransferases utilising GDP-mannose as the donor sugar. This lipid-linked heptasaccharide is then translocated across the ER membrane into the luminal side by a putative flippase. Here, the oligosaccharide is further extended by the addition of four mannoses and three glucoses via the action of unique glycosyltransferases utilizing dolichol-linked monosaccharide donors to yield $Glc_3Man_9GlcNAc_2$-P-P-dolichol (Kornfeld et al. 1985).

Once the LLO is assembled, the glycan portion is transferred to the protein co-translationally, whereby the innermost GlcNAc is attached in a *β*-linkage to the amide group of asparagine. This reaction is catalysed by the multi-subunit protein oligosaccharyltransferase. Following the transfer, the oligosaccharide branches undergo several processing reactions by glycosidases and glycosyltransferases. In the lumen of the ER, the lectin chaperone molecules calnexin and calreticulin (Zhang et al. 1997) bind to the glycoprotein's monoglucosylated oligosaccharide and help the

glycoprotein to fold correctly. Further trimming and addition of sugar moieties onto the *N*-glycans continues in the Golgi apparatus resulting in different modifications to the core structure. Thus, depending on the destination of the glycoprotein, its glycans may exist as high- mannose, hybrid and complex types. Misfolded glycoproteins that fail to achieve their correct folding state are removed from the calnexin/calreticulin cycle and encounter a membrane-bound lectin which triggers the ER-associated degradation (ERAD) (Hebert et al. 2007).This involves the export of the glycoproteins out of the ER via the translocon to the proteasome for degradation (Lord et al. 2000).

O-glycans

In *O*-linked glycosylation the carbohydrate chains are covalently linked to the hydroxyl group of serine (Ser), threonine (Thr) or hydroxylysine (hLys). In contrast to *N*-glycosylation, which occurs co-translationally in the ER and is initiated by the transfer of a large LLO, *O*-glycosylation is post-translational, occurs in the ER and the Golgi, and is synthesized by the step-wise addition of monosaccharides directly to the protein. The *O*-glycan is not trimmed by glycosidases during its synthesis. Furthermore, a consensus amino acid sequence for *O*-glycosylation attachment is poorly defined in most cases.

In the following pages, some examples of *O*-glycosylated proteins will be mentioned with an emphasis on the unusual *O*-fucosylation.

Common *O*-glycans

Mucins and proteoglycans are the two most representative glycoprotein groups bearing large numbers of *O*-linked sugars. Mucin polypeptides are extremely long and contain numerous tandem repeats with a relatively simple, uncharged amino acid composition rich in Pro, Ser and Thr residues. The synthesis of *O*-glycans is initiated by the attachment of GalNAc in an *α*-linkage to the hydroxyl of Ser/Thr. This initial step is catalysed by a large family of the polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAc-T). The ppGalNAc-Ts are an exception to the “one enzyme-one linkage” paradigm, which has been a wide-spread theory in the field of glycobiology. So far, the transferase activity of sixteen isozymes has been shown (Wandall et al. 2007) and according to *in silico* analysis 24 ppGalNAc-Ts exist (Ten Hagen et al. 2003). The isoforms are expressed in a tissue- and cell-specific and also in a time-dependent manner (Young et al. 2003). They also show a preference for different amino acids surrounding the target Ser/Thr residues. Subsequent sugar addition to the GalNAc linked protein, by other glycosyltransferases in the Golgi, gives rise to several core structures which often terminate in sialic acids. This clustering of negatively charged oligosaccharides in mucins can result in high water retention. Mucins are the major component of mucus and primarily function in the water retention and lubrication of surfaces found in the digestive, genital and respiratory tracts (Ten Hagen et al. 2003).

Proteoglycans consist of a core protein covalently modified with glycosaminoglycans (GAGs). These GAGs are linear polysaccharides ranging from 40 to over a 100 sugar residues (~10KDa to over 25KDa). Some proteoglycans, such as aggrecan, can have more than a 100 of these GAGs chains attached to its protein core. The GAGs are composed of disaccharide repeats, of amino sugars (GlcNAc and GalNAc) and uronic acids (GlcA and IdoA), which can be decorated with sulphate groups. In the GAG, hyaluronan (HA), there may be 10^4 GlcNAc*β*1,3GlcA*β*1,4 repeats. Exceptionally,

HA is never covalently linked to a protein and it is synthesised at the plasma membrane. In addition, the HA polymer is assembled from the reducing end at a rate of a 100 monosaccharide per second. HA holds a high water content which increases its volume. This hydration effect allows HA to act as a biological lubricant by reducing friction during movement and providing resistance under compression. Chondroitin sulphate (CS), dermatan sulphate and heparan sulphate *O*-glycan synthesis is initiated by the attachment of xylose onto a Ser in Ser-Gly rich sequences often surrounded by acidic residues. In contrast, keratan sulphate (Galβ1,4GlcNAcβ1,3)$_n$ has its *O*-glycans (KSII) linked via GalNAc on Ser/Thr as described above for mucins. Keratan sulphate (KSI) GAGs are *N*-linked to asparagine via GlcNAc. In the cornea, the KSI proteoglycan lumican, maintains the collagen fibril organisation and is essential for corneal transparency (Chakravarti et al. 1998). Large proportion of the extracellular matrix (ECM) of the central nervous system (CNS) is composed of proteoglycans. Increasing evidence shows that CS and HA play essential roles in the development of the CNS but inhibit CNS repair (Sherman et al. 2008). The cell surface proteoglycans have other distinct functions, such as binding to soluble polypeptide growth factors. The best studied example is the activation of fibroblast growth factor (FGF) through its interaction with heparin and heparan sulphate. This interaction is essential for the FGF to activate its receptor, FGFR, by stabilising the dimerised FGFR-FGF complex (Ornitz et al. 1995).

Unusual *O*-glycans

In addition to the common *O*-glycans, a wide variety of simpler *O*-glycans exists. Biosynthesis of these shorter glycan structures is initiated by the attachment of GlcNAc, fucose (Fuc), glucose (Glc), galactose (Gal) or mannose (Man) to a Ser or Thr.

O-GlcNAc modification

O-GlcNAc modification is the addition of a single GlcNAc in a β-linkage to Ser/Thr and occurs in the nucleus and the cytoplasm (Torres et al. 1984). Apart from its localisation, there are other reasons why this modification is so unique. With its cycling mechanism it is different from the stable glycan modifications and is more similar to the dynamic phosphorylation (Slawson et al. 2006). Site mapping of c-Myc (Kamemura et al. 2003), oestrogen receptor-β (Cheng et al. 2000), SV40 large T antigen (Medina et al. 1998) and endothelial nitric oxide synthase (Du et al. 2000) revealed that the attachment site of the GlcNAc and the phosphate were the same. This led to the hypothesis that *O*-GlcNAc and *O*-phosphate are reciprocal and competes for the same Ser/Thr residues.

Two enzymes are involved in *O*-GlcNAcylation, the *O*-GlcNAc transferase (OGT), which adds the GlcNAc to the Ser or Thr, and a glycosidase, *O*-GlcNAcase, which removes the sugar from the protein. More than 500 nuclear and cytoplasmic proteins are modified with *O*-GlcNAc (Hart et al. 2007), including nuclear core proteins, transcription factors, polymerases, RNA binding proteins, phosphatases, kinases, cytoskeletal proteins and chaperones (Zachara et al. 2006). Cells lacking the *O*-GlcNAc transferase gene are not viable, which indicates that *O*-GlcNAcylation has an essential role in fundamental cell processes such as protein phosphorylation, protein turnover and cell signalling (Shafi et al. 2000). *O*-GlcNAc can act as nutrient sensor responding to the glucose flux through the hexosamine biosynthesis pathway (Wells et al. 2003; Zachara et al. 2004). Recently, a novel

phosphoinositide binding domain on OGT was identified (Yang et al. 2008). Upon insulin stimulation PIP_3 can recruit OGT to the plasma membrane where it modifies various insulin signalling pathway proteins with *O*-GlcNAc, affecting their phosphorylation state and activity, and resulting in a reduced insulin response (Yang et al. 2008). The dysregulation of *O*-GlcNAcylation can lead to diseases such as insulin resistance, obesity, type 2 diabetes (Liu et al. 2000; Clark et al. 2003; Yang et al. 2008) and Alzheimer's disease (Ludemann et al. 2005).

O-fucosylation pathways

Fucose usually appears as a terminal sugar on various glycoconjugates, where it has important biological roles. For example, the core *α*1,6-fucose on the *N*-glycans of TGF-β1 receptor is required for receptor activation (Wang et al. 2006). Fucose can also be linked directly to the hydroxyl group of Ser/Thr in three different protein domains: epidermal growth factor-like domains (EGFs), thrombospondin type 1 repeats (TSRs) and the protease inhibitor PMP-C (Nakakura et al. 1992) (Figure 4). These domains are between 36 to ~60 amino acids long and contain three disulfide bonds in a module specific pattern.

PMP-C has a compact structure, consisting of three short anti-parallel *β* sheets (Mer et al. 1996). Although the presence of fucose on Thr-9 in PMP-C does not affect its ability to inhibit serine proteinases (Kellenberger et al. 1995), it does increase the thermostability of the molecule (Mer et al. 1996). Also, the non-fucosylated PMP-C can inhibit high voltage-activated Ca^{2+} currents, whereas the fucosylated form does not (Scott et al. 1997).

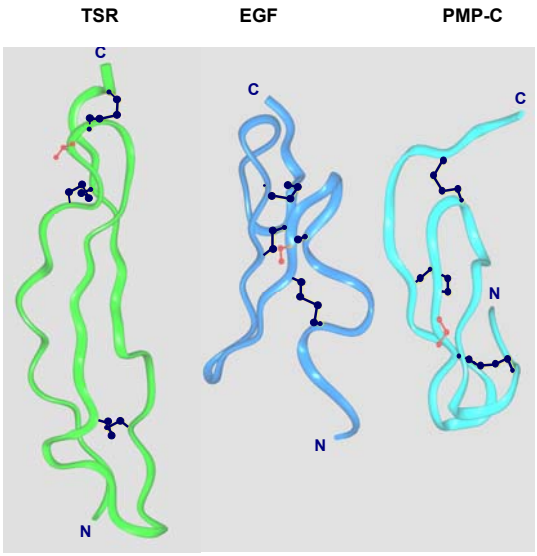


Figure 4. Three-dimensional structures of TSR, EGF and PMP-C

Crystal structures of TSR2 from thrombospondin-1 (PDB: 1LSL), EGF from factor VII (PDB: 1FF7) and PMP-C (PDB: 1PMC). Disulfide bridges are shown in dark blue, *O*-fucosylation sites are indicated in red.

The residue for the *O*-fucosylation is found within a *β*-sheet in EGF domains, whereas in TSRs it is located in the loop between the first and second strands (Figure 4).The biosynthetic pathway of *O*-fucosylation in PMP-C is unknown but has been elucidated for EGFs and TSRs (Figure 5).

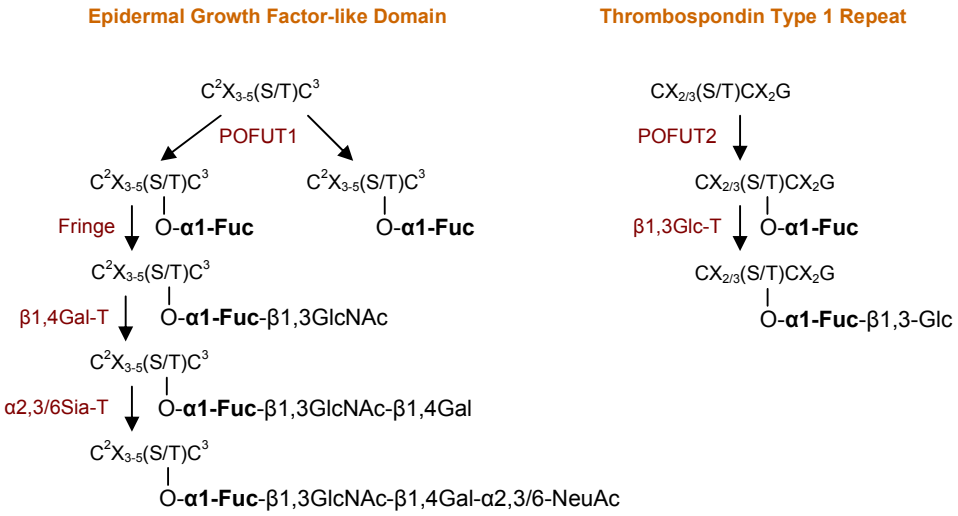


Figure 5. The three O-linked fucosylation pathways in EGF and TSR domains

In the putative consensus sequences X can be any amino acid. The superscript next to the cysteines indicates the position of the conserved cysteines counting from the N-terminal end. S/T is the *O*-fucosylated residue. The glycosyltransferases are shown in red. *In mammals there are three Fringe homologues: Manic, Lunatic and Radical Fringe. In *Drosophila* there is a single Fringe that extends the EGF fucosylation pathway to *O*-*α*1-Fuc-*β*1,3GlcNAc. There is no evidence of longer *O*-fucosylated glycans in *Drosophila*.

EGF *O*-fucosylation

EGF domains can be *O*-fucosylated in two different ways. The Ser/Thr residue in the putative consensus sequence $C^2X_{3\text{-}5}(S/T)C^3$ carries either a single *α*1-Fuc residue or a tetrasaccharide (Figure 5). The *O*-linked fucose monosaccharide occurs in urinary-type plasminogen activator (uPA) (Buko et al. 1991); tissue-type plasminogen activator; human clotting factors VII, IX and XII (Bjoern et al. 1991); Notch receptors (Harris et al. 1993) and its ligands, Delta1 and Jagged1 in mammals (Panin et al. 2002) and Delta and Serrate in *Drosophila* (Okajima et al. 2002; Panin et al. 2002); and Cripto (Schiffer et al. 2001; Yan et al. 2002; Shi et al. 2007). The fucose transfer is catalysed by the protein *O*-fucosyltransferase 1 (POFUT1) (Wang et al. 1998; Wang et al. 2001). The *O*-fucose may be extended with a GlcNAc by Fringe, a *β*1,3 *N*-acetylglucosaminyltransferase (Moloney et al. 2000; Panin et al. 2002), and further elongated with a galactose and then a sialic acid to yield the tetrasaccharide, Sia-*α*2,6Gal-*β*1,4-GlcNAc-*β*1,3-Fuc*α*1, as found on coagulation factor IX (Nishimura et al. 1992; Harris et al. 1993) and Notch receptor in mammals (Harris et al. 1993; Moloney et al. 2000).

O-fucosylation of EGF domains can have a ligand-receptor modulating role. In uPA, the *O*-fucose is required for the activation, but not the binding of uPA receptor, a mitogenic receptor involved in various kinase cascades (Rabbani et al. 1992). Extensive studies, summarized below, demonstrate how glycosylation can also affect the Notch signalling pathway. Notch receptors control several short-range signalling events through cell-cell contact such as lineage decision, lateral inhibition and inductive signalling. Mutations in the components of the Notch signalling pathway are associated with several human disorders including cerebrovascular dementia (CADASIL) (Joutel et al.

1996), cancer (Bolos et al. 2007), and developmental disorders of liver, heart, skeleton, eye, and kidney (Li et al. 1997; Oda et al. 1997).

Notch receptors are a family of single transmembrane glycoproteins containing up to 36 EGF repeats in their extracellular domain. As the receptor matures through the secretory pathway, many of its EGF repeats are modified with *O*-fucose and *O*-glucose in the ER (Moloney et al. 2000; Shao et al. 2002) The elongation of the *O*-linked glycans occurs in the Golgi. In the *trans*-Golgi, a furin-like convertase cleaves Notch into two fragments that heterodimerise into a structure which, in mammals, is essential for its activity. Notch translocates to the cell membrane and is activated upon binding its ligands that are expressed on apposing cells. The receptor-ligand interaction triggers the cleavage of Notch by a cell surface metalloprotease, ADAM10, leaving the extracellular domain of Notch bound to its ligand on the neighbouring cell. The truncated membrane-bound Notch is further cleaved within the membrane by a γ-secretase, releasing the intracellular domain of Notch which transits to the nucleus, where it acts as a transcription factor inducing the expression of target genes.

Ablation of the mouse *pofut1* gene is embryonic lethal at ~E9.5 with defects in somitogenesis, vasculogenesis, cardiogenesis, and neurogenesis. These phenotypes are consistent with a complete lack of all Notch signalling. Loss of the *pofut1* homologue in *Drosophila*, *Ofut1*, also results in a Notch loss of function phenotype. Whilst the importance of *pofut1/Ofut1* in Notch signalling is not in doubt, its mechanism of action is. POFUT1/Ofut1 is a soluble ER glycosyltransferase, which in the case of CHO-derived POFUT1, requires correctly folded EGF domains in order to recognise the consensus site (Wang et al. 1998). In *Drosophila*, the knockdown of OFUT1 causes Notch to accumulate in the ER, and an inactive form of the fucosyltransferase supports Notch secretion (Okajima et al. 2005). These experiments, from the Irvine laboratory, propose a chaperone-like role of OFUT1. Recently, work from the same laboratory has shown that the chaperone-like activity of OFUT1 is sufficient to generate functional Notch receptors (Okajima et al. 2008). Contrary to these conclusions, the Matsuno group do not find Notch accumulating in the ER of Ofut1 mutants, but rather in a novel endocytic compartment (Sasamura et al. 2007). This group proposed two trafficking models for OFUT1. The first, which is independent of its fucosyltransferase activity, argues that secreted OFUT1 associates with Notch at the cell surface and promotes the constitutive endocytosis of the Notch receptors (Sasamura et al. 2007). The second model, which is dependent on the fucosyltransferase activity, suggests OFUT1 promotes the trafficking of Notch from the plasma membrane to the sub-apical and adherens junction (Sasaki et al. 2007). The chaperone model and trafficking models for OFUT1 function in Notch signalling are apparently restricted to *Drosophila* as Notch receptors are equivalently expressed on the surface of mammalian cells lacking Pofut1 (Stahl et al. 2008).

An alternative way to determine whether the real cause of the Notch loss of function phenotypes lies in the loss of the fucosyltransferase activity in general, or the loss of the fucose residue in the EGF repeats of Notch receptors, is to mutate the *O*-fucose consensus sites. Using this approach, the effect of the absent *O*-fucose as well as the elimination of the elongated fucosylated glycan can be examined. A point mutation, replacing the Thr with an Ala, in the most conserved *O*-fucose site found in EGF 12 ligand-binding domain was created. The mutant mice show weaker Notch signalling during embryogenesis, post-weaning growth and T-cell development (Ge et al. 2008). To prove that it is the absence of *O*-fucose and not the amino acid change, which leads to the

observed signalling inhibition, Thr was replaced by Ser in EGF 12 with no effect in Notch signalling (Shi et al. 2007). In contrast, mutation of the *O*-fucose site in *Drosophila* EGF 12 leads to a hyperactive Notch response to its ligands (Lei et al. 2003). One reason for these opposite effects of the EGF 12 *O*-fucosylated glycans may be due directly to the different glycan structures found in *Drosophila* and the mouse. In *Drosophila* there is only evidence for the *O*-fucose and *O*-*α*1-Fuc-*β*1,3GlcNAc glycans (Xu et al. 2007), whereas in mammals the tetrasaccharide, *O*-*α*1-Fuc-*β*1,3GlcNAc-*β*1,4Gal-*α*2,3/6-NeuAc,also occurs (Figure 5).

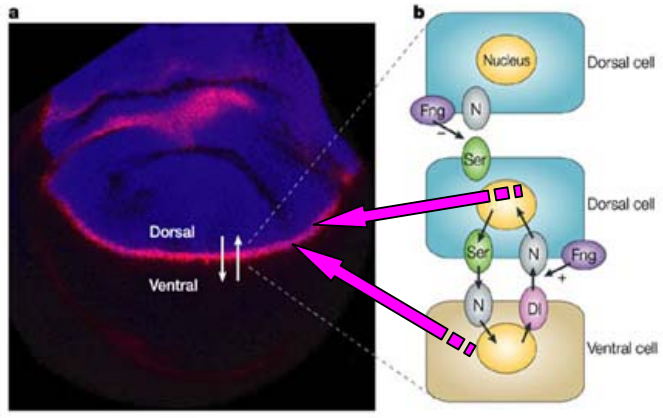
Protein type	*Drosophila*	Mammals	*C. elegans*
Receptors	Notch	Notch1 Notch2 Notch3 Notch4	Lin-12 Glp-1
Ligands	Delta Serrate	Delta1 Delta3 Delta4 Jagged1 Jagged 2	Lag-2 Apx-1
Fucosyltransferase	OFUT1	POFUT1	C15C7.1
Fringe	Fringe	Lunatic Fringe Manic Fringe Radical Fringe	-
Galactosyltransferase	*β*4GalNAcA *β*4GalNAcB	*β*4GalT-1 *β*4GalT-2 *β*4GalT-3 *β*4GalT-4 *β*4GalT-5 *β*4GalT-6	n.d.?
Sialyltransferase	n.d.?	*α*2,3SiaT *α*2,6SiaT	n.d.?

Table 2. Components of the Notch pathway

The interest in the glycosylation state of Notch was stimulated by the identification of a known modulator of Notch signalling, Fringe (Panin et al. 1997), as the *β*1,3-*N*-acetylglucosaminyltransferase that attaches GlcNAc in a *β*1,3-linkage to the *O*-fucose on Notch EGF domains (Bruckner et al. 2000; Moloney et al. 2000; Panin et al. 2002). In the *Drosophila* wing disc, Fringe potentiates Notch signalling induced by its ligand, Delta, and inhibits Notch signalling induced by its ligand, Serrate (Okajima et al. 2003). The outcome of this restricted Notch signalling is clearly observed along dorsal/ventral boundary (Figure 6) (Haines et al. 2003).

Fringe mutants show less severe phenotypes than O*fut1*$^{-}$ indicating that Fringe only affects a subset of Notch signalling events. In vertebrates, the influence of Fringe is more complicated, since there are four Notch receptors, five ligands and three Fringe proteins (Table 2). The three Fringe homologues, Lunatic (Lfng), Manic (Mfng) and Radical Fringe (Rfng) function in the Golgi but they are also secreted (Johnston et al. 1997; Lu et al. 2006). In general the effects of Lfng and Mfng is similar to *Drosophila* Fringe, they enhance Delta1-induced Notch signalling and inhibit Jagged1-induced Notch signalling. However, differences between Lfng and Mfng on Notch2 signalling have been reported (Shimizu et al. 2001), and Rfng can enhance both Delta1-induced and Jagged-1 induced Notch signalling (Yang et al. 2005). The vertebrate Fringes are differently expressed throughout

development and exhibit differences in their *in vitro* specific activities which suggest their distinct role in Notch signalling (Qiu et al. 2004; Rampal et al. 2005). Lfng is expressed in many cell types during embryogenesis and somitogenesis (Johnston et al. 1997). Mice with inactivated *Lfng* gene display severely disorganized axial skeleton and die perinatally or in early adulthood (Evrard et al. 1998; Zhang et al. 1998). Recently, a human patient with spondylocostal dystosis (defective somitogenesis) has been diagnosed to carry an autosomal recessive mutation (F188L) in the *Lfng* gene (Sparrow et al. 2006). The equivalent mutation of this conserved amino acid in mouse LFGN results in its mislocalisation and inactivity (Sparrow et al. 2006). In contrast, mice with mutations in *Rfng* show no visible phenotype (Moran et al. 1999).



Adapted from Haines et al., 2003

Figure 6. The role of *O*-fucose glycans in Notch inductive signalling

(a) The developing tissue that will give rise to the wing (the wing imaginal disc) is subdivided by the expression of *apterous* (blue stain) into dorsal and ventral cells. Normally, a peak of Notch activation occurs along the interface between dorsal and ventral cells. Notch signalling can be visualized by examining the expression of downstream target genes such as *wingless* (red stain). (b) Notch activation is positioned at the dorsal–ventral interface by the modulatory effect of Fringe (Fng) on signalling by the two Notch (N) ligands. Fringe (Fng) and Serrate (Ser) are expressed in dorsal cells (blue). Fringe potentiates Delta (Dl) signalling (shown by +), allowing it to signal to dorsal cells, and inhibits Serrate signalling (as shown by -), which limits Serrate to signalling from dorsal cells to ventral cells. The expression of Delta and Serrate is initially broad, but their later expression is stimulated by Notch activation. So, a positive-feedback loop is established between the two Notch ligands along the edge of Fringe expression in the wing.

The GlcNAc-Fuc disaccharide on Notch EGF repeats can be extended into a tetrasaccharide by the subsequent actions of a *β*1,4-galactosyltransferse (*β*4GalT) and a sialyltransferase (Moloney et al. 2000; Chen et al. 2001; Shao et al. 2002). Co-culture assays in CHO cells revealed that *β*4GalT-1 is required for Lfng and Mfng inhibition of Jagged1-induced Notch signalling. The subsequent addition of sialic acid was not necessary suggesting that the trisaccharide, Gal*β*1,4GlcNAc*β*1,3Fuc, supports Fringe-dependent modulation of Notch signalling in mammalian cells (Chen et al. 2001). However, mice lacking *β*4GalT-1 are viable with no obvious skeletal defects as would be expected in the inhibition of Notch1 signalling (Chen et al. 2006). This may be due to genetic redundancy, since in mammals there are six *β*4GalT genes and they all transfer Gal to GlcNAc. In the *Drosophila* genome only two *β*4GalT exist, and flies with mutations in both genes do not exhibit the *fng*-like phenotype (Haines et al. 2005). This raises the possibility that in *Drosophila*, the GlcNAc*β*1,3Fuc disaccharide is sufficient for Notch signalling (Xu et al. 2007).

EGF domains can undergo another type of glycosylation, *O*-glucosylation of serine residues. Bovine blood coagulation factors VII and IX were the first proteins identified with this modification (Hase et al. 1988). The structural analysis by NMR revealed Xyl-*α*1,3-Xyl-*α*1,3-Glc-*β*1-*O*-Ser53 on the first EGF domain in bovine factor IX (Hase et al. 1990). This is the only occasion the complete trisaccharide structure has been determined. The structure of an *O*-glucose containing trisaccharide attached to Notch1 isolated from CHO cells was not determined but is presumed to be the same as

that found on bovine coagulation factor IX (Moloney et al. 2000). The identification of several EGF containing proteins carrying *O*-glucosylated glycans (Nishimura et al. 1989) lead to the putative consensus sequence for *O*-glucosylation, C^1XSXPC^2 (C^1 and C^2 are the first and second conserved cysteines, respectively) (Harris et al. 1993; Moloney et al. 2000).

Protein *O*-glucosyltransferase activity has been demonstrated in many different cell lines and the mammalian enzyme was shown to require properly folded EGF domains for activity (Shao et al. 2002). Recently, in a genetic screen looking for genes involved in the development of mechanosensory organs on the body of the fly, *rumi* was identified (Acar et al. 2008). Mutations in *rumi* caused a temperature-sensitive Notch loss of function phenotype. The *rumi* gene encodes an *O*-glucosyltransferase that attaches glucose to serine residues in EGF repeats on Notch. Interestingly, Rumi is acting in the ER of the signal-receiving cell. The results suggest that the *O*-glucosylation of the EGF repeats in Notch mediated by Rumi is essential for Notch receptor folding and signalling (Acar et al. 2008). The *O*-glucose may also be a substrate for xylosyltransferases, which can extend the *O*-glucose monosaccharide to a trisaccharide. Since the xylosyltransferases have not been cloned, it is unclear if the *rumi* phenotype changes are caused by the absence of *O*-glucose or of a di- or trisaccharide containing xyloses.

TSR *O*-fucosylation

In 1975 an unusual disaccharide Glc-*β*1,3-Fuc-*O* attached to a Thr was discovered in human urine (Hallgren et al. 1975). Later, the same structure was isolated from several unidentified proteins expressed in Lec1-CHO cells labelled with ^{3}H-fucose (Moloney et al. 1997). Then in 2001 Hofsteenge and co-workers, whilst analysing another rare modification, *C*-mannosylation on human thrombospondin-1, demonstrated that the Glc*β*1,3Fuc disaccharide is present on thrombospondin type 1 repeats (TSRs) (Hofsteenge et al. 2001). This clearly established a new fucosylation modification on a distinct protein module (Figure 4). Several other TSR containing proteins, including properdin, F-spondin (Gonzalez de Peredo et al. 2002), ADAMTS13 (Ricketts et al. 2007) and ADAMTS-like 1/punctin-1 (Wang et al. 2007) have since been shown to carry this *O*-fucosylation (For further information about the TSR superfamily and its glycosylation see Chapter 5.3). The disaccharide is attached to the Ser/Thr residue in the putative consensus sequence $CX_{2/3}(S/T)CX_2G$ (Gonzalez de Peredo et al. 2002). It is now apparent that distinct glycosyltransferases act either in the *O*-fucosylation of EGF or TSR and there is no cross-talk between these two pathways (Luo et al. 2006a). Protein *O*-fucosyltransferase 2 (POFUT2) modifies correctly folded TSR (Luo et al. 2006b) and the subsequent glucose addition is catalysed by the *β*1,3-glucosyltransferase, *β*3Glc-T (Kozma et al. 2006; Sato et al. 2006). Both POFUT2 and *β*1,3-glucosyltransferase are soluble enzymes localised in the ER.

Little is known about the function of the enzymes participating in the *O*-fucosylation of TSRs. Canevascini et. al. (manuscript in preparation) have shown that deletion of the *C. elegans* homologue of POFUT2, causes an early dorsal migration phenotype of the anterior distal tip cell. *O*-fucosylation of the TSR domains in ADAMTS-like 1/punctin (Wang et al. 2007) and ADMATS-13 (Ricketts et al. 2007) is required for their secretion in mammalian cells. Currently it is unclear if the fucose alone or the Glc*β*1,3Fuc disaccharide is necessary for the secretion of TSR containing proteins. Recently it was

shown that POFUT2 is ubiquitously expressed in mouse embryos and required for normal gastrulation. *Pofut2* null embryos die by mid-gastrulation. Histological analysis suggest that POFUT2 promotes the epithelial to mesenchymal transition (Du et al. 2007). Not only the fucose, but also the glucose is needed for normal development. Mutations in the gene encoding *β*3Glc-T causes a severe developmental disease, Peters Plus syndrome (for detailed information see Chapter 6.4).

Glycosyltransferase	*C. elegans*	*Drosophila*	Mammals
Fucosyltransferase	PAD-2	OFUT2	POFUT2
Glucosyltransferase	ZC250.2*	CG9109-PA*	*β*3Glc-T

Table 3. Conservation of the TSR *O*-fucosylation pathway

Glycosyltransferases participating in the *O*-fucosylation of TSR repeats. *The enzymatic activity of the putative *β*1,3-glucosyltransferases in *C. elegans* and *Drosophila* needs to be determined.

C-glycosylation

A novel type of protein glycosylation, where the *C*-1 atom of mannose is covalently linked in an *α*-linkage to the *C*-2 atom of the indole ring of tryptophan, was discovered using a combination of mass spectrometry and NMR spectroscopy (Hofsteenge et al. 1994; de Beer et al. 1995; Löffler et al. 1996). *C*-mannosylation differs from *N*- and *O*-glycans not only in the nature of its linkage, but also in its fragmentation pattern by tandem mass spectrometry. In the case of *O*-linked hexose, the glycosidic bond between the amino acid and the sugar residue fragments first, resulting in a 162 Da hexosyl residue loss. In the case of carbon-linked mannose, it is not the glycosidic bond that fragments, but a 120 Da fragment which brakes off from the hexosyl residue. This 120 Da loss was previously observed in the fragmentation of flavone *C*-glycopyranosides (Li et al. 1991) and became the “fingerprint” of *C*-linked mannose. Trp is the only amino acid residue modified with a *C*-*C* linked mannose in mammalian cells (Zanetta et al. 2004). No further elongation of the mannose has been observed.

Ribonuclease 2 (RNase 2), from human urine, was the first mammalian protein shown to be *C*-mannosylated (Hofsteenge et al. 1994). In RNase 2 only the first tryptophan in the sequence WXXW is mannosylated. (Krieg et al. 1998) The motif occurs in a number of different protein families including the type I cytokines receptors, mucins and notably, the thrombospondin type 1 repeat (TSR) superfamily. In all cases examined, members of these families are *C*-mannosylated (Doucey et al. 1999; Hofsteenge et al. 1999; Hartmann et al. 2000; Gonzalez de Peredo et al. 2002; Furmanek et al. 2003; Perez-Vilar et al. 2004). Interestingly, in many TSR-containing proteins there is an extended putative recognition motif, WXXWXXW, in which one, two or even all three tryptophans may be modified with mannose. Moreover, a deca neuropeptide from the stick insect *Carausius morosus*, that does not contain any sequence similarity to the WXXW motif, is also *C*-mannosylated (Gade et al. 1992; Munte et al. 2008). Either there are undetermined residues that support *C*-mannosylation or there are multiple C-mannosyltransferases.

C-mannosylation is carried out by an unidentified membrane-bound *C*-mannosyltransferase that uses dolichyl-phosphomannose (Dol-P-Man) as the sugar donor (Doucey et al. 1998). The enzyme activity has been detected in many organisms from *C. elegans* to man, but not in bacteria, yeast or insects (Krieg et al. 1997; Zanetta et al. 2004). Clearly there is a discrepancy between the

observed modification in insects and their apparent lack of enzyme activity that still needs to be resolved. The biological function of *C*-mannosylation remains to be determined. Nevertheless, the modification is necessary for the secretion of mucins, MUC5AC and MUC5B Cys subdomains, in Chinese hamster ovary cells (Perez-Vilar et al. 2004).

Lectins

Lectins are proteins that bind glycans and function in decoding this sugar information into a biological effect. They were discovered more than 100 years ago in plants but are now known to exist throughout nature, including microorganisms where they are called adhesins and toxins in bacteria, and hemagglutinins in viruses. Lectins exist as integral membrane proteins, as well as soluble proteins, hence they can interact with glycoconjugates located either in the extracellular or luminal spaces (Varki et al. 1999).

The binding between a lectin and its counter partner glycan is reversible, non-covalent, and tends to be of low-affinity. Animal lectins are usually multivalent either through their multi-subunit organisation or due to the presence of several carbohydrate recognition domains. The multivalent binding of low-affinity single sites generates high avidity, enabling the lectin to function in diverse biological processes.

The major lectin families are: C-type (collectins, selectins); calnexin group (calnexin, calreticulin, calmegin); L-type (lectins in ER-Golgi pathway); P-type (mannose-6-phosphate receptors); S-type (galectins); and I-type lectins (siglecs).

Selectins take part in the recruitment of leukocytes to inflammatory sites. As a response to the chemokines released by the injured or inflamed tissues, endothelial cells start to express selectins on their cell surface. These transmembrane adhesion molecules mediate the leukocytes to tether and roll along the endothelial layer. The rotational movements of the white blood cells are possible by the forming and breaking of the weak glycan-selectin interactions. The stronger adherence of the leukocytes to the endothelial layer is facilitated by protein-protein interactions between integrins and their receptors. Finally the leukocytes migrate through the endothelium into the tissue via extravasation (Kansas 1996).

Lectins participate in the glycoprotein quality control mechanism in the ER, in which the misfolded proteins are separated from the correctly folded ones (Ellgaard et al. 1999). As described earlier, nascent glycoproteins are *N*-glycosylated in the ER through the attachment of the $Glc_3Man_9GlcNAc_2$ oligosaccharide (Figure 3). Glucosidase I and II trim off the outer two glucose residues to yield $Glc_1Man_9GlcNAc_2$ which is recognised by the lectin chaperones, calnexin and calreticulin. This binding exposes the maturing glycoprotein to ERp57 (a glycoprotein-dedicated oxidoreductase), and generally slows the folding reaction, enhancing its overall efficiency (Helenius et al. 1992). When the substrate is released from the lectin chaperones, glucosidase II removes the final glucose, thereby preventing substrate rebinding to calnexin/calreticulin. If the glycoprotein is folded correctly it proceeds into the secretory pathway, otherwise the nearly native folded glycoproteins are re-glucosylated by the UDP-glucose:glycoprotein glucosyltransferase and the glycoprotein enters another round of folding interactions with calnexin/calreticulin (Hauri et al. 2000; Caramelo et al. 2004). Extensively misfolded glycoproteins are destined for ER-associated protein degradation

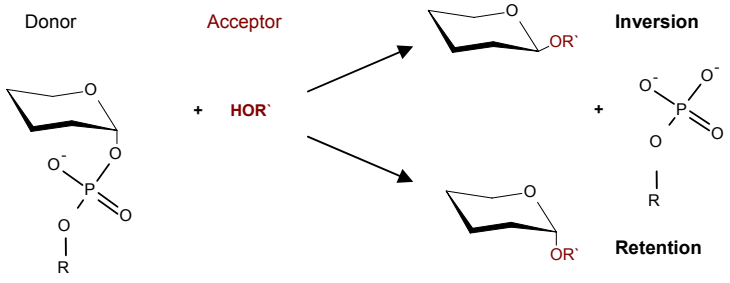
(ERAD). The *N*-glycans on these ERAD candidates are de-mannosylated, in particular the *α*1,2-linked mannoses are removed to yield $Man_{5-6}GlcNAc_2$. This excludes the misfolded glycoprotein from futile folding cycles and facilitates its retro-translocation into the cytosol and subsequent removal by the proteasome. The extensive de-mannosylation of misfolded glycoproteins also prevents their interaction with the lectin cargo receptors ERGIC53/VIP36 (L-type lectins) thus reducing the protein load in the ER.

Lysosomes are intracellular organelles accommodating lysosomal enzymes that degrade various macromolecules. In order to avoid the release of these harmful enzymes from the cell, they are uniquely modified with a targeting signal. As the lysosomal proteins pass through the secretory pathway, they acquire a phosphomannosyl residue on their *N*-glycans in the *cis*-Golgi. Mannose-6-phosphate receptors (P-type lectins), recognise the mannose-6-phosphate on the lysosomal enzymes in the *trans*-Golgi and direct the protein traffic to the late endosomes via clathrin-coated vesicles. In the acidic endosomes, the receptor releases the enzyme, which is then packaged into the lysosomal compartment. The free mannose-6-phosphate receptor can either recycle back to the *trans*-Golgi to cargo more enzymes or it can cycle to the plasma membrane to re-capture and endocytose any mislocalised hydrolases (Kaplan et al. 1977; Taylor et al. 2003).

Lectins also participate in the clearance of glycoproteins e.g. glycoprotein hormones from the circulatory system, adhesion of infections agents to host cells, signalling, and in cell interactions in the immune system, malignancy and metastasis (Ambrosi et al. 2005).

Carbohydrate Active Enzymes

Glycosyltransferases (GTs) and glycosidases (GHs) account for ~1-3% of all the encoded proteins in most organisms (Davies et al. 2005). GTs are responsible for the transfer of the glycosyl residue from activated high-energy sugar donors (e.g. GDP-fucose, Dol-P-mannose) to an acceptor (e.g. sugar, lipid, protein or small molecule), forming a glycosidic bond (Figure 7). GHs catalyse the hydrolysis of the glycosidic bonds.



Adapted from Coutinho *et al*., 2003

Figure 7. Glycosyltransferase reaction schemes

Both GTs and GHs are subdivided into two main groups as either inverting or retaining enzymes, depending on the stereochemical outcome at the anomeric centre relative to that of the sugar donor (Lairson et al. 2004). The Carbohydrate Active enZyme database (CAZy, http://www.cazy.org) classifies the enzymes into families based on their amino acid sequence similarities (Campbell et al. 1997; Coutinho et al. 2003). This database was set up more than 15 years

ago and is continuously updated. Currently there are over 200 families in CAZy including GTs, GHs, polysaccharide lyases and carbohydrate esterases. The GT section of the database contains more than 16000 open reading frames organised into 90 families. Within the human genome, there are 240 sequences listed, which fall into 42 distinct families.

Crystal structures have been solved for 28 different GT families. In contrast to the GHs that show a large variety of 3D folds including all *α*, all *β* or mixed *α*/*β* structures, the GTs have two main folds, termed GT-A and GT-B. These folds are similar to the Rossmann-type fold, which is a six-stranded parallel *β*-sheet linked together with *α*-helices and found in many nucleotide-binding proteins (Lesk 1995). The GT-B fold contains two Rossmann-like folds connected with a linker region. There is no strictly conserved residue shared in the GT-B family members (Hu et al. 2002). The GT-A fold consists of a seven-stranded *β*-sheet, with one sheet running anti-parallel. The central *β*-sheet is flanked by a smaller one and association of both, via a short loop, creates the active site. A DXD motif, located in this short loop, interacts with the phosphate groups of the nucleotide donor through the coordination of a divalent cation. A third fold has emerged with the crystal structure of the sialyltransferase CstII from *Campylobacter jejuni*, which is structurally close to the GT-A fold but lacks the DXD motif (Chiu et al. 2004). There are no structures solved for GTs using lipid-phosphate donor sugars.

Since the glycosylation reactions occur in the secretory pathway and in the cytosol, the GTs exist as membrane-bound, secreted, and soluble proteins. The location of a GT dictates where it acts in the glycan biosynthesis pathway. Golgi-resident GTs tend to adopt a type II membrane topology (Paulson et al. 1989), consisting of a short cytosolic N-terminal tail with a single transmembrane domain, linked to a long stem region extending into the Golgi lumen followed by the globular catalytic domain at the C-terminus. Proteases acting on the stem region can release the catalytic domain, leading to secretion of the GTs from the cell. Since these circulating GTs do not have access to their donor substrates, they do not act as GT. GTs residing in the ER may be multi-transmembrane or soluble enzymes. Soluble GTs are retained in the ER via their C-terminal KDEL-like ER retrieval sequence (Pelham 1988; Pidoux et al. 1992; Zhen et al. 1993).

Sugar donors are made in the cytosol and have to be transported into the appropriate secretory compartment for glycosylation to occur. Distinct nucleotide sugar transporters are present in the Golgi/ER membrane and act as antiporters (Martinez-Duncker et al. 2003), in which the entrance of the sugar nucleotide into the Golgi/ER lumen is coupled by the exit of the corresponding nucleoside monophosphate from the Golgi/ER (Hirschberg et al. 1998). Curiously, although fucosylation has been shown to occur in the ER (Luo et al. 2005; Luo et al. 2006b), the existence of an ER GDP-fucose transporter has not been demonstrated yet.

There are several exceptions to the “one-enzyme-one-linkage” paradigm, which was associated to the action of glycosyltransferases for a long time. Redundant glycosyltransferase families, (e.g. polypeptide *N*-acetylgalactosaminyltransferases (ppGalNAc-T) and *α*2,8-sialyltransferases), are known where the same glycosidic bond is formed by several, structurally and genetically related enzymes. Other glycosyltransferases, such as fucosyltransferase III, can catalyse the formation of *α*1,3 and *α*1,4 glycosidic linkages (Kukowska-Latallo et al. 1990). Moreover the acceptor specificity of a certain glycosyltransferases may be modified by their interaction with other proteins.

*β*1,4-galactosyltransferase switches its substrate specificity from GlcNAc to Glc in the presence of *α*-lactalbumin to allow lactose synthesis during milk production (Brew et al. 1975; Elling et al. 1999). Furthermore, there are polymerizing GTs such as chondroitin synthases with GlcAT-II and GalNAcT-II activities, which have two active sites and catalyse two consecutive sugar transfers (Kitagawa et al. 2001).

5.2. Dysregulation of Glycosylation

It has become evident that defects in monosaccharide synthesis, nucleotide sugar transport, assembly and attachment of carbohydrate structures to proteins and lipids or glycan recognition can lead to misglycosylation. This has been implicated in several inherited and acquired human diseases.

Acquired Human Diseases

The number of non-inherited diseases with aberrant glycosylation is increasing. This is not surprising considering the diverse roles of the glycans, carbohydrate active enzymes and lectins in various cellular mechanisms.

Cancer

Over expression or loss of expression of certain glycan structures can occur in malignant cells. The altered glycosylation can result in loss of adhesion or initiate new cell adhesion and therefore promote the formation of secondary tumours at new sites.

The metastatic potential of tumour cells has been correlated with the increased sialylation of certain cell surface glycans. Upon malignant transformation of the cell, the activity of GlcNAc-transferase-V (GnT-V) is increased, which is caused by the transcriptional activation of its promoter by the Ras-Raf-Ets signal transduction pathway. This leads to an increase of *N*-linked *β*(1,6)-branching. The additional branching creates extra sites for terminal sialic acid residues, which consequently results in upregulation of sialyltransferases increasing the global sialylation (Dennis et al. 1987; Orntoft et al. 1999). Endogenous lectins such as siglecs and selectins bind to sialic acid residues and are known to mediate cell adhesion and extravasation of normal cells. This property can facilitate invasion of tumour cells. Interestingly, GnT-V is a bifunctional protein, participating not only in tumour metastasis but also in angiogenesis, the earlier phase of tumour progression. This angiogenic potential is not mediated by the GnT-V enzyme activity, but instead by the translocation of its cleaved soluble form to the extracellular matrix. The basic domain of the secreted soluble form of GnT-V promotes the release of fibroblast growth factor-2 resulting in tumour angiogenesis (Saito et al. 2002).

Inherited Human Diseases

A large number of inherited disorders are caused by dysfunctional carbohydrate active enzymes and lectins. These diseases are biochemically heterogeneous and exhibit a broad clinical spectrum of symptoms within a particular disorder, and even between patients having a defect in the same gene. Most of the diseases are currently untreatable, but some do have simple and effective therapies.

Lysosomal storage disorders

Biomolecules, such as carbohydrates, lipids, proteins and sterols have finite life times and are degraded after being transported into the lysosomes. There are over 40 known lysosomal storage diseases (LSD) that result from a deficiency in the activity of a lysosomal hydrolase and a rapid accumulation of the corresponding substrates (Vellodi 2005). The degree of the severity of the disorder depends on the residual enzyme activity, the amount of the accumulated material and even on the time of the onset of the disease (Butters 2007).

The lysosomal endo- and exoglycosidases are responsible for the degradation of the oligosaccharides. Exoglycosidases hydrolyse the terminal sugar from the non-reducing end of the glycan chain, whereas the endoglycosidases cleave internal glycosidic bonds. The large family of glycolipid LSD includes Gaucher, Fabry, Tay-Sachs and Sandhoff diseases and the disorders have often been difficult to treat. However, in the last two decades the development of different therapies has made significant progress. Apart from bone marrow transplantation, enzyme replacement therapy (ERT) is also used to correct the deficient enzyme activity. The administration of recombinant *β*-glucocerebrosidase to more than 4000 patients with Gaucher disease is an example of a very successful therapy. However, ERT is expensive and cannot treat all the symptoms associated with LSD. Therefore, alternative therapeutics are being developed. A non-enzymatic way to treat LSD is by substrate reduction therapy. This approach attempts to inhibit the synthesis of glycolipids and thus diminish the load in the lysosomes. A number of small molecule inhibitors against the glucosylceramide synthase, which catalyses the first step in glycolipid biosynthesis, are currently in clinical trials. An alternative therapeutic approach is to correct the malfunctioning lysosomal enzyme during its biosynthesis through chaperone-mediated therapy. Some small molecular chaperones appear to bind to the active site of the mutated lysosomal enzyme during its biosynthesis in the ER and assist in protein folding, reducing ERAD, and increasing the trafficking of the more active glucosylceran synthase to the lysosome (Futerman et al. 2004; Vellodi 2005). The level of enzyme activity required to prevent substrate accumulation is ~10% of the normal level, which is achievable with the chaperone-mediated therapy. Furthermore, the small molecular chaperones, in contrast to ERT, can cross the blood-brain-barrier, opening up the possibility of treating LSD with neurological involvement (Fan 2008).

Congenital Disorders of Glycosylation

Several pathways are involved in the assembly and the attachment of the numerous oligosaccharides linked to proteins and lipids. To date, more than 37 genetic disorders result from a deficiency in the glycosylation of glycoproteins or glycolipids, including malfunctioning enzymes which participate in the *N*- and *O*-glycan synthesis pathways (Freeze 2006) as well as in GPI-anchor biosynthesis (Kranz et al. 2007). These autosomal recessive metabolic disorders are clustered into a family termed, congenital disorders of glycosylation (CDG). These disorders are very rare, less than 1000 patients are known worldwide (Grunewald 2007; Jaeken et al. 2007). Although it is believed that the actual number of CDG cases far exceeds those diagnosed as around 20% of patients die in the first 5 years of childhood. In addition, there appears to be a lack of awareness of these diseases amongst clinicians.

Fortunately, efforts are underway in the United States to include CDG in the next generation of infant screening tests. The advances in diagnostic testing are also uncovering many new types of CDG.

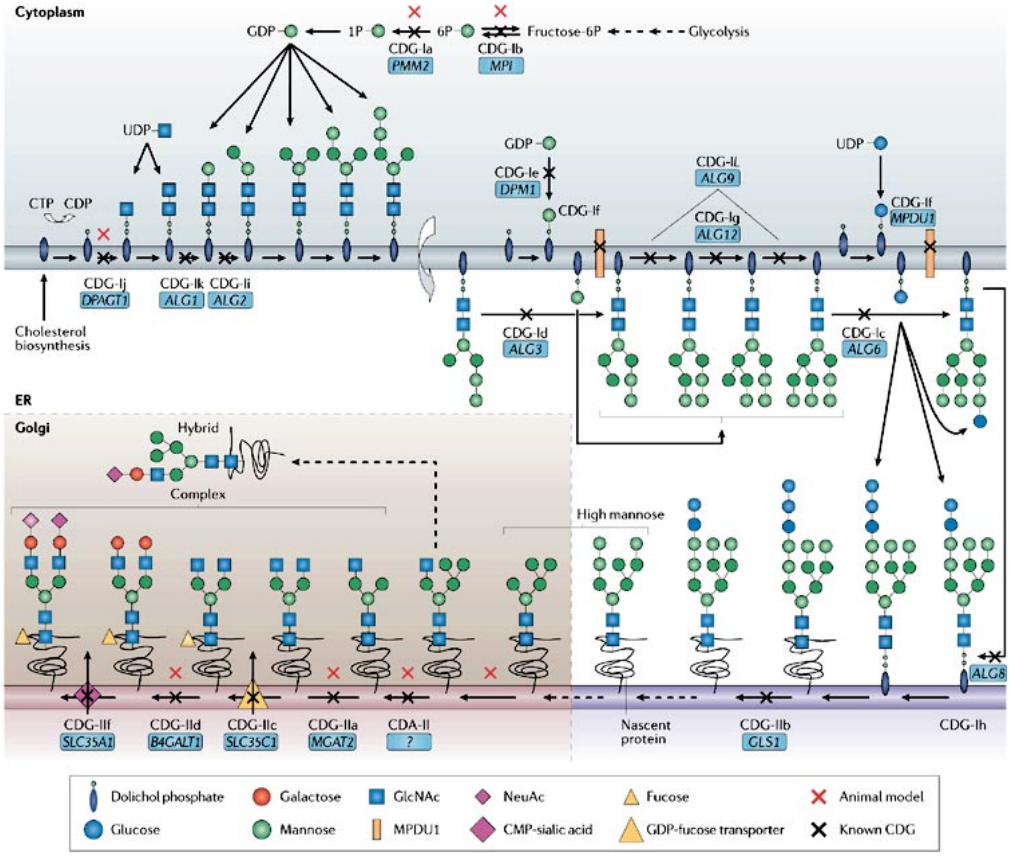
CDG types

Since the first patient with CDG was reported in 1980, several new types of CDG have been identified. To date there are over twenty *N*-glycosylation defects, ten *O*-glycosylation defects and two defects in lipid glycosylation (Freeze 2007; Jaeken et al. 2007). The known CDG are divided into two groups depending on the nature and the location of the defect. Type I disorders (CDG-I) involve the disrupted biosynthesis of the LLO precursor leading to a lack of *N*-glycans attached to the proteins. Type II disorders (CDG-II) exhibit defects in either the trimming, remodelling, extension reactions or the intracellular trafficking of *N*-glycans (see Figure 8) (Aebi et al. 1999). The disorders caused by defective *O*-glycosylation have not been formally included into the groups of CDG and still have their names which are descriptive to the disease. The number of patients, where the glycosylation deficiency has not yet been identified is continuously growing (Jaeken et al. 2001) and therefore a new group, CDG-X, has recently been added.

Considering the multiple steps involved in glycosylation and the diversity of proteins that are modified or interact with glycosylated proteins, it is unsurprising that the clinical manifestations of the inherited disorders are heterogeneous. Clinicians should screen their patients for CDG if they present the following symptoms: failure to thrive, hypotonia, inverted nipples, unusual fat deposits, mental and psychomotor retardation, stroke-like episodes, protein-losing enteropathy, hypoglycemia, generalized demyelination and optic atrophy (Freeze 2006). Symptoms also vary between individuals with the same disorder. This can be partially explained by differences in their residual enzymatic activities and/or additional genetic factors.

Defects in the biosynthetic pathway of the core *N*-glycan prevent the completion of the glycan synthesis, which result in decreased *N*-glycosylation. The complete loss of *N*-glycosylation is lethal (Marek et al. 1999). Mutations in PMM2 cause the most common CDG, CDG-Ia. PMM2 encodes a phosphomanno-mutase that converts mannose-6-phosphate to mannose-1-phosphate, the precursor of GDP-mannose required for all mannosylation reactions. Over 90 mutations have been identified in PMM2 that result in loss of the enzyme's thermostability or in reduced substrate binding. In CDG patients there is always some residual enzymatic observed. Interestingly, the correlation between mild phenotype and higher residual PMM2 activity is only observed in some cases, which suggest that other genetic factors are involved in CDG-Ia.

CDG-Ib is different from most of the other CDG types. Strikingly, it does not affect the central nervous system and most of its phenotypic changes, such as protein-losing enteropathy, hypoglycemia and hyperinsulinemia can be corrected by supplementing mannose in the diet. In CDG-Ib, mutations in phosphomannose-isomerase (MPI) result in a decreased level of mannose-6-phosphate derived from fructose-6-phosphate. However, dietary mannose can be converted into mannose-6-phophate by a hexokinase and therefore by-pass the metabolic block caused by the deficient MPI (Leroy 2006).



Adapted from Freeze, H. H., 2006

Figure 8. Sites of genetic defects in the biosynthetic pathway for *N*-linked glycans

In this schematic representation of the *N*-glycan biosynthetic pathway, the assembly, the attachment and the processing of the LLO can be followed. The steps, where the genetic disorder appears, are indicated by names of the genes involved. Above the gene name, the official CGD nomenclature is written.

Genetic defects concerning the processing of the *N*-glycans in the Golgi apparatus and other glycosylation pathways often manifest in different clinical features than the ones that can be found in type I CDGs. CDG-IIc, also known as leukocyte adhesion deficiency type II (LADII), is characterised by recurring bacterial infections, deficient physical growth, microcephaly, abnormal body proportions and psychomotor retardation. The disorder is caused by a defect in the GDP-fucose transporter in the Golgi. This causes a decrease in fucosylated glycans, such as the fucosylated selectin ligands on the surface of neutrophils required for leukocyte adhesion. Oral fucose therapy in LADII patients can restore the fucosylation of cells and correct their leukocyte adhesion phenotype and immunodeficiencies. Apparently fucose supplementation in LADII works via an alternative GDP-fucose transporter (Helmus et al. 2006).

The clinical manifestations of *O*-glycosylation defects are different from those observed in *N*-glycosylation disorders. *O*-mannosylation defects cause congenital muscular dystrophies, such as Walker-Warburg syndrome where defective glycosylation of *α*-dystroglycan leads to the disruption of *α*-dystroglycan interactions with the extracellular matrix resulting in severe muscular dystrophies, structural eye abnormalities, brain malformations and short life span. Defects in *O*-xylose pathways usually cause bone, cartilage and other extracellular matrix abnormalities, such as Ehler-Danlos syndrome.

Detection of CDGs

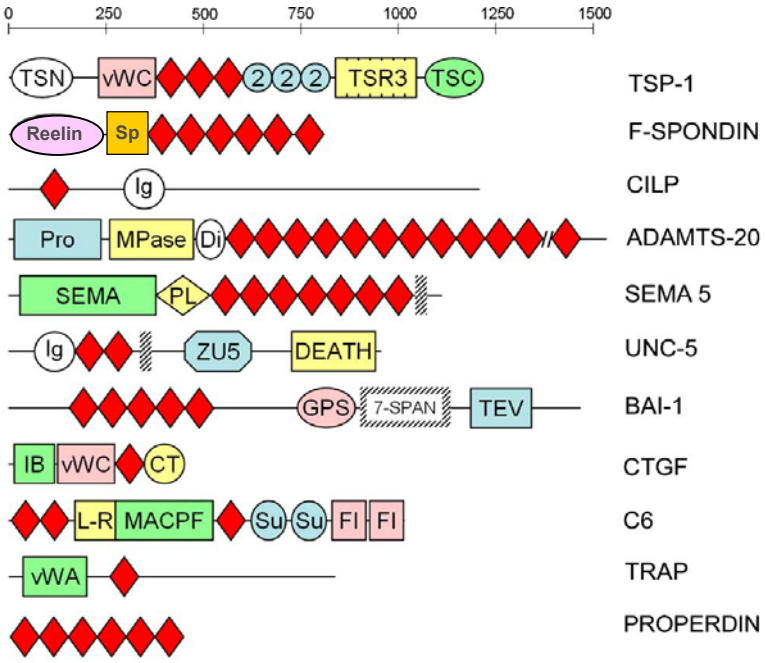
The common screening test for CDG is the isoelectric focusing of the serum protein, transferrin (Tf). Tf has two *N*-glycosylation sites, both occupied with biantennary or triantennary disialylated glycans, resulting in negatively charged species. In the case of type I CDG, a subpopulation of Tf molecules lacking one or both *N*-glycan chains leads to a reduction in the negatively charged species (carbohydrate deficient Tf). In type II CDG the *N*-glycosylation sites are occupied but the glycosylation is incomplete, resulting in different glycoforms with 0, 1, 2, 3, 4 or 5 sialic acids. The Tf serum test is fast, cheap and generally reliable. Recently HPLC and mass spectrometric analysis of carbohydrate deficient Tf by either electrospray ionization or matrix-assisted laser desorption/ionization, have proved to be efficient analytical methods. Often a number of protein variants of transferrin are present in the serum which can affect the isoelectric point of the molecule (Ohno et al. 1992). There are some disorders, such as uncontrolled galactosemia (Charlwood et al. 1998; Sturiale et al. 2005), fructose intolerance (Adamowicz et al. 1996; Jaeken et al. 1996), and heavy alcohol consumption (del Castillo Busto et al. 2005) which can mimic CDG profiles leading to false positive results. To avoid these pitfalls of diagnosis a multiple test can be used to screen CDG patients. Here the Tf isoelectric focusing screen is combined with MALDI-TOF analysis for glycosylation site occupancy and analysis of the glycan structures on Tf glycopeptides (Wada 2006). To identify the subtype of CDG, the most time-consuming step, the identification of the defective gene in each patient is needed. This is usually done by assaying for the altered enzymatic activity.

Since the nature of the *O*-glycosylation defects are so diverse, testing only one specific glycoprotein would not lead to a diagnosis. So far, the only screening method that is available for the detection of defects in *O*-mannosylation, is by immunohistochemical staining of the *α*-dystroglycan with monoclonal antibodies (Wopereis et al. 2006). Recently a novel immunopurification-mass spectrometry method was developed to confirm the suspected glycosylation defects, occurring in Peters Plus syndrome (Hess et al. 2008). Patients with this syndrome have biallelic truncating mutations in the gene encoding *β*3Glc-T (Lesnik Oberstein et al. 2006), that is responsible for the glucosylation of *O*-fucosylated TSR repeats (Kozma et al. 2006; Sato et al. 2006). This approach highlights the importance of analysing discrete protein domains, TSR in this case, to detect changes in rare forms of glycosylation.

5.3. Thrombospondin Type 1 Repeats

Thrombospondin type 1 repeats (TSRs) were first identified in the multi-domain extracellular matrix protein thrombospondin 1 (TSP-1) (Lawler et al. 1986). TSRs are conserved in evolution with 31 *Drosophila*, 28 *C. elegans* and 141 human proteins. These proteins often consist of multiple modules with the TSRs secreted extracellularly. The TSRs are functionally important in the regulation of extracellular matrix organization, cell-cell interaction, axonal and cell guidance (Tucker 2004).

There are only two thrombospondin TSR containing proteins, TSP-1 and TSP-2, that each of them has three tandem TSRs located between the vWC and the EGF-like domains. TSRs are not only found in TSPs but also in other multidomain proteins, including F-spondin, SCO-spondin, UNC-5, semaphorin-5, thrombospondin-related anonymous protein (TRAP), thrombospondin-related sporozoites protein (TRSP) and circumsporozoite protein (CSP) from the *Plasmodium falciparum* parasite, properdin, complement proteins C6, C7, C8, C9 and the ADAMTS (A Disintegrin And Metalloproteinase with ThromboSpondin type 1 repeats) family (Adams et al. 2000).

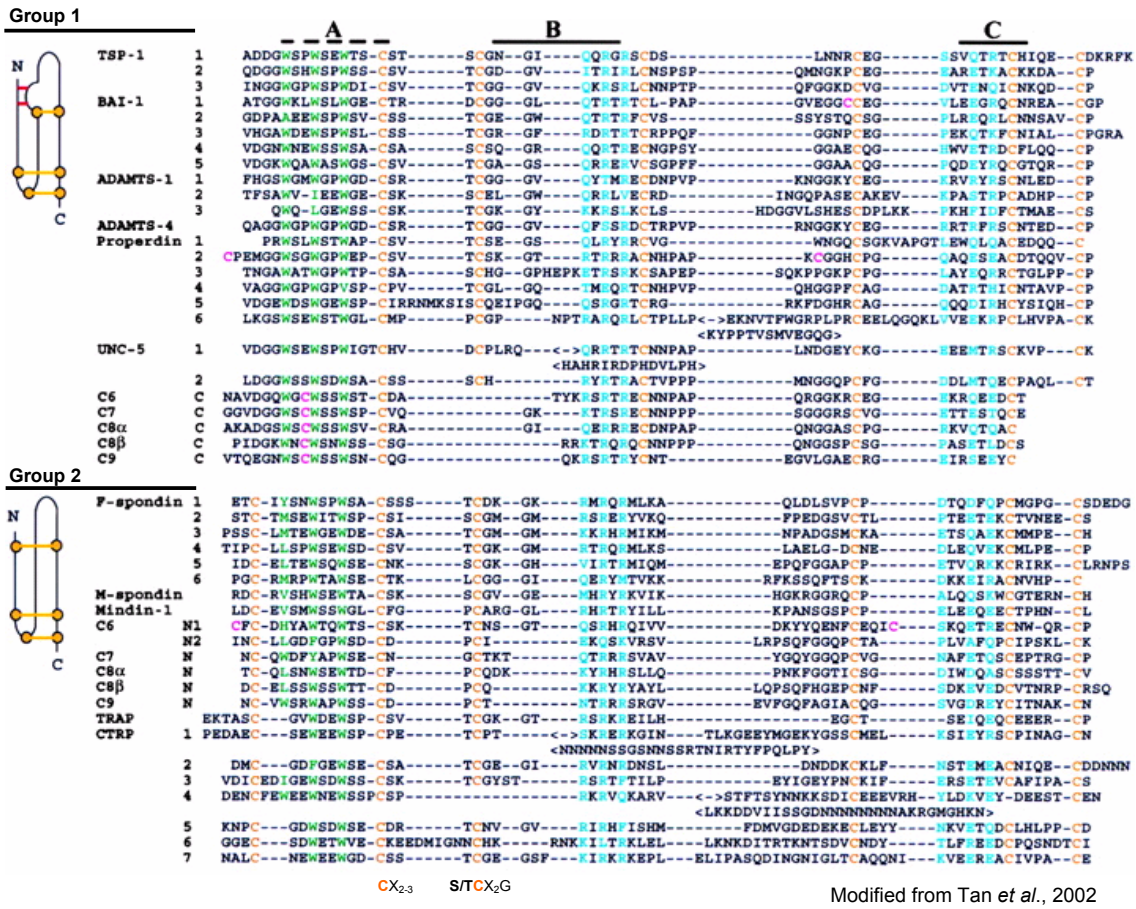


Adapted from Tucker, 2004

Figure 9. Domain structure of some TSR-superfamily members

The length of the protein corresponds to the amino acid scale at the top of the figure. The TSR domains are indicated by red diamonds. The cross-hatched boxes represent the transmembrane domains. vWC, von Willebrand factor type C domain; 2, calcium-binding EGF-like domain; Reelin, Reelin domain; Sp, Spondin domain; 7-SPAN, seven transmembrane spanning domain; vWA, von Willebrand factor type A domain; CT, C-terminal cysteine knot-like domain; DEATH, death apoptosis domain; Di, disintegrin-like domain; FI, factor I membrane attack complex; GPS, latrophilin/CL-1-like G-protein-coupled receptor proteolytic site; IB, insulin growth factor-binding protein domain; Ig, immunoglobulin domain; L-R, LDL receptor class A; MACPF, membrane attack complex/perforin domain; MPase, metalloprotease domain; PL, plexin repeat; Pro, reprolysin family propeptide; SEMA, semaphorin domain; Su, Sushi (SCR) domain; TEV, domain that interacts with PDZ-containing proteins; TSC, thrombospondin C-terminal domain; TSN, thrombospondin N-terminal domain; TSR3, thrombospondin type 3 domain; ZU5, domain present in ZO-1 and UNC-5.

TSRs are ~60 amino acids long with around 12 conserved residues (highlighted in Figure 10). The crystal structure of TSR2-TSR3 from human TSP-1 revealed that TSRs fold as three anti-parallel strands (Tan et al. 2002) (Figure 11). The B and C strands form limited regular β-sheets, whereas the A strand has a unique rippled conformation and contains the conserved sequence motif WXXWXXW. The tryptophan side chains form a layer with the conserved arginines guanidinium groups from the B strand filling in the spaces between the indole moieties. The top and bottom of this stacked core is capped by disulphides.



Modified from Tan *et al.*, 2002

Figure 10. Multiple sequence alignment of the two major groups of TSRs

The disulphide patterns of groups 1 and 2 are drawn schematically as yellow lines on the left-hand side. For group 1, two hydrogen bonds between the jar handle and the N-terminus are drawn as red lines. In the alignment the paired cysteines are also shown in yellow. The putative recognition motif for *O*-fucosylation (in bold) is indicated under the alignment.

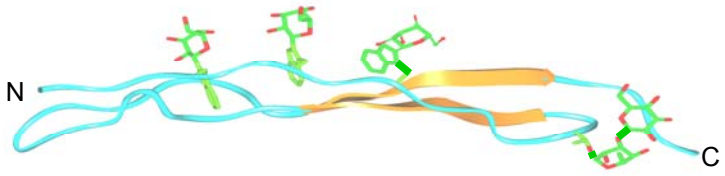
Interestingly, it appears that there are two main groups of TSR-containing proteins based on their disulphide bond pairings (Figure 10). The main difference occurs at the top of the TSR layered structure. For TSRs in group 1, which includes members of TSP, BAI, ADAMTS and properdin, the top cysteine bridge is formed by cysteines from the end of the B strand to the beginning of the C strand (Cys^3-Cys^4). In contrast, TSRs in group 2, which includes F-spondin and TRAP, the top disulphide is formed at the start of the A and C strands (Cys^1-Cys^4). This suggests that the N-terminus in group 2 TSRs is stabilised by this disulphide bridge rather than the jar handle found in group 1 TSR (Figure 10). At the bottom of the TSR domain in both groups 1 and 2, the AB loop is stabilised by a

disulphide bond to the carboxyl terminus of the C strand. In addition, the most N-terminal tryptophan is poorly conserved in TSRs from group 2.

The NMR solution structures of rat F-spondin TSR1 and TSR4, as well as TRAP-TSR from *P. falciparum* have also been solved (Paakkonen et al. 2006; Tossavainen et al. 2006). There is a similar fold between these group 2 TSR containing proteins and the group 1 TSR2-TSR3 from TSP-1, with the overall core and AB loop superimposing well. Surprisingly there is some increase in flexibility in the BC loop region of F-spondin TSR4 compared to F-spondin TSR1 (Paakkonen et al. 2006).

Glycosylation and Biological Role of TSRs

TSRs undergo two unusual types of glycosylation, *C*-mannosylation and *O*-fucosylation (Hofsteenge et al. 2001; Gonzalez de Peredo et al. 2002). In *C*-mannosylation, the mannose is covalently linked to the indole ring of tryptophan in a *C*-*C* linkage. Depending on the particular TSR, one, two or all three tryptophans found in the WXXWXXW sequence in strand A may be modified. TSR *O*-fucosylation occurs in the putative consensus sequence $CX_{2\text{-}3}\underline{S}/\underline{T}CX_2G$ in the AB loop (Figure 11). This modification is carried out by POFUT2 and requires a properly folded module. Further extension of the fucose by *β*3Glc-T yields *O*-*α*1-Fucβ1,3Glc, which has only been identified in TSRs.



Modified from Tan et al, 2002

Figure 11. Schematic representation of a TSR domain

The three *C*-mannosylated tryptophans are located on strand A (starting from the N-terminus). The Glc*β*1,3Fuc-*O*- disaccharide occurs in the AB loop. The short *β*-sheets in strands B and C are drawn as orange arrows.

Although the *Drosophila* S2 expression system used in the production of TSR2-TSR3 from TSP-1 for crystallography did not allow for *C*-mannosylation modification (Hofsteenge et al. 2001; Tan et al. 2002), the side chains of all tryptophans in TSR2 and TSR3 modules are orientated so that the Cδ1 atoms are exposed (Figure 11). The interleaved arginines are also exposed along the front face of the domain forming a positively charged groove. Tan and colleagues proposed that this molecular surface represents the “recognition face” of the TSR domain. *C*-mannosylation of the exposed tryptophans does not significantly alter the surface electrostatic potential but may enhance the groove in the recognition face (Tan et al. 2002).

TSRs can bind to GAGs, which are repeating units of negatively charged disaccharides found on proteoglycans. Differences in TSRs binding affinities to GAGs may be explained by the overall electrostatic potential on the surface of the particular TSR module. For example, the front face of the ADAMTS4 TSR and that of TSR5 and TSR6 from F-spondin are almost completely covered in positive charges compared with the positive charge groove on TSR2 from TSP-1. The extra density of positive charges in the former TSRs is due to their more basic C strands. The high affinity binding of

ADAMTS4 TSR (aggrecanase-1) to the GAG side chains of aggrecan (the major proteoglycan in cartilage) appears to be necessary for the enzymatic cleavage of aggrecan during pathogenic cartilage destruction (Tortorella et al. 2000). Similarly, the binding of F-spondin to the extracellular matrix and neuron growth cones involves GAGs and is mediated by the positively charged TSR5 and TSR6 domains (Klar et al. 1992; Tzarfati-Majar et al. 2001). In the case of the malarial parasite, *P. falciparum*, there are several TSR containing proteins including CSP, TSRP and TRAP, which mediate host cell invasion. CSP and TSRP are involved in the attachment and entry of sporozoites to liver cells, respectively (Rathore et al. 2002; Labaied et al. 2007), whereas TRAP binds to GAGs on the surface of salivary glands in the mosquito and liver hepatocytes in the vertebrate host (Matuschewski et al. 2002) and is essential for sporozoites gliding motility (Sultan et al. 1997). The solution structure of TRAP-TSR indicates a front surface of N-terminal positive charges and C-terminal negative charges which may account for its weak binding to heparin. Hence, TRAP probably depends on its neighbouring vWA domain for its high affinity binding to heparin (Tossavainen et al. 2006).

There is a twist between TSR2 and TSR3 of TSP-1 due to a proline residue between the two domains. This results in a 180° turn so that TSR2 and TSR3 face in opposite directions (Tan et al. 2002). Comparing the alignment of TSRs in Figure 10 suggests that a similarly positioned proline occurs more frequently in the group 1 multi-TSR containing proteins than those belonging to group 2. The functional implications of changing the orientation of the TSR recognition face complete with its glycosylation modifications remains to be determined.

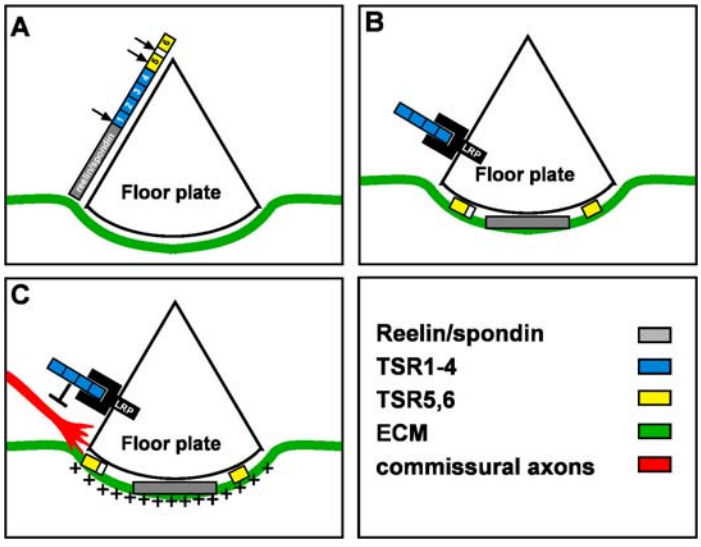
TSRs mediate many of the biological effects of TSP-1 and TSP-2 including the regulation of angiogenesis, wound healing, cell adhesion and migration (Lawler 2000). The anti-angiogenic effect of TSP-1 and TSP-2 occurs via its interaction with the endothelial cell surface receptor CD36 and the subsequent induction of apoptosis (Dawson et al. 1997; Jimenez et al. 2001). The CSVTCG sequence (derived from TSR2 and TSR3 of TSP-1) was identified as the binding region (Asch et al. 1992). These inhibition studies used synthetic peptides, yet this sequence naturally occurs in TSRs with the two cysteines involved in separate disulphide bonds and the threonine modified with the Glc*β*1,3Fuc disaccharide. Therefore, the different glycoforms of thrombospondin TSR modules should be analysed for a binding effect on CD36. TSP-1 can also activate TGFβ. The WSHWSPW and RFK sequences in TSP-1 TSR2 mediate the binding and activation of TGFβ, respectively (Young et al. 2004a; Young et al. 2004b). As the first two tryptophans in the WSHWSPW sequence are fully *C*-mannosylated in the native molecule (Hofsteenge et al. 2001) it would be worthwhile testing if these small sugar additions in the TSR module alter its binding and subsequent activation of TGFβ.

One member of the ADAMTS family, ADAMTS13, is a plasma zinc-containing metalloproteinase that cleaves large multimers of von Willebrand factor into smaller less thrombogenic fragments. A lack of ADAMTS13 activity leads to the life-threatening condition thrombotic thrombocytopenic purpura (Soejima et al. 2005). Six of the eight TSRs in ADAMTS13 are modified with the Glc*β*1,3Fuc disaccharide (Ricketts et al. 2007). Mutation of any of the modified serine sites to alanine reduced the expression of ADAMTS13 in mammalian HEK293 TREx cells. No effect on von Willebrand factor cleavage was observed with the mutants. The secretion defects were

more pronounced when more than one *O*-fucosylation site was mutated. Curiously, over-expression of an active but not an inactive form of POFUT2, could rescue the secretion defect of an ADAMTS13 protein that had mutations in two of its TSR *O*-fucosylation sites (Ricketts et al. 2007). The secretion of another ADAMTS superfamily member, ADAMTS-like-1/punctin-1, which lacks metalloprotease activity but contains four TSRs, was similarly regulated by *O*-fucosylation (Wang et al. 2007).

F-spondin is an extracellular matrix protein that consists of a reelin and spondin domain at the N-terminus and six TSR domains at the C-terminus (Klar et al. 1992; Wang et al. 2007). TSR1-4 are *O*-fucosylated and TSRs1-5 are *C*-mannosylated (Gonzalez de Peredo et al. 2002). Between the fifth and sixth TSR domain a non TSR-element is present which is targeted for proteolysis by the serine protease plasmin (Tzarfati-Majar et al. 2001).

F-spondin is mostly expressed by the floor plate cells of the developing embryo and plays a dual role in the central nervous system. On the one hand it promotes the outgrowth of the commissural axons (Burstyn-Cohen et al. 1999) but on the other hand it inhibits the outgrowth of the motor axons and is involved in restricting the migration path of neural crest cells in somites (Tzarfati-Majar et al. 2001). The proteolytic processing of F-spondin generates four functional protein fragments, the reelin/spondin domain (Burstyn-Cohen et al. 1999), the inhibitory non-adhesive TSR1-4 and adhesive TSR5 and TSR6. The reelin/spondin domain, TSR5 and TSR6, bind the extracellular matrix, which underlies the floor plate, whereas the TSR1-4 binds to the apical floor plate cells via lipoprotein receptor-related protein receptors. The TSR1-4 attachment prevents the penetration of commissural axons into the floor plate cells whereas the adhesive TSR5 and 6 domains permit it (Zisman et al. 2007) (Figure 12).



Taken from Zisman et al., 2007

Figure 12. The role of F-spondin in midline crossing

F-spondin is also expressed in developing peripheral nerves and it is likely to play role in axonal regeneration after nerve injury (Burstyn-Cohen et al. 1998). F-spondin can also act as an extracellular linker protein between another member of the low-density lipoprotein receptor family,

ApoEr2, and amyloid β-precursor protein. The reelin/spondin domains bind to the extracellular domain of amyloid *β*-precursor protein and inhibit its cleavage by *β*-secretase (Ho et al. 2004), while the TSR1-4 bind to the ligand binding domain of ApoEr2 (Hoe et al. 2005). This clustering of the two transmembrane proteins (ApoEr2 and amyloid *β*-precursor protein) by full-length F-spondin results in increased amyloid *β*-precursor protein levels but decreased processing and A*β* levels (Hoe et al. 2005).

In summary, the various biological activities of many TSR containing proteins resides within their TSR domains. The next challenge is to determine if the unusual sugar structures found in these modules regulates their biological activity in an analogous manner to Notch glycosylation.

6. RESULTS

6.1. Large Scale Production of Rat F-Spondin TSR4 and TSR4-Fucose

Introduction

Three different types of protein domains are known to be *O*-fucosylated directly on the protein backbone: EGF domains; TSRs; and PMP-C. The glycan biosynthesis pathways have been elucidated for EGF domains and more recently for TSRs. The fucosylation reactions on EGF repeats and TSRs are performed by protein *O*-fucosyltransferase 1 (POFUT1) and protein *O*-fucosyltransferase 2 (POFUT2), respectively (Luo et al. 2006a; Luo et al. 2006b). Although POFUT1 and POFUT2 share 22% amino acid sequence identity, they are specific for their substrates and do not cross-talk between the two pathways (Figure 5) (Luo et al. 2006a). Interestingly, both enzymes are localised in the ER and they both require a properly folded substrate for their activity (Wang et al. 1998; Luo et al. 2006a). The second step in the EGF glycosylation pathway is performed by a *β*1,3-*N*-acetylglucosaminyltransferase, Fringe, of which there are three homologues in vertebrates (Johnston et al. 1997). One of the aims of this thesis was the identification of the *β*1,3-glucosyltransferase acting on fucosylated TSRs.

To achieve this aim, a sensitive enzyme activity assay was needed to monitor *β*3Glc-T activity. Prior to the identification of proteins modified with Glc*β*1,3Fuc, the original substrate used in detecting *β*3Glc-T activity was the synthetic small molecule *para*-nitrophenyl-*α*-L-fucose (Moloney et al. 1999). However, it has since been determined that the POFUT2 requires properly folded TSR and thus the *β*Glc-T encounters the properly folded fucosylated form of TSR as its substrate *in vivo*. Therefore, for the optimal *in vitro* detection of *β*3Glc-T activity, milligrams of properly folded and highly pure fucosylated TSR modules were required. Previously in our laboratory TSR4 from rat F-spondin had been studied as a model TSR. A His_6-tagged form of rat F-spondin TSR4 could be expressed in mammalian cells, purified and analysed by LC-MSMS, to reveal modifications with C-mannoses on its tryptophans and the O-Fuc*β*1,3Glc disaccharide on the threonine in the sequence CSVTGC (Gonzalez de Peredo et al. 2002). Attempts to express non-glycosylated TSR4 in bacteria have also been made. These experiments showed that the major fraction of TSR4 generated multimeric complexes due to improper disulfide bridge formation. As cysteine scrambling occurs at neutral pH and the concatemer formation could not be circumvented, a three-step purification protocol has been established to obtain the TSR4 module in the correctly folded state. Subsequently, the TSR4 domain was fucosylated in large amounts and purified in a two-step chromatographic procedure to yield a suitable substrate for the glucosylation reaction.

Materials and Methods

Expression and purification of rat F-spondin TSR4

The rat and human F-spondin TSR4 modules are highly conserved. There is a single conservative amino acid change in the C strand where the rat TSR4 has an alanine and the human a valine, three residues before the 5th cysteine. The bacterial expression construct, pET22b-F-spondin-TSR4-TEV-Myc-His$_6$, encoding a tagged form of rat F-spondin TSR4 was generated as described (Kozma et al. 2006). Jan Hofsteenge optimised the small scale *E. coli* expression and Ni-NTA purification of rat F-spondin TSR4 (Kozma et al. 2006). The large scale bacterial protein expression was performed using eleven batches of eight litres bacterial culture with the help of Jasmin Althaus. Details of the optimised three-step purification scheme are described in Chapter 6.2 (Kozma et al. 2006). Aliquots of purified TSR4 (0.78 mg/ml) were stored at -80 °C. This non-fucosylated TSR4 was used as a substrate for the *in vitro* protein *O*-fucosyltransferase assay, as described in Canevascini et al. (manuscript in preparation).

Preparation and purification of TSR4-fucose

To prepare the substrate for the *β*1,3-glucosyltransferase assay, aliquots of rat F-spondin TSR4 were fucosylated as described (Kozma et al. 2006). The standard assay reaction volume of 25 µl was scaled up 50-400x for the preparation of milligram stocks. To obtain 100% fucosylated TSR4, the reaction was performed for 2 hours at room temperature and stopped by placing on ice. The degree of TSR4 fucosylation was monitored by either reversed-phase HPLC-MS or MALDI-TOF-MS (Kozma et al. 2006).

The 100% fucosylated TSR4 was immediately purified by FPLC using a prepacked 1 ml HisTrap HP (GE Healthcare) affinity column. For optimal binding, the imidazole concentration in the sample was reduced to 20 mM by adding five volumes of 20 mM Tris-HCl, pH 7.4, 500 mM NaCl buffer. The sample was filtered through a 0.22 µm filter (Steriflip, Millipore) and loaded onto the equilibrated HisTrap HP column using either a superloop or a peristaltic pump at a flow rate of 0.3 ml/min. The column was washed with 2 volumes of 20 mM Tris-HCl, pH 7.4, 20 mM imidazole, 500 mM NaCl (buffer A), and the His-tagged TSR4-fucose was eluted stepwise with 4 column volumes of 20 mM Tris-HCl, pH 7.4, 500 mM imidazole, 500 mM NaCl (buffer B) at the flow rate of 0.4 ml/min. The fractions containing TSR4-fucose were pooled and desalted into 20 mM Tris-HCl, pH 7.4, 30 mM NaCl buffer using a 5 ml HiTrap desalting (GE Healthcare) column at a flow rate of 2 ml/min. Fractions containing TSR4-fucose protein were pooled and the concentration determined by SDS-PAGE with purified TSR4 as standard (Kozma et al 2006). The quantification was done by Dr. Reto Portman using the Proteomweaver (Definiens) software. The concentration of purified TSR4 was determined from its absorbance at 280 nm using a calculated molar absorption coefficient of 12962 $M^{-1}cm^{-1}$. Aliquots from the different batches of TSR4-fucose production were stored at -80ºC.

Results

Large scale TSR4 production and purification in E. coli

Properly folded, homogenous F-spondin TSR4 was required for the fucosyltransferase, POFUT2. In the large scale production 22 mg of the TSR4 monomer was recovered after the three purification steps (Figure 13). The overall purification yield was 0.25 mg/litre of culture.

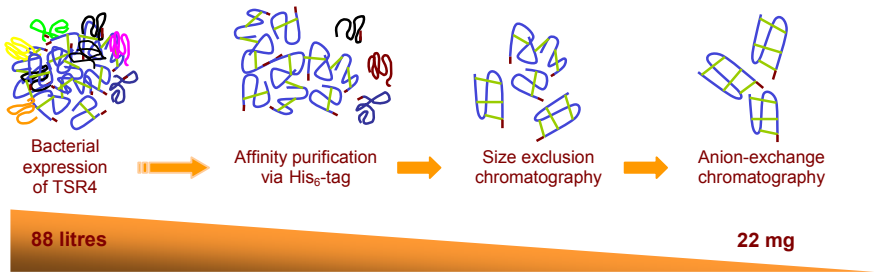


Figure 13. The three-step purification scheme of bacterial TSR4

Following a mechanical lysis of the bacterial cultures, the clarified lysates were incubated overnight with Ni-NTA agarose (for details see Chapter 6.2). The eluates containing TSR4 were pooled and subjected to size-exclusion and ion-exchange chromatographic steps. TSR4 expressed in *E.coli* has the tendency to form inter-molecular disulfide bridges resulting in TSR4 concatemers. In order to separate the monomeric TSR4 from such concatemers, size-exclusion chromatography on a Sephacryl-FF S-200 column was performed (Figure 14).

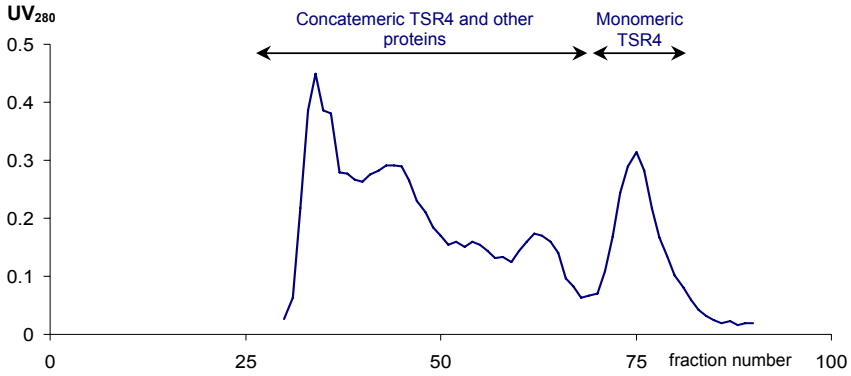


Figure 14. Gel-filtration of TSR4-Myc-His_6 on a Sephacryl FF S-200 column in 50 mM NH_4HCO_3

The major portion (~80%) of the fractions consisted of TSR4-concatemers and contaminating proteins. The underlying cause of the TSR4 concatemer formation may simply be the overproduction of recombinant protein or the spontaneous, incorrect formation of inter-molecular disulfide bridges. The high pH of the Ni-NTA elution may have contributed to the formation of incorrect disulphide bridges as cysteine-scrambling starts to occur at neutral pH (Chang 1994). Refolding of the protein in a redox buffer system could have remedied the disulphide-scrambling and resulted in the formation of

the natural disulfide bonds (data not shown). However, this procedure would have to be performed using a low protein concentration (0.1 mg/ml) requiring a vast buffer volume.

The monomeric TSR4 fractions from the gel-filtration column were lyophilised. Further purification was performed by ion-exchange chromatography on a MonoQ 5/50 (GE Healthcare) FPLC column in 20 mM Tris-HCl, pH 8.0 (Figure 15A).

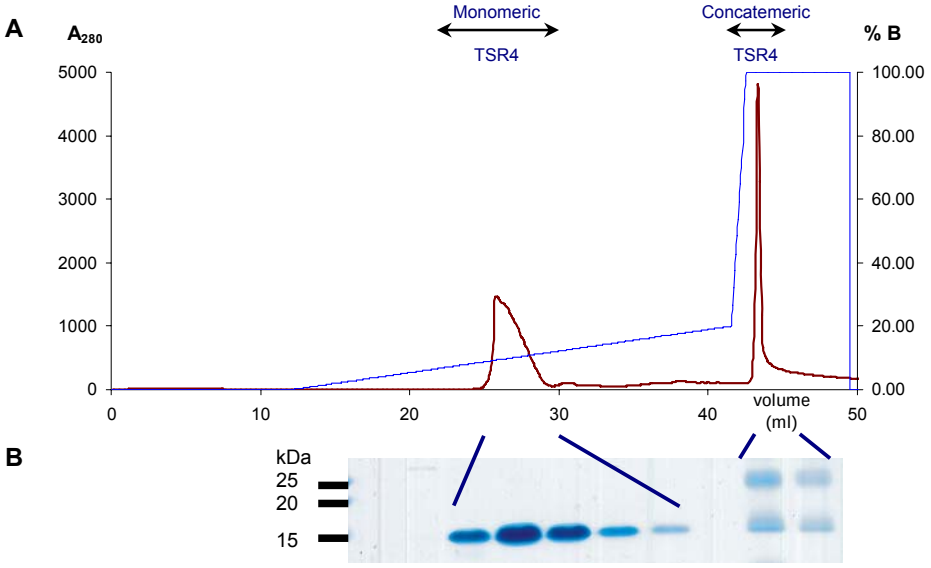


Figure 15. Purification of gel-filtered TSR4 by anion-exchange chromatography

(A) A_{280} trace of TSR4 on an anion-exchange MonoQ 5/50 FPLC column. TSR4 elutes at ~120 mM NaCl. The monomeric and concatemeric TSR4 fractions are indicated by black arrows. (B) SDS-PAGE analysis of protein fractions (26-30 and 43-44) stained with Colloidal Coomassie Brilliant blue.

The monomeric TSR4 eluted at 120 mM NaCl from the MonoQ column (Figure 15A), migrated as a single band on a non-reducing SDS-PAGE (Figure 15B) and as a single peak on a reverse-phase C_{18} HPLC column (Figure 16). Out of 143 mg of TSR4 F-spondin only 22 mg was recovered in the properly folded form.

An indication for the presence of the three-disulfide bridges is seen from the observed molecular mass of TSR4 (9960 Da, Figure 16), which is the same as the calculated one. To decide whether the protein was in a native-like conformation, we analysed the folding/unfolding behaviour by far-UV CD and fluorescence spectroscopy (Kozma et al 2006). In summary, the purified TSR4 appears to be correctly folded by all the criteria tested including, electrophoretic, chromatographic and spectroscopic analysis (Kozma et al 2006). In addition, the TSR4 proved to be a suitable substrate for the protein *O*-fucosyltransferase POFUT2 (shown below).

Small scale in vitro fucosylation of TSR4

The *in vitro* fucosylation assay was initially established in our laboratory using extracts from HEK293T cells expressing *C. elegans* POFUT2 as the enzyme source. By increasing the amount of enzyme source it is possible to obtain 100% substrate conversion. Reverse-phase HPLC shows the incorporation of the fucosyl residue into the TSR4 module (Figure 16). TSR4 (5 µg) was purified from fucosylation reaction mixtures or control reactions, where the TSR4 was added after the reaction was

stopped, using a RP-C_{18} column. The TSR modules eluting at 45.9 minutes (blue trace in Figure 16) and 46.6 minutes (green trace in Figure 16) were analysed by MALDI-TOF-MS and the masses of the fucosylated (10106 Da) and non-fucosylated (9960 Da) TSR4 modules were observed, respectively, in agreement with the expected masses. The mass difference of 146 Da, corresponds to the mass of a single fucosyl residue. The leading "shoulder" in each HPLC chromatogram is due to oxidized methionine residues in the TSR4, whereas the trailing "shoulder" in the TSR4-fucose trace is caused by one of the components in the protease inhibitor cocktail used in the fucosylation reaction (Figure 16). The presence of the fucosyl residue also shows that the purified TSR4 is correctly folded as it serves as a substrate for *C. elegans* POFUT2 and human POFUT2 (data not shown).

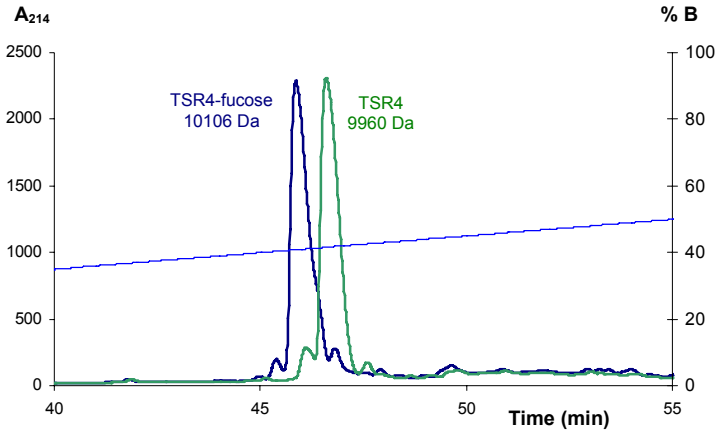


Figure 16. Product proof of TSR4-fucose by RP-HPLC

TSR4 modules were purified, from fucosylation reaction mixtures (blue trace) or control reactions where TSR4 was added to the mix after the reaction was stopped (green trace), using a RP-C_{18} HPLC column. The TSR4-fucose elutes ~40 seconds earlier than the TSR4. The indicated masses of TSR4 and TSR4-fucose were determined by an internally calibrated MALDI-TOF MS (Brucker Daltronics Ultraflex II). The theoretical masses of the TSR4 species were calculated using the protein analysis software GPMAW (Lighthouse Day). Buffer B consists of 80% acetonitrile, 0.085% triflouroacetic acid).

Peptide mapping of TSR4-fucose

In the intact module of TSR4-fucose only the presence of a 146 Da fucose can be deduced by MALDI-TOF-MS but not its position. In order to determine the latter, the HPLC purified TSR4-fucose was digested with endoproteinase Lys-C. This resulted in a mixture of peptides (K1, K3-K6) with the expected masses and a glycopeptide (K2*) which was 146 Da heavier than the non glycosylated K2 (Figure 17). Since K2* peptide contains five hydroxyl residues which can be potentially fucosylated, further cleavage of the peptide was necessary. An endopeptidase Asp-N digest of the K2* glycopeptide yields an eight amino acid-long peptide, ^{14}DCSVTCGK^{21}, with 146 Da greater mass than expected for the unmodified peptide (see red arrow, Figure 18A). By using tandem mass spectrometry we could fragment the ^{14}DCSVTCGK^{21} (+146 Da) peptide and observe the ion for ^{18}TCGK^{21} (+146 Da) (Figure 18B).

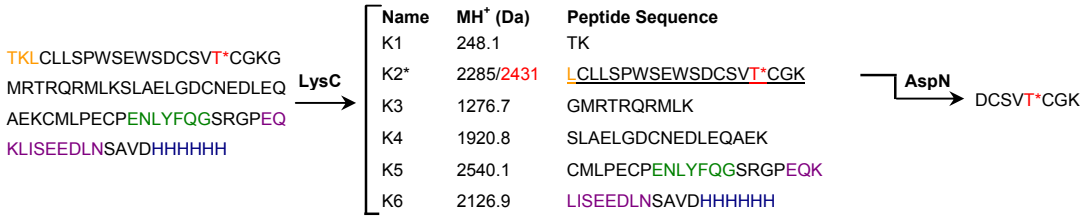
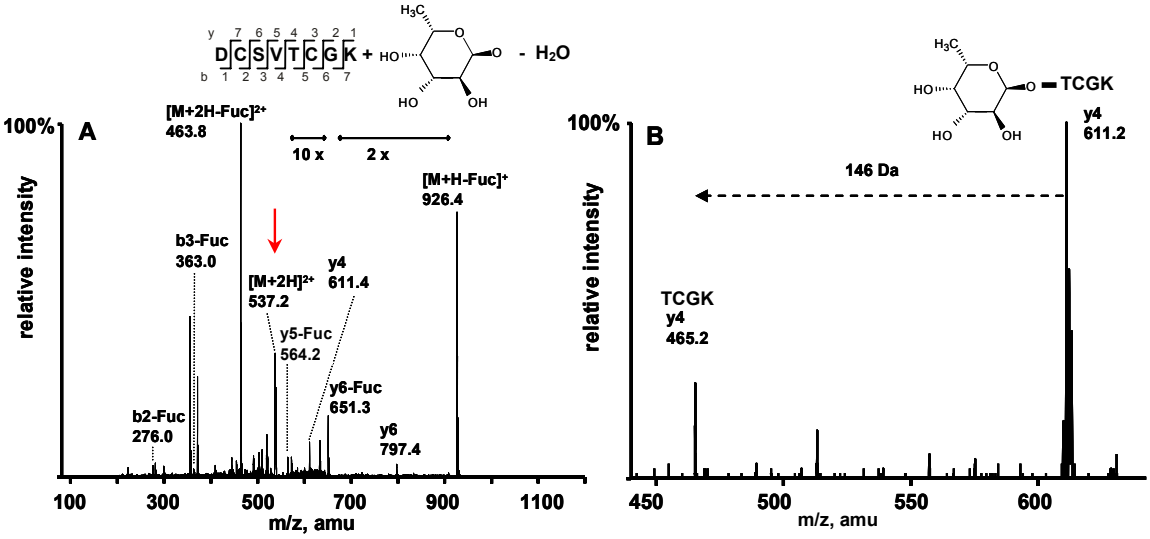


Figure 17. Sequential digest of TSR4-fucose

The intact TSR4-myc-His_6 protein is on the left hand side. Additional sequence from pET22b vector is marked in yellow; theoretical *O*-fucosylation site is in red and an asterisk; tobacco etch virus-site is in green, c-myc epitope is in purple; His_6-tag marked in blue. The two endoproteinase reactions (LysC digest of the intact TSR4; ApsN digest of the glycopeptide (underlined)) are indicated by arrows.

Since Thr-18 is the only hydroxyl residue in this fragment we conclude that it is the site of fucose modification, the same site that is modified in the native protein (Gonzalez de Peredo et al. 2002). This data shows that not only *Drosophila* POFUT2 (Luo et al. 2006b), but also the *C. elegans* homologue can perform the *O*-fucosylation reaction on an isolated TSR module.



Adapted from Canevascini et al.

Figure 18. Identification of the *O*-fucosylation site in TSR4-fucose

The fucosylated peptide (K2*) was purified by RP-HPLC and cleaved with endopeptidase Asp-N. (A) Tandem mass spectrometry of the C-terminal peptide (DCSVTCGK + 146 Da) was performed to confirm its identity. Its double charged 537.2 kDa ion is indicated by red arrow. Ions that lost the fucosyl residue are shown as "-Fuc". (B) Secondary fragmentation of the y4 fragment was performed to verify the site of fucose (146 Da) attachment. The loss of the 146 kDa is shown by dashed arrow.

Large scale TSR4-fucose production and purification

After verifying the fucosylation site on the TSR4 acceptor, TSR4-fucose could be produced in large quantities. For this a reproducible, fast and easy purification method was developed. TSR4 was *in vitro* fucosylated using cell extracts from HEK293T cells expressing *C. elegans* POFUT2 and purified via its His_6-tag followed by a desalting step.

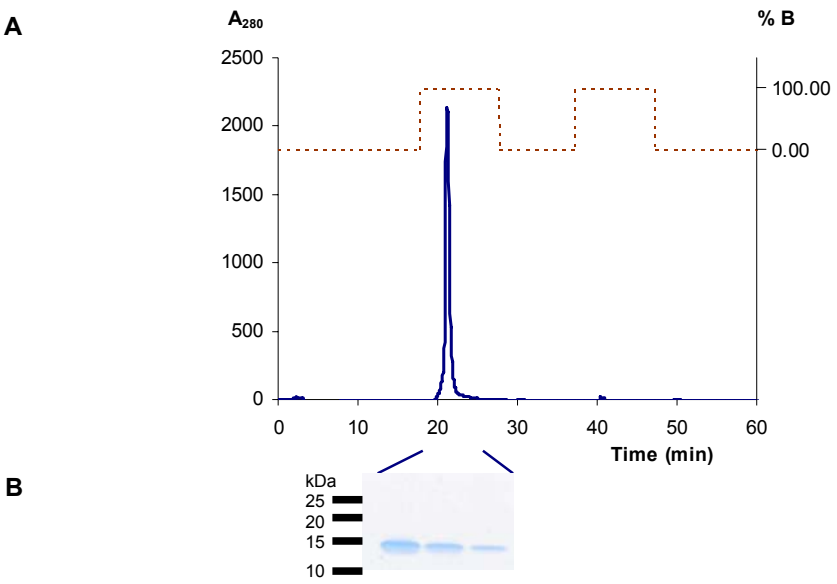


Figure 19. Elution profile of TSR4-fucose from the Ni-NTA affinity column

(A) A_{280} trace of the eluted His-tagged TSR4-fucose. The depicted signal is the subtraction of the signal of the reaction sample and the blank. Elution with 500 mM imidazole is marked by dark red brackets. (B) A non-reducing SDS-PAGE stained with colloidal Coomassie Brilliant Blue of the protein containing fractions.

All the His_6-tagged TSR4-fucose (1 mg batch, Figure 19A) eluted in the first elution step from the Ni-NTA affinity column in highly pure form (Figure 19B).

The imidazole had to be removed from the TSR4-fucose for its application in downstream assays. After this desalting step ~85% of the pure TSR4-fucose was recovered. Three consecutive batches of TSR4-fucose were produced using 1 mg, 2 mg and 3.5 mg of TSR4. In the 3.5 mg batch a small amount of high molecular weight contamination was observed. This is due to the increased amount of His_6-tagged POFUT2 from *C. elegans* present in the reaction mixture. At present, this problem can be overcome as high amount of purified, active FLAG-tagged human POFUT2 is now available in our laboratory.

The secondary structure of TSR4-fucose was characterised by circular dichroism (see Chapter 6.2) and shown to be the same as that of TSR4.

Discussion

The production of pure fucosylated TSR modules is necessary for the *in vitro* measurement of *β*3Glc-T activity, as well as for future kinetic studies on this enzyme. In this chapter the production of highly pure, monomeric and properly folded TSR4 with intact disulfide bonds has been demonstrated. The recombinant TSR4 module could be fucosylated on the same site identified in the full-length F-spondin protein and is thus a suitable substrate for subsequent *in vitro* glucosylation assays. The developed purification protocol is reproducible but very laborious. Discussed below are some areas where the production of TSR may be improved.

Enhancing the bacterial expression levels and purification yields

The overproduction of recombinant proteins often stresses the host cells which can lead to misfolding or aggregation of the expressed protein. In engineered *E. coli* strains overexpressing a combination of six cytosolic chaperones and two small heat shock proteins, a novel two-step strategy was developed which resulted in a 42-fold increase in protein solubility in 70% of the 64 examined proteins (de Marco et al. 2007). Initially, the chaperone system was co-overexpressed with the recombinant target protein. Then the protein synthesis was inhibited, enabling the chaperones to mediate refolding of misfolded and aggregated proteins (de Marco et al. 2007). Addition of the highly soluble NusA domain between the His-tag and the target protein can also increase the solubility and stability of the recombinant protein (de Marco et al. 2004). The introduction of a specific protease recognition site downstream of the NusA, facilitates the cleavage of the target protein from the rest of the construct during purification.

Since the cytoplasm of *E. coli* is a reducing environment, due to the presence of reduced glutathione (Hwang et al. 1995), most of the disulfide bonds do not form properly. As a large proportion of the TSR4 produced in our *E. coli* protein production system was concatemerised (Figures 14 and 15), efforts should be made to avoid the polymerisation process. One possibility would be to prepare a construct encoding TSR4 that targets the protein expression to the periplasm of *E. coli*, enabling a more rapid folding/oxidation of disulfide bridges (Missiakas et al. 1993). Another solution may be to use the *E. coli trxB gor* strain which have a mutation in both the thioredoxin reductase and glutathione reductase genes. This changes the redox potential of the cytoplasm to one that resembles the mammalian endoplasmic reticulum and allows for efficient folding of disulphide bridges (Bessette et al. 1999). The expression levels of certain proteins with complex disulfide bonds patterns, such as human tissue plasminogen activator, was increased 10- to 15-fold in the *E. coli trxB gor* mutant strain compared to the wild type strain (Bessette et al. 1999). The yield of functional Fab antibody in this mutant strain is further enhanced when co-expressed with cytoplasmic chaperones (Levy et al. 2001).

Applying one of these options may results in larger amounts of properly folded TSR4 leading to a less complex and faster purification scheme.

Changing the expression system

For the production of milligram quantities of protein in a short time, the expression in *E. coli* is the simplest option. Nevertheless, there are alternative expression systems, such as insect or yeast cell lines, that may produce high yields of correctly folded proteins.

When TSR1-3 from human TSP-1 was expressed in *Drosophila* S2 cells, the modules were only modified with *O*-fucose and not further elongated to Glc*β*1,3Fuc (Hofsteenge et al. 2001; Tan et al. 2002). The S2 expression system was apparently efficient for the production of these TSR modules (24 µg/ml) and combined with the simple purification scheme (Miao et al. 2001) lead to the solving of the TSR crystal structure (Tan et al. 2002). Encouraged by these results, the initial attempts of large scale production of properly folded TSR4-fucose were investigated in S2 cells. Although most of the expressed TSR4 was in a monomeric form, the levels of protein expression (0.5 µg/ml) and poor purification yields meant that the S2 expression system was not pursued further.

Recently, Chigira and colleagues. engineered a mammalian *O*-fucosylation system in *Saccharomyces cerevisiae* (Chigira et al. 2008). In their system, the *O*-fucosylation pathway was reconstituted with genes encoding an EGF domain, the protein *O*-fucosyltransferase 1 (OFUT1) and two genes whose product converts GDP-mannose to GDP-fucose. The production level of the EGF domain in the engineered yeast was 5 mg/litre culture. The recombinant module was modified with *O*-fucose, and could be further elongated with a GlcNAc sugar by the introduction of the Fringe gene into the engineered *in vivo* O-fucosylation system, suggesting that the module was correctly folded. The ratio between the *O*-fucosylated and non-glycosylated EGF domain was estimated to be 9:1 in the culture media after four days. It would be worthwhile to test a similar system geared for the *in vivo* production of TSR *O*-fucosylation. If similar yields of TSR and fucosylated TSR modules could be produced in one system *in vivo*, then the *in vitro* fucosylation step of the TSR module could be omitted. After the purification of the tagged protein, the non glycosylated and glycosylated forms could be separated by preparative HPLC as shown in Figure 16.

Changing the TSR module

A search of the Pfam database (Finn et al. 2006) in April 2008, identified 431 TSRs in 141 human proteins. However, only ~5% of the TSRs from five proteins, including F-spondin, have so far been identified with the Glc*β*1,3Fuc modification. Previous work in our laboratory showed that the individual TSR4 module from F-spondin could be expressed in mammalian cells and the Glc*β*1,3Fuc disaccharide easily analysed by LC-MSMS. Hence, this protein module was considered as a good *in vitro* acceptor substrate for the POFUT2 and *β*3Glc-T radiochemical activity assays. Using TSR4 from F-spondin as a model TSR we can begin to ask questions about some of the requirements for the *O*-fucosylation and *β*3Glc-T reactions. Besides the sequence motif, $CX_{2\text{-}3}\underline{S}/\underline{T}CX_2G$, the correct three-dimensional structure of the AB loop is essential for *O*-fucosylation. For example, insertion of two glycine residues before the modification site abolishes *O*-fucosylation when expressed in HEK 293T cells (Klein et al., unpublished observations). The results presented in this chapter show that the bacterial expression of TSR4 from F-spondin yields a correctly folded non-glycosylated protein module that can be efficiently *O*-fucosylated *in vitro*. This indicates that C-mannosylation, which occurs on the A-strand of TSRs, is not required at the single module level, for the *O*-glycosylation

modifications in the AB loop. We do not know if it is possible to synthesise a minimal TSR AB loop, stabilised by cysteine bridges, which may act as a suitable acceptor for POFUT2 and *β*3Glc-T.

It appears that individual TSR domains are expressed at different levels in various mammalian expression systems. For example, TSR4 from properdin is expressed at much greater levels than either TSR1 from properdin or TSR4 from F-spondin in mammalian cells (Keusch J.J., unpublished observations). As TSR modules can vary in their sequence, length and disulphide pattern, it is possible that their expression levels may also change in other expression systems.

6.2

Identification and Characterization of a β1,3-Glucosyltransferase That Synthesizes the Glc-β1,3-Fuc Disaccharide on Thrombospondin Type 1 Repeats*

Received for publication, June 20, 2006, and in revised form, September 6, 2006 Published, JBC Papers in Press, October 10, 2006, DOI 10.1074/jbc.M605912200

Krisztina Kozma‡, Jeremy J. Keusch‡, Björn Hegemann‡[1], Kelvin B. Luther§, Dominique Klein‡, Daniel Hess‡, Robert S. Haltiwanger§, and Jan Hofsteenge‡[2]

From the ‡Friedrich Miescher Institute, Maulbeerstrasse 66, CH-4058 Basel, Switzerland and §Department of Biochemistry and Cell Biology, Institute for Cell and Developmental Biology, Stony Brook University, Stony Brook, New York 11794-5215

Thrombospondin type 1 repeats (TSRs) are biologically important domains of extracellular proteins. They are modified with a unique Glcβ1,3Fucα1-*O*-linked disaccharide on either serine or threonine residues. Here we identify the putative glycosyltransferase, B3GTL, as the β1,3-glucosyltransferase involved in the biosynthesis of this disaccharide. This enzyme is conserved from *Caenorhabditis elegans* to man and shares 28% sequence identity with Fringe, the β1,3-*N*-acetylglucosaminyltransferase that modifies *O*-linked fucosyl residues in proteins containing epidermal growth factor-like domains, such as Notch. β1,3-Glucosyltransferase glucosylates properly folded TSR-fucose but not fucosylated epidermal growth factor-like domain or the non-fucosylated modules. Specifically, the glucose is added in a β1,3-linkage to the fucose in TSR. The activity profiles of β1,3-glucosyltransferase and protein *O*-fucosyltransferase 2, the enzyme that carries out the first step in TSR *O*-fucosylation, superimpose in endoplasmic reticulum subfractions obtained by density gradient centrifugation. Both enzymes are soluble proteins that efficiently modify properly folded TSR modules. The identification of the β1,3-glucosyltransferase gene allows us to manipulate the formation of the rare Glcβ1,3Fucα1 structure to investigate its biological function.

Protein glycosylation is a common co- and post-translational modification that has important biological functions at the cellular and organism level. In eukaryotes, *O*-glycosylation is the most diverse class of protein glycosylation as any amino acid with a hydroxyl group (Ser, Thr, Tyr, hydroxylysine, and hydroxyproline) can be modified with one of eight different monosaccharides (GalNAc, GlcNAc, Gal, Glc, Man, Fuc, xylose, arabinose) in either an α or β configuration (1). The unusual protein *O*-linked fucosylation has been described on two different extracellular protein domains, epidermal growth factor-like (EGF)[3] repeat (2–8) and thrombospondin type 1 repeat (TSR) domains (9, 10).

These domains are cysteine knot motifs of between 30–40 and ~60 amino acids in length for an EGF repeat and TSR, respectively. The TSR superfamily contains different extracellular matrix and transmembrane proteins involved in cell-cell and cell-matrix interactions (11). There are TSRs present in 41 different human proteins and 27 different *Caenorhabditis elegans* proteins including thrombospondin-1 (TSP-1), multiple ADAMTS proteases, F-spondin, and UNC-5 (12). Significantly, the TSR domain within these diverse proteins has been shown to be the site of protein-protein or protein-carbohydrate interaction and is nearly always essential for function. The crystal structure of TSR2 and TSR3 from human TSP-1 expressed in *Drosophila* S2 cells shows that each domain consists of three antiparallel strands (12). The first is a longer irregular strand containing the three conserved tryptophans, which are known to be *C*-mannosylated (9). This A strand is connected to two regular β-strands, B and C. Three disulfide bridges and, alternately, stacked layers of tryptophan and arginine residues define this unique structure. Recently the solution structures of TSR1 and TSR4 domains of rat F-spondin were determined (13).

We have identified three members of the TSR superfamily (human TSP-1, human properdin, and rat F-spondin) as the first proteins that carry the disaccharide Glc-Fuc on Ser/Thr (9, 10). The disaccharide structure observed in TSRs was confirmed as Glcβ1,3Fuc (14), a structure first discovered as an amino acid glycoside in human urine (15). TSR *O*-fucosylation occurs in the putative consensus sequence $WX_5CX_{2-3}S/TCX_2G$ (10). The crystal structure revealed that the Ser or Thr carries a fucosyl residue and is located in the short loop between the A and B strands (12). This modification is synthesized by protein *O*-fucosyltransferase 2 (POFUT2) in *C. elegans* (16), mammalian cells (14), and in *Dro-*

* This work was supported in part by National Institutes of Health Grant GM61126. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Present address: Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, 1030 Vienna, Austria.

[2] To whom correspondence should be addressed. Tel.: 41-61-6974722; Fax: 41-61-6973976; E-mail: jan.hofsteenge@fmi.ch.

[3] The abbreviations used are: EGF, epidermal growth factor-like; Asp-N, endoproteinase of *Pseudomonas fragi* mutant; β3Glc-T, β1,3-glucosyltransferase; B3GTL, β3-glycosyltransferase-like; Fuc, fucose; GalNAc, *N*-acetylgalactosamine; Lys-C, endoproteinase from *Achromobacter lyticus*; MS, mass spectroscopy; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; *m/z*, mass over charge; POFUT, protein *O*-fucosyltransferase; TSP, thrombospondin; TSR, TSP type 1 repeat; TSR4, fourth TSR repeat in rat F-spondin; ER, endoplasmic reticulum; HEK cells, human embryonic kidney cells; PIC, protease inhibitor mixture; LC, liquid chromatography; HPLC, high performance LC.

sophila S2 cells, where it has been shown to be localized within the endoplasmic reticulum (ER) (17).

In an alternative glycosylation pathway, *O*-fucosylation has been identified in a number of proteins containing EGF repeats including the Notch receptors (8) and its ligands (7). This modification is carried out in the ER by protein *O*-fucosyltransferase 1 (POFUT1) (18, 19), a paralogue of POFUT2. The role of POFUT1 is essential in Notch signaling, and mutant mice defective in POFUT1 expression are embryonic lethal with a phenotype similar to Notch1 deficiency (20). We recently reported that an enzymatically inactive mutant of the gene for POFUT2 in *C. elegans* results in an abnormal shape of the anterior gonadal arm (16).

Furthermore, the *O*-fucose on Notch serves as a substrate for a β1,3-*N*-acetylglucosaminyltransferase, Fringe, a known modulator of Notch function (21, 22). The product of the Fringe reaction is then further extended by the addition of a galactose and sialic acid to form the tetrasaccharide (NeuAcα2,6Galβ1,4GlcNAcβ1,3Fuc-α1-O-Ser, NeuAc is *N*-acetyl neuraminic acid) (8, 23). This mature Notch glycoform leads to strong activation of Notch signaling via its interaction with the Delta ligand, whereas there is inhibition of Notch signaling via the Jagged ligand (24). These data demonstrate that the particular *O*-glycosylation in an EGF repeat or TSR module on the protein can have significant functional and biological consequences.

Although there are similarities between these two *O*-fucosylation pathways in that both fucosyltransferases are present in the ER, use GDP-fucose as the donor sugar, and require properly folded acceptor substrate, there does not appear to be any biosynthetic cross-talk (14). In contrast to *O*-fucosylation of the EGF repeat, the second and final step in the TSR *O*-fucosylation pathway involves the addition of a capping β1,3-glucose residue.

As a first step to understanding the functional significance of the Glcβ1,3Fuc disaccharide structure in TSR, we wanted to identify the missing β1,3-glucosyltransferase. A soluble enzyme activity capable of forming a Glcβ1,3Fuc linkage has been reported in cell lines from different species and multiple rat tissues (25). We postulated that the β1,3-glucosyltransferase may share sequence similarity with known β1,3-glycosyltransferase members of family 31 in the carbohydrate active enzymes data base (CAZy) (26). One putative human β1,3-glycosyltransferase, B3GTL, that shares 28% sequence identity to the human radical Fringe was recently reported (27). In this paper we identify and characterize this protein as the missing β1,3-glucosyltransferase and we will refer to it as β3Glc-T.[4] This enzyme is localized in the ER compartment of HEK 293T cells and shows high specificity toward TSR-fucose in comparison to fucosylated EGF repeat *in vitro*.

EXPERIMENTAL PROCEDURES

Materials—UDP-glucose, UDP-galactose, UDP-*N*-acetylglucosamine, UDP-*N*-acetylgalactosamine, and GDP-β-L-fucose were purchased from Sigma-Aldrich. UDP-[6-^{3}H]glucose (60 Ci/mmol) and GDP-[2-^{3}H(N)]fucose (17.5 Ci/mmol) were from American Radiolabeled Chemicals (St. Louis, MO) and PerkinElmer Life Sciences, respectively. Oligonucleotides for PCR were synthesized by Microsynth (Balgach, Switzerland). Oasis® HLB cartridges (30 mg) were purchased from Waters Associates (Rupperswil, Switzerland). Antibodies used in Western blots included mouse anti-myc (Sigma-Aldrich), rabbit anti-calnexin (Abcam, Cambridge, UK), and horseradish peroxidase-conjugated secondary antibodies (GE Healthcare). All other chemicals were of the highest quality available.

Subcloning of β3Glc-T and Site-directed Mutagenesis—We identified a full-length cDNA clone of 3467 bp (accession number BC068595, IMAGE 4837250, obtained from Geneservice, Cambridge, UK) that has a single base substitution of an A to a G at position 852 in its 1497-bp coding sequence compared with the B3GTL sequence (accession number AY190526) (27). This results in a conservative change from an isoleucine to a methionine at position 284. The open reading frame encoding the 498-amino acid-long human β3Glc-T (protein ID AAH68595.1) was amplified by PCR from the plasmid pBluescript-B3GTL (IMAGE 4837250) with the forward primer (5′-TGCTCTAGAACCACCATGCGGCCGCCCGCCTGCTG-3′) and the reverse primer (5′-GCGGATCCTAACTCCTCTCGAAAACCTTTCTG-3′) to create XbaI and BamHI restriction sites (underlined) at the 3′- and 5′-end of the gene, respectively. The PCR product was subcloned into pcDNA3.1^{TM} (−)-myc-His A (Invitrogen) to give the plasmid pcDNA3.1-β3Glc-T-Myc-His. The ^{349}DDD^{351} motif in β3Glc-T was mutated to ^{349}ADD^{351} and to ^{349}ADA^{351} by inverse PCR using *Pfu* DNA polymerase (Promega, Wallisellen, Switzerland) with overlapping primers containing the desired point mutations (28). The coding regions of all expression plasmids were sequenced to verify the presence of the desired mutations and the absence of unwanted ones.

Cell Culture—HEK 293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2 mM glutamine, penicillin, and streptomycin. HEK 293T cells that were stably transfected with a plasmid encoding *C. elegans* POFUT1 or POFUT2[5] were grown in the same medium supplemented with 200 μg/ml hygromycin B.

Preparation of Cell Lysates—The plasmids encoding either the wild type or mutant forms of β3Glc-T were transiently transfected into HEK 293T cells using Lipofectamine (Invitrogen) following the manufacturer's instructions. Sixty hours post-transfection the cells were washed and scraped from the plates in ice-cold 10 mM Tris-HCl, 150 mM NaCl, pH 7.5, complemented with EDTA-free Complete® protease inhibitor mixture (PIC, Roche Applied Science). The cells were centrifuged at 170 × *g* for 5 min and resuspended in 10 mM Tris-HCl, 30 mM NaCl, pH 7.4, buffer containing PIC to a density of ~3 × 10^7 cells/ml and disrupted by sonication at 4 °C with three 30-s bursts (35% probe energy, using a 3-mm Microtip from Branson). The lysate was centrifuged at 3320 × *g* for 10 min at 4 °C. The supernatant was ultracentrifuged at 100,000 × *g* for 1 h at 4 °C. The high speed supernatant was collected, separated into

[4] During the revision of this manuscript a paper describing the characterization of B3GTL as a novel β1,3-glucosyltransferase appeared (58).

[5] S. Canevascini, D. Klein, J. Althaus, K. Kozma, R. Chiquet and J. Hofsteenge, manuscript in preparation.

aliquots, and stored at −80 °C until use. Protein concentration was determined using the Bradford Protein Assay (Bio-Rad). Proteins were separated on a 10% SDS-polyacrylamide gel and transferred to nitrocellulose membrane (Bio-Rad). The membranes were developed using the enhanced chemiluminescence (ECL Western blotting, GE Healthcare) kit.

Expression and Purification of TSR4—The protein rat F-spondin TSR4 (residues 615–666 (29)) fused to a tobacco etch virus protease cleavage site followed by a Myc epitope and a His affinity tag (TKLCLLSPWSEWSDCSVTCGKGMRTRQRMLKSLAELGDCNEDLEQAEKCMLPECPENLFQGSRGPEQLISEEDLNSAVDHHHHHH, T indicates Thr at position 18) was prepared by standard PCR methods using a rat F-spondin cDNA as the template (a gift from Dr. A. Klar, Hebrew University, Jerusalem, Israel). The PCR product was subcloned into the bacterial expression plasmid pET22b. Bacteria, strain BL21 (DE3), that had been transformed with pET22b-F-spondin-TSR4-tobacco etch virus-Myc-His were grown to an optical density of 1.0 at 600 nm and induced with 1 mM isopropyl-β-thiogalactopyranoside for 3 h. The cells were harvested by centrifugation, resuspended in lysis buffer (20 mM Tris-HCl, pH 7.5, containing 300 mM NaCl, 0.5% Tween 20, PIC, and 10 mM imidazole), and lysed in a French pressure cell. The lysate was cleared by centrifugation and loaded onto nickel nitrilotriacetic acid beads (Qiagen). The resin was washed in lysis buffer with 20 mM imidazole, and the TSR4 was eluted in lysis buffer containing 500 mM imidazole. The protein was gel-filtered on a Sephacryl FF S-200 column in 50 mM NH_4HCO_3, and the fractions containing monomeric TSR4 were lyophilized. Final purification was achieved by ion-exchange chromatography on a Mono Q 5/50 fast protein liquid chromatography column (GE Healthcare) in 20 mM Tris-HCl, pH 8.0. The protein eluted at 120 mM NaCl in a linear gradient of 0–200 mM NaCl over 30 min in the same buffer. The concentration of the purified protein was determined from its absorbance at 280 nm using a calculated molar absorption coefficient of 12962 $M^{-1}cm^{-1}$. The overall yield of the procedure was 0.3 mg/liter of culture. Aliquots were stored at −80 °C.

Characterization of the TSR4 Protein—The molecular mass of purified TSR4 and of TSR4 after treatment with 102 mM iodoacetamide in 500 mM Tris-HCl, pH 8.6, containing 6 M guanidinium-HCl and 10 mM EDTA with and without prior reduction with 54 mM dithiothreitol were determined by LC-MS (9).

Far-UV CD spectra of TSR4 and TSR4-fucose (0.3–0.8 mg/ml) in 20 mM Tris-HCl, pH 7.5, were measured on an Aviv 62 DS or a Shimadzu circular dichroism spectropolarimeter in 1- or 2-mm cuvettes, respectively. The reported data (molar ellipticity mean residue weight, $[\theta]_{MRW}$) are the average of six spectra and have been base line-corrected. Thermal denaturation and renaturation curves were measured in 1 °C steps at the maxima of ellipticity.

Fluorescence spectra of TSR4 (0.06 mg/ml) in the same buffer were measured on an Aminco Bowman Series 2 luminescence spectrometer in a 1-cm quartz cuvette. Excitation was at 295 nm with a band pass of 2 nm, whereas the emission was recorded with a width of 4 nm. The spectra were quantum-corrected, and contributions from the buffer solution were subtracted.

Production of Fucosylated TSR and EGF Repeat—High speed supernatant from sonicated HEK 293T cells that stably express *C. elegans* POFUT2 (25 μg of total protein) was incubated with 20 μM TSR4 and 100 μM GDP-fucose in 10 mM imidazole-HCl, pH 6.9, 5 mM $MnCl_2$ for 2 h at room temperature. The fucosylated product was purified by fast protein liquid chromatography using a HisTrap HP (GE Healthcare) affinity column and desalted. The mass of the TSR4-fucose was 10,106 Da, as determined using an internally calibrated MALDI-TOF-MS (Brucker Daltonics, Ultraflex II) and sinapinic acid as the matrix. Recombinant EGF repeat from human factor VII was produced in *Escherichia coli* and purified as described (30). Complete fucosylation of the EGF repeat (EGF-fucose) was performed in a similar way except that the enzyme source was from HEK 293T cells stably expressing *C. elegans* POFUT1 and 20 μM EGF repeat was used instead of TSR4 as the acceptor substrate. The fucosylated product was purified by reversed-phase HPLC on a C_{18} column (Zorbax, 1.0 × 150 mm, 3.5 μm, Agilent Technologies) with a 30-min linear gradient of 0–30% buffer B (buffer A, 2% CH_3CN and 0.1% trifluoroacetic acid; buffer B, 80% CH_3CN and 0.085% trifluoroacetic acid). The purified glycoproteins were examined by MALDI-TOF-MS using 2,5-dihydroxybenzoic acid as the matrix. The masses of the unmodified and fucosylated EGF repeat were 5688.6 and 5834.4 Da, respectively.

Glucosyltransferase Assays—Reaction mixtures for assaying the glucosyltransferase activity of recombinant human β3Glc-T contained 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM $MnCl_2$, PIC, 2.64 μM UDP-glucose, 0.33 μM UDP-[6-^{3}H]glucose, 10 μM TSR4-fucose, and high speed supernatant (25 μg of total protein) in a volume of 25 μl. After incubation for 1 h at 37 °C the reaction was terminated by adding 200 μl ice-cold 0.2% trifluoroacetic acid. In the case of blank reactions, the acceptor substrate was added after the reactions were stopped. The reaction mixtures were loaded onto Oasis® HLB cartridges (30 mg) that were previously equilibrated in 0.2% trifluoroacetic acid and washed with 1.2 ml of 0.2% trifluoroacetic acid followed by 1.2 ml of 0.2% trifluoroacetic acid, 20% CH_3CN. The glucosylated product was eluted with 600 μl of 0.2% trifluoroacetic acid, 40% CH_3CN. With the unmodified and fucosylated EGF repeats (EGF-fucose) as acceptor substrates, the cartridges were washed with 0.2% trifluoroacetic acid and 0.2% trifluoroacetic acid, 5% CH_3CN, and the product was eluted as above. Incorporation of [^{3}H]glucose into the acceptor substrate was determined by scintillation counting. Product analysis of glucosylated TSR4-fucose by reversed-phase HPLC, alkali-induced β-elimination, gel filtration, and high pH anion exchange chromatography on a Dionex MA1 column was performed as described (17, 31).

Donor Substrate Specificity of β3Glc-T—The glucosyltransferase assay was carried out as above, except that 20 μM concentrations of the different UDP-sugar donors were used. The products were analyzed by reversed-phase HPLC using the C_{18} Zorbax column with a 35-min linear gradient of 0–35% buffer B (buffer A, 2% CH_3CN, 0.1% formic acid; buffer B, 80% CH_3CN, 0.1% formic acid) on an Agilent 1100 HPLC system interfaced

to an upgraded API 300. The product peak in the 280-nm absorbance trace was integrated.

In Vitro Glucosylation of TSR4-fucose and Mass Spectrometric Analysis—The glucosylation reaction was carried out as above, except that 200 μM UDP-glucose was used, and the incubation was for 2 h. The nearly fully glucosylated product was purified by reversed-phase HPLC on the C_{18} Zorbax column with a 30-min linear gradient of 0–48% buffer B using the trifluoroacetic acid system described above. Fractions containing glucosylated TSR4-fucose were neutralized, dried, reduced, and carboxymethylated as described (32), except that the EDTA was omitted from the buffer and digested with endoproteinase Lys-C (20% w/w, Wako) in 50 mM sodium phosphate, pH 8, 700 mM guanidinium-HCl for 12 h at 37 °C. The peptide of interest was isolated by reversed-phase HPLC-MS on a C_{18} column (GraceVydac 1.0 × 230 mm) with a 2-h linear gradient of 0–50% buffer B using the formic acid system, neutralized, dried, and digested overnight at 37 °C with endopeptidase Asp-N (Roche Applied Science) in 50 mM ammonium bicarbonate, pH 8.

Nanospray Mass Spectrometry of O-Fucosylated and Glucosylated Peptide—Static nanoelectrospray mass spectrometry was performed on a quadrupole linear 4000 Q TRAP instrument (Applied Biosystems, Foster City, CA) in the positive ion mode. In the mass spectrometry with collision-induced dissociation experiments the precursor ion was selected in the quadrupole analyzer and fragmented in the collision cell. Masses were analyzed in the linear ion trap. For MS^3 analysis, the daughter ion was isolated in the linear trap and fragmented, and the granddaughter ions were analyzed using Analyst software. The mass accuracy of all measurements was better than 0.3 *m/z* units.

Subcellular Fractionation—Subcellular fractions of wild type HEK 293T cells were obtained by ultracentrifugation on a nonlinear Nycodenz gradient, and marker proteins for ER and Golgi subfractions were determined as described.[6] Fractions were assayed for β3Glc-T activity as described above. POFUT2 activity was determined as described elsewhere.[5]

RESULTS

Production of Properly Folded TSR4 and TSR4-fucose—Previous studies have shown that *O*-fucosylation of EGF repeats (33) and TSRs (14) requires their correct three-dimensional structure. Therefore, we prepared properly folded recombinant TSR4 from rat F-spondin in *E. coli* as a suitable acceptor substrate for the assays described below. The highly purified protein showed a single band both on reducing and non-reducing SDS-polyacrylamide gels that had been stained with colloidal Coomassie Brilliant Blue. Furthermore, a single peak was detected by reversed-phase HPLC (not shown).

F-spondin TSR4 consists predominantly of β-strands connected by turns and contains three disulfide bridges (13). Mass spectrometric analysis of TSR4 in our purified preparations revealed a molecular mass of 9960 Da, which would be consistent with the presence of three disulfides and the absence of free thiol groups (expected mass 9960 Da). This interpretation was confirmed by the observation that the mass of TSR4 did not change upon treatment with iodoacetamide in the presence of 6 M guanidinium-HCl. In contrast, reduction with dithiothreitol before alkylation resulted in a mass of 10,308 Da, consistent with the incorporation of six carboxamidomethyl moieties (not shown).

We examined the secondary structure of TSR4 by spectroscopic techniques that have previously been used to ascertain the correct folding of recombinant TSRs from human TSP-1 (34). At 25 °C the far-UV CD spectrum of TSR4 displayed positive ellipticity with maxima at 212 and 231 nm (Fig. 1*A*). As expected for a fully denatured protein, the spectrum obtained at 81 °C was dominated by strongly negative ellipticity (Fig. 1*A*). Determination of the thermal denaturation curves at either ellipticity maximum yielded very similar transition temperatures, 47 and 48 °C for 212 and 231 nm, respectively (Fig. 1*B*). This indicated that the maxima arise from the same structural motif. Furthermore, the denaturation process was fully reversible based on the observation that the renaturation curves have the same shape as the ones for denaturation. The positive ellipticity of the far-UV CD spectrum of native TSR4 is in agreement with a structure that largely consists of β-strands and turns and contains clustered aromatic amino acids. In fact, the spectra are very similar to that found for properdin, a protein isolated from human plasma that consists of six TSRs (35) and for recombinant TSRs from TSP-1 (34).

The tryptophan side chains in TSRs form alternately stacked layers with arginines and are partially shielded from solvent. We examined the local environment of the tryptophans in folded and unfolded TSR4 by fluorescence spectroscopy. A clear red shift of the emission maximum from 330 to 344 nm was observed (Fig. 1*C*), indicating that the tryptophans move from a rather non-polar environment to a more solvent-exposed environment upon unfolding of TSR4. In summary, the electrophoretic, chromatographic, chemical, and spectral data provide strong evidence that TSR4 in our preparations contains three disulfide bridges and is correctly folded.

TSR4 could be 100% converted into TSR4-fucose (10,106 Da) using GDP-fucose and recombinant POFUT2 from *C. elegans*. Importantly, we determined the modified amino acid residue by peptide mapping and tandem MS analysis to be Thr-18 (TSR4 numbering), the same residue as in the full-length protein (10) (data not shown).[5] The far-UV CD spectrum of TSR4-fucose was very similar to that of TSR4 (Fig. 1*A*), indicating that the *O*-fucosylation did not cause a significant change in secondary structure.

β3Glc-T Transfers Glucose onto the Fucosyl Residue in TSR4-fucose—B3GTL was identified by Heinonen *et al.* (27) as a putative β1,3-glycosyltransferase-like protein. We tested whether the overexpressed protein[7] in HEK 293T cells has activity for TSR4-fucose. Incubation of TSR4-fucose with UDP-[^{3}H]glucose and high speed supernatant from HEK 293T cells overexpressing β3Glc-T resulted in glucose incorporation into the acceptor (Fig. 2). The amount of activity was 4.7-fold higher than the endogenous activity found in cells transfected with the empty control vector. Importantly, this increase fully

[6] J. J. Keusch, T. Smilda, J. Krieg, and J. Hofsteenge, manuscript in preparation.

[7] Please note that there is a single point mutation in the cDNA used here as compared to the clone identified by Henionen *et al.* (27). See "Experimental Procedures" for details.

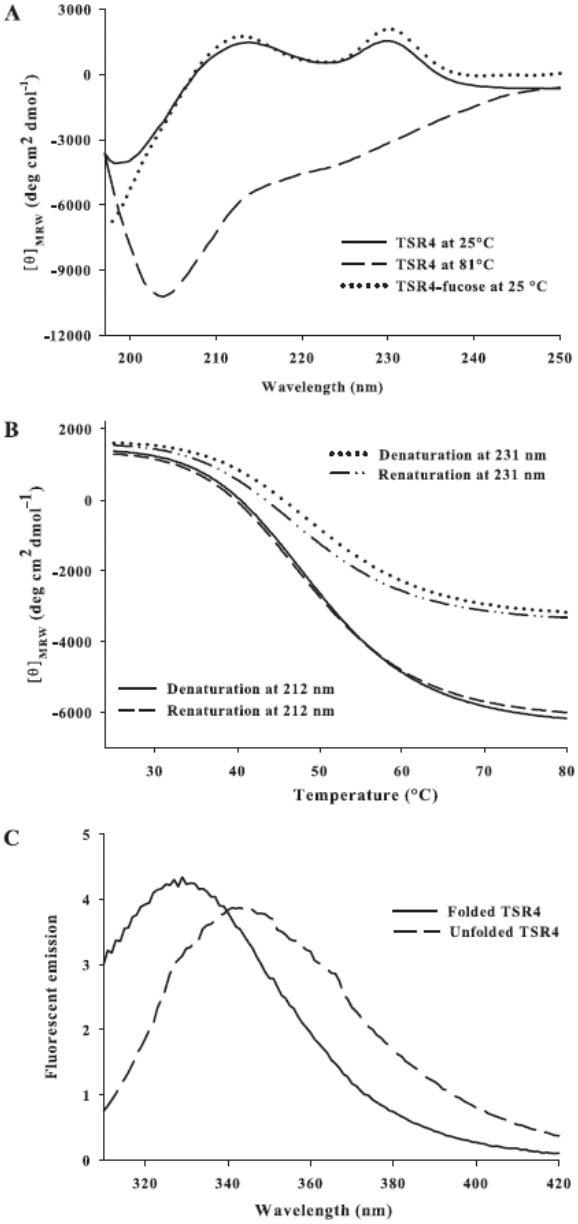


FIGURE 1. **Circular dichroism and fluorescence spectra of folded and unfolded TSR4.** *A*, the far-UV CD spectra of native recombinant TSR4 (*solid line*) and TSR4-fucose (*dotted line*) were measured at 25 °C and that of the completely unfolded TSR4 at 81 °C (*dashed line*). The molar ellipticity data are the average of six spectra. *B*, the thermal unfolding and refolding of TSR4 was monitored at the maxima of ellipticity (212 and 231 nm). The data were fitted to the Hill equation. *C*, the fluorescence spectra of folded TSR4 (*solid line*) and reduced and carboxamidomethylated TSR4 (*broken line*) were determined at 25 °C at an excitation wavelength of 295 nm. $[\theta]_{MRW}$, molar ellipticity mean residue weight.

depended on the presence of a fucosyl residue in the acceptor, since it was not observed with unmodified TSR4 (Fig. 2).

This observation was confirmed by experiments in which we extended the incubation time and increased the concentration

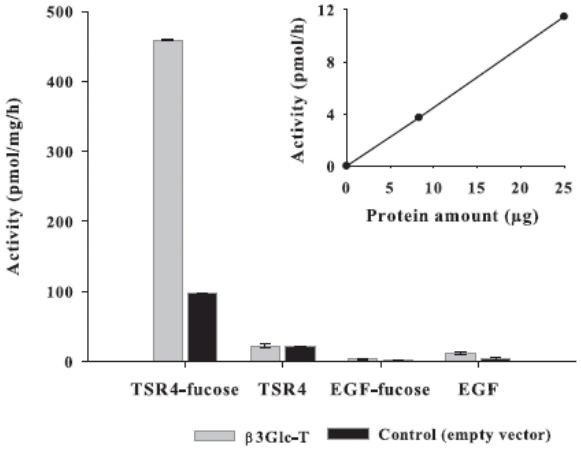


FIGURE 2. **β3Glc-T transfers glucose to TSR4-fucose.** Glucosyltransferase activity of β3Glc-T with either unmodified or fucosylated factor VII EGF repeat or TSR4 as the acceptor substrates (10 μM) was determined with the radiochemical assay described under "Experimental Procedures." The activity in high speed supernatants of HEK 239T cells overexpressing β3Glc-T (*gray bars*) and from cells transfected with empty vector (*black bars*) was assayed. Background activity, determined by adding the acceptor substrate after stopping the reaction (<5%), has been subtracted. Values are the mean (±S.E.) of triplicate determinations from one of three independent experiments. The *inset* shows the linearity of the assay as a function of total protein concentration.

of (unlabeled) UDP-glucose. This resulted in the nearly complete conversion of TSR4-fucose into product exhibiting a mass (10,271 Da, as determined using a triple quadrupole mass spectrometer) in close agreement with that expected for glucosylated TSR4-fucose (Fig. 3*A*, *solid line*). No such product was observed when the acceptor substrate was added after stopping the reaction (Fig. 3*A*, *dotted line*), and comparatively little of it was formed by the endogenous enzyme alone (Fig. 3*A*, *dashed line*). The purified product was used to determine the site of attachment of the glucosyl residue by LC-MS peptide mapping. The peptide containing the Thr-18 had a mass of 2593 Da (Fig. 3*B*), consistent with the presence of a fucosyl and a glucosyl residue. Because this peptide contained multiple hydroxyl amino acids that could carry the glucose it was necessary to obtain a smaller peptide suitable for analysis by tandem MS. Cleavage with endopeptidase Asp-N yielded a peptide with a mass of 1234.5 Da, whose tandem MS spectrum confirmed it to have the sequence DCSV**T**CGK (**T** is Thr-18) and to be modified with fucose and glucose (+308 Da; Fig. 3*C*). Due to the extreme lability of the fucosyl-peptide linkage under the MS-MS conditions of the triple-quadrupole mass spectrometer, the vast majority of the parent ion as well as most of the fragments had lost their 308-Da modification. However, weak ions corresponding to y4 and y5 fragments were observed, suggesting the presence of a disaccharide on Thr-18. We confirmed this by MS^3 analysis of the y4 fragment, which showed a prominent loss of 308 Da and a minor one of 162 Da. This behavior is typical of Glc-Fuc-peptides and has previously been observed in tandem MS experiments with peptides obtained from natural proteins (9). Because Thr-18 is the only hydroxyl amino acid in the y4 fragment, we conclude that a Glc-Fuc disaccharide is attached to this residue.

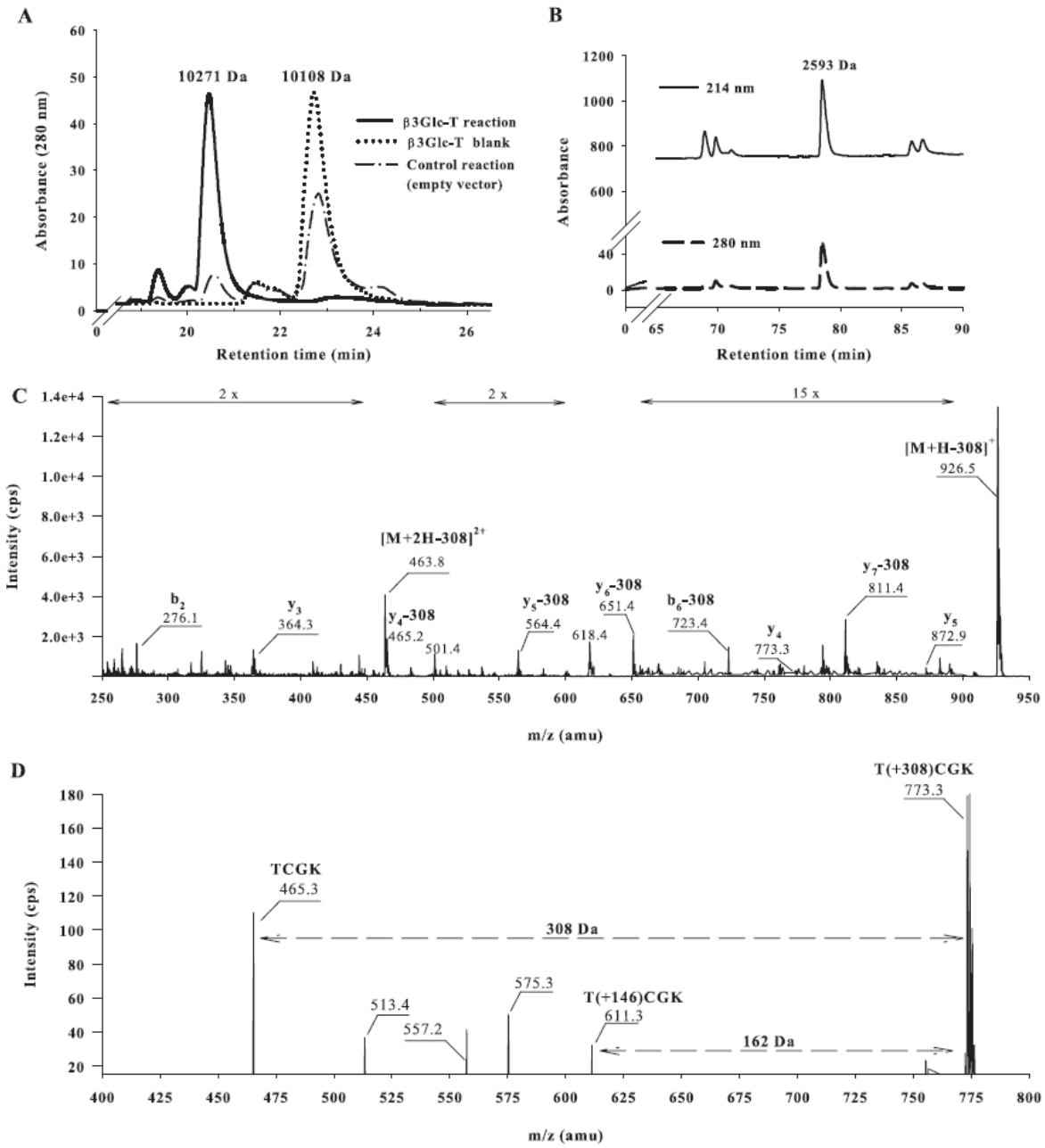


FIGURE 3. **β3Glc-T transfers glucose onto the fucosyl residue in TSR4-fucose.** *A*, purification of TSR4-fucose-glucose on a C_{18} reversed-phase HPLC column (*solid line*). TSR4-fucose (10 μM), modified with fucose on Thr-18, was glucosylated *in vitro* as described under "Experimental Procedures." None or only a small amount of product formation was observed in the enzyme blank (*dotted line*) or control reaction (*dashed line*). *B*, the reaction product eluting at 20.4 min in *panel A* was reduced, alkylated, and cleaved with endoproteinase Lys-C. The digest was fractionated by LC-MS on a reversed-phase C_{18} column. The eluate was monitored by absorbance at 214 and 280 nm. The glycosylated peptide, eluting at 79.5 min, was identified by its mass. *C*, static nanoelectrospray MSMS of the glycosylated subpeptide obtained by cleavage of the 2593-Da peptide with endopeptidase Asp-N. *D*, MS^3 analysis of the y_4 fragment (*m/z* 773) from *panel C*. The losses of Glc (162 Da) and the disaccharide Glc-Fuc (308 Da) have been indicated by *dashed arrows*.

To determine the stereochemistry of the linkage between the glucosyl and fucosyl residue, the disaccharide was released from ^{3}H-glucosylated TSR4-fucose by reductive β-elimination, isolated by gel filtration, and compared with disaccharides of known configuration by high pH anion exchange chromatography (14). This analysis showed a single radioactive peak that

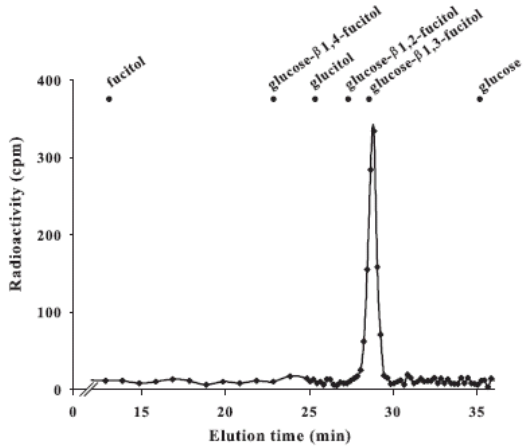


FIGURE 4. **The disaccharide formed by the β3Glc-T-catalyzed reaction is Glc-β1,3-Fuc.** TSR4-Fuc was incubated with UDP-[6-^{3}H]Glc and high speed supernatants of HEK 293T cells overexpressing β3Glc-T. The product of the reaction was purified by reversed-phase HPLC, and the disaccharide was released by alkali-induced β-elimination. The radiolabeled and reduced disaccharide was purified by gel filtration and compared with sugar standards by high pH anion exchange chromatography.

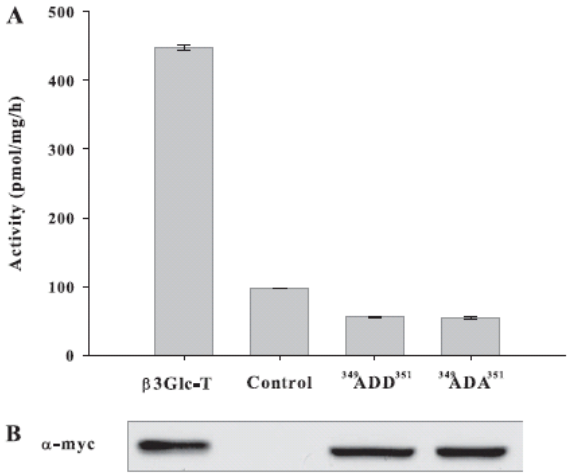


FIGURE 5. **Effect of mutations in the catalytic domain of β3Glc-T.** *A*, wild type and two mutated (^{349}ADD^{351} and ^{349}ADA^{351}) forms of β3Glc-T were expressed in HEK 293T cells. Approximately equal amounts of enzyme, as judged from Western analysis (*B*), were assayed for glucosyltransferase activity with TSR4-fucose (10 μM) as the acceptor. The amount of product was determined with the radiochemical assay as described under "Experimental Procedures." To compensate for the different β3Glc-T protein expression levels, 3-fold more total protein was used in assays from the mutants compared with wild type cell extracts. As a control, high speed supernatant from cells transfected with empty vector was analyzed. The reported values are the mean (±S.E.) of triplicate determinations from one of two independent experiments. Values have been corrected for background (<5%) as described in the legend to Fig. 2.

co-migrated with Glcβ1,3Fucitol (Fig. 4). The absence of [^{3}H]glucitol in the gel filtration profile (not shown) excluded the possibility that a glucosyl monosaccharide is present in addition to the disaccharide. Taken together, the results in Figs. 3, *C* and *D*, and 4 demonstrate that β3Glc-T attaches a glucose residue in a β1,3-linkage to the fucosyl residue on Thr-18 in TSR4-fucose.

Mutation of the DXD Motif Abolishes the Enzymatic Activity of β3Glc-T—To ascertain that β3Glc-T itself contains the glucosyltransferase activity and that the increase in activity we observed in overexpression experiments (Fig. 2) was not resulting from an endogenous transferase, we mutated the DXD motif (^{349}DDD^{351}) in its putative catalytic domain. This sequence occurs in most glycosyltransferases that use nucleotide-activated sugars and is crucial for activity (36–38). Extracts from cells expressing the single (^{349}ADD^{351}) or double mutant (^{349}ADA^{351}) of β3Glc-T did not exhibit glucosyltransferase activity above background level (Fig. 5). This indicates that the increased activity is due to the protein encoded by the transgene.

β3Glc-T Glucosylates TSR-fucose and Strongly Prefers UDP-Glc as the Sugar Donor—As pointed out in the Introduction, EGF repeats can also be modified with an *O*-linked fucosyl residue, which can undergo subsequent elongation reactions. It was, therefore, of interest to examine whether β3Glc-T would transfer glucose to a fucosylated EGF repeat. We prepared and purified the fully fucosylated EGF repeat (EGF-fucose) from human factor VII (5834.4 Da) and tested it as an acceptor substrate for β3Glc-T with UDP-glucose as the donor in the radiochemical assay. No activity was found (Fig. 2).

In addition to *O*-fucosylation, EGF repeats can be core O-glucosylated by an as yet unidentified glucosyltransferase (30). Using EGF repeat from human factor VII as the acceptor substrate, we did not find any evidence for such an activity in β3Glc-T (Fig. 2).

To test the sugar donor specificity of β3Glc-T, we determined its activity in high speed supernatants of transfected

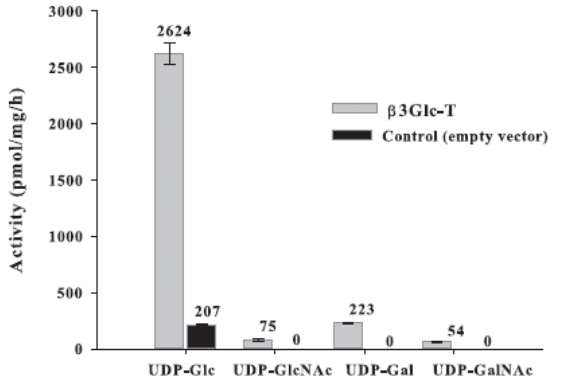


FIGURE 6. **Sugar donor specificity of β3Glc-T.** The glucosyltransferase activity of β3Glc-T in high speed supernatants of transfected HEK 293T cells was determined with different sugar donors (20 μM) and TSR4-fucose (10 μM) as the acceptor. The products of the transferase reaction were quantified by reversed-phase HPLC as described under "Experimental Procedures." High speed supernatant from cells transfected with empty vector was used as a control. The reported values are the mean (±S.D.) of duplicate determinations from one of two independent experiments.

HEK 293T cells with different nucleotide sugars using the HPLC-based assay. As shown in Fig. 6, β3Glc-T has a strong preference for UDP-glucose. There is less than 10% of the activity with UDP-Gal, whereas it is hardly above background with UDP-GlcNAc and UDP-GalNAc.

β3Glc-T Activity Is Localized in the Endoplasmic Reticulum—Contrary to expectation, it has recently been found that POFUT2 is located in the ER (17). Therefore, it was of interest

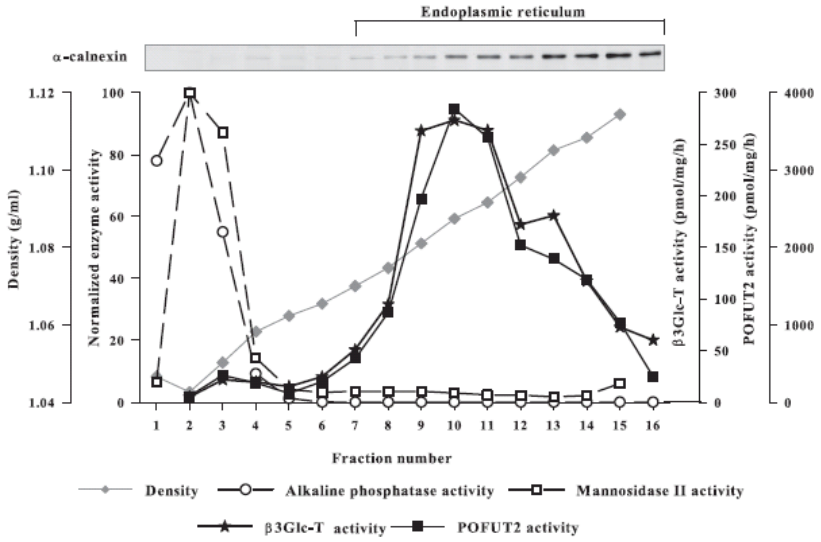


FIGURE 7. **β3Glc-T and POFUT2 activities colocalize in the endoplasmic reticulum.** Isotonic homogenates of wild type HEK 293T cells were fractionated on a non-linear Nycodenz density gradient. The activity of β3Glc-T and POFUT2 in the fractions was determined with the radiochemical assay described under "Experimental Procedures." The position of markers for the Golgi (mannosidase II) and plasma membrane (alkaline phosphatase) subfractions was determined by enzymatic activity, whereas that of ER marker calnexin was determined by Western blot analysis.

to investigate whether β3Glc-T is present in the same subcellular compartment. To avoid possible artifacts of overexpressing the enzyme, we determined the endogenous activity along with that of POFUT2 in cellular subfractions that were obtained from HEK 293T cells by density gradient centrifugation. The gradient that we used has been specifically developed to separate ER, intermediate compartment, and Golgi complex (39). The ER sedimented as a wide peak, as judged by the position of its marker, calnexin (39), that was well separated from Golgi and plasma membranes (Fig. 7). We found the specific activity of both β3Glc-T and POFUT2 in the ER fractions as a double peak. Significantly, the profiles were essentially superimposed. The reason for the double peak is unknown but could be due to separation of the rough and smooth ER.

DISCUSSION

In this report we have identified and characterized β3Glc-T as the β1,3-glucosyltransferase that adds a single glucose from UDP-glucose to a specific α-linked fucose in properly folded TSRs. The closest relatives to β3Glc-T are β1,3-*N*-acetylglucosaminyltransferase (Fringe), core 1 β1,3-galactosyltransferase, and its chaperone COSMC-1, proteins involved in the addition of the second sugar in different protein *O*-glycosylation pathways. The β3Glc-T protein sequence contains 12 amino acids in the putative catalytic domain that are invariant in other members of the CAZy family 31 β1,3-glycosyltransferases (27). This includes the highly conserved D*X*D motif, which is involved in coordinating the divalent cation and UDP-sugar in many different glycosyltransferases (40, 41) and is also necessary for β3Glc-T catalytic activity.

The putative recognition sequences for POFUT1, $CX_{4-5}S/TC$, and POFUT2, $WX_5CX_{2-3}S/TCX_2G$, are clearly different (7, 10), with the modification of the EGF repeats occurring in a β-strand, whereas TSR domains are modified in the loop connecting the A and B strands. The *O*-fucose in an EGF repeat can be extended by one of three Fringe glycosyltransferases (lunatic, manic, and radical) in the Golgi (23, 42). However, the sites are incompletely modified and exhibit distinct effects on Notch function (43). In contrast, almost all the TSR modules expressed in mammalian systems are fully substituted with a Glc-Fuc disaccharide. β1,3-Glucosyltransferase activity was originally identified using the small artificial acceptor *para*-nitrophenyl-α-1-fucose (25). Here, we show that TSR-*O*-α-1-fucose is a natural acceptor for β3Glc-T but not EGF-*O*-α-1-fucose. This suggests that the β3Glc-T recognizes a particular fold rather than a specific amino acid sequence.

We found the β3Glc-T and POFUT2 enzyme activities to be soluble rather than membrane-bound, consistent with other reports (17, 25). Both enzymes appear to have a signal sequence, the first 24 amino acids in the case of β3Glc-T rather than a single transmembrane domain as predicted (27). The *in vitro* POFUT2 and β3Glc-T reactions can be easily driven to completion on fully folded TSR modules. We also observed full occupancy of the glycosylation sites in mammalian protein *in vivo* (9). This and the observation that the activity profiles of POFUT2 and β3Glc-T closely superimpose in ER subfractions (Fig. 7) suggests that the two glycosylation reactions may be tightly coupled *in vivo*. Unusual for a glycosyltransferase, β3Glc-T contains a KDEL-like REEL sequence at its C terminus. We are currently examining whether the ER localization of POFUT2, which lacks such an ER retrieval signal, is dependent on the presence of β3Glc-T.

Because both the POFUT1 and POFUT2 enzymes are localized in the ER and recognize properly folded protein domains, it has been proposed that these glycosyltransferases may be involved in quality control of protein folding (17, 19). The localization of POFUT1 is essential for its function *in vivo*. Furthermore, the ability of POFUT1 to bind to its substrate, Notch, in the ER and to help its folding does not require fucosyltransferase activity (19). Because *O*-fucosylation of a TSR domain requires proper folding (14), β3Glc-T also acts on properly folded TSR. Therefore, if β3Glc-T were involved in quality control of folding, this would not be at the level of the individual TSR domain. Rather it could help in the assembly of the supradomain structure of multiple TSR domains. In this context it is important to note that multiple TSRs often occur in proteins

that contain other types of domains. Some of the most extreme examples occur in the ADAMTS-9 and ADAMTS-20, human metalloproteases, where there are 15 TSRs (44). These are orthologs of GON-1, a protease required for gonadal morphogenesis in *C. elegans* (45).

The *pofut2* gene in *C. elegans* (*pad2*) has been deleted, and we observed an altered shape of the anterior gonadal arm resulting from abnormal distal tip cell migration (16). This is in contrast to a study using RNA-mediated interference where gross morphological changes were reported (46). We are currently investigating whether the phenotype we observed in the *pofut2* null worm is due to loss of functional POFUT2 protein, the fucosyl moiety, or the disaccharide Glcβ1,3Fuc on a particular class of proteins.

Now that the pertinent glycosyltransferases have been cloned, it is also possible to test whether these structures in TSR domains modulate protein-protein interactions. For example, the anti-angiogenic activity of TSP-1 and TSP-2 is mediated through the interaction of their TSR domains with the cell surface receptor, CD36 (47, 48). These studies used synthetic peptides and identified the CSVTCG sequence (derived from TSP-1 TSR2, and TSR3) as the binding region (49). In addition, the TSP-1 and -2/CD36 interaction is antagonized by a histidine-rich glycoprotein that also binds to the CSVTCG sequence (50, 51). Naturally this sequence occurs in TSRs with the cysteine residues engaged in disulfide bonds and the Thr residue modified with the disaccharide Glc-Fuc (9). It is, therefore, worthwhile testing if the glycosylation of the TSR affects the binding to CD36.

TSRs of TSP have been implicated in the activation of TGFβ through their interaction with the inhibitory latency binding protein. Consequently, it is interesting to note that the β3Glc-T gene was identified in a differential display-PCR screen in which T84 human intestinal epithelial cells were induced to differentiate in response to TGFβ (27). It remains to be determined if β3Glc-T activity modulates TGFβ activation by TSP.

It is worth remarking on the changes that have been reported in the level of expression of the two transferases involved in the biosynthesis of the Glcβ1,3Fucα1 disaccharide. The β3Glc-T gene is located on human chromosome 13q12.3, and its promoter region contains several potential binding sites for Smads, proteins that transduce TGFβ signals (27). β3Glc-T is widely expressed in human tissues as two transcripts of 4.2 and 3.4 kilobases whose relative levels differ in a tissue-specific manner (27). Changes in both β3Glc-T and POFUT2 expression have also been found in cancer. Jacques *et al.* (52) observed that the level of β3Glc-T transcript increases in human thyroid cancer, whereas POFUT2 is overexpressed in human glioblastomas (53). Interestingly, the genes for a number of tumor suppressors (BRCA2, RB1, and FLT3) are in close proximity to the β3Glc-T gene (54), and a critical region (marker D13S893) in chromosome 13q12 that excludes the BRAC2 gene but includes the β3Glc-T gene has been implicated in tumorigenicity (55).

Protein *O*-fucosylation is often accompanied by other unusual forms of glycosylation. Within the EGF repeats of factors VII, IX, Notch, and TSP, an *O*-β-linked glucose capped with either one or two Xylα1,3 (Xyl, xylose) residues is found just nine amino acids away from the O-fucosylated glycans (56, 57). Similarly, we reported the disaccharide Glc-Fuc in TSR domains of TSP-1 while studying *C*-mannosylation, another unusual type of glycosylation that is located just five amino acids N-terminal to the *O*-fucosylation site (9). This paper and our previous observations show that glucosylation of fucosylated TSR can occur in the absence of *C*-mannosylation (and vice versa) (9).

Clearly distinct glycosyltransferases act in the EGF repeat and TSR glycosylation pathways. The ability to detect and modify different forms of glycosylation in these biologically important protein domains will help to determine their functional significance.

Acknowledgments—We thank Alexandra Bezler for the initial experiments on the glucosylation of TSR-fucose, Jasmin Althaus for setting up the POFUT2 radiochemical assay and help in the TSR4 purification, and Ragna Sack for help with the MALDI-TOF-MS analysis. The cDNA encoding rat F-spondin was a generous gift from Dr. A. Klar (Hebrew University, Jerusalem, Israel). We are thankful to Drs. T. Ahrens and A. Bachmann (Biocenter, University of Basel, Switzerland) for help with and advice on measuring the CD and fluorescence spectra. The Friedrich Miescher Institute is part of the Novartis Research Foundation.

REFERENCES

1. Spiro, R. G. (2002) *Glycobiology* **12,** 43–56
2. Bjoern, S., Foster, D. C., Thim, L., Wiberg, F. C., Christensen, M., Komiyama, Y., Pedersen, A. H., and Kisiel, W. (1991) *J. Biol. Chem.* **266,** 11051–11057
3. Buko, A. M., Kentzer, E. J., Petros, A., Menon, G., Zuiderweg, E. R., and Sarin, V. K. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3992–3996
4. Harris, R. J., Leonard, C. K., Guzzeta, A. W., and Spellman, M. W. (1991) *Biochemistry* **30,** 2311–2314
5. Harris, R. J., Ling, V. T., and Spellman, M. W. (1992) *J. Biol. Chem.* **267,** 5102–5107
6. Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., and Iwanaga, S. (1992) *J. Biol. Chem.* **267,** 17520–17525
7. Panin, V. M., Shao, L., Lei, L., Moloney, D. J., Irvine, K. D., and Haltiwanger, R. S. (2002) *J. Biol. Chem.* **277,** 29945–29952
8. Moloney, D. J., Shair, L. H., Lu, F. M., Xia, J., Locke, R., Matta, K., and Haltiwanger, R. S. (2000) *J. Biol. Chem.* **275,** 9604–9611
9. Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F., and Peter-Katalinic, J. (2001) *J. Biol. Chem.* **276,** 6485–6498
10. Gonzalez de Peredo, A., Klein, D., Macek, B., Hess, D., Peter-Katalinic, J., and Hofsteenge, J. (2002) *Mol. Cell. Proteomics* **1,** 11–18
11. Tucker, R. P. (2004) *Int. J. Biochem. Cell Biol.* **36,** 969–974
12. Tan, K., Duquette, M., Liu, J. H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J., and Wang, J. H. (2002) *J. Cell Biol.* **159,** 373–382
13. Paakkonen, K., Tossavainen, H., Permi, P., Rakkolainen, H., Rauvala, H., Raulo, E., Kilpelainen, I., and Guntert, P. (2006) *Proteins* **64,** 665–672
14. Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006) *J. Biol. Chem.* **281,** 9385–9392
15. Hallgren, P., Lundblad, A., and Svensson, S. (1975) *J. Biol. Chem.* **250,** 5312–5314
16. Canevascini, S., Kozma, K., Grob, M., Althaus, J., Klein, D., Chiquet-Ehrismann, R., and Hofsteenge, J. (2006) *European Worm Meeting,* April 29–May 3, 2006, p. 71, Hersonissos, Crete, Greece
17. Luo, Y., Koles, K., Vorndam, W., Haltiwanger, R. S., and Panin, V. M. (2006) *J. Biol. Chem.* **281,** 9393–9399
18. Luo, Y., and Haltiwanger, R. S. (2005) *J. Biol. Chem.* **280,** 11289–11294
19. Okajima, T., Xu, A., Lei, L., and Irvine, K. D. (2005) *Science* **307,** 1599–1603
20. Shi, S., and Stanley, P. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5234–5239

21. Haines, N., and Irvine, K. D. (2003) *Nat. Rev. Mol. Cell Biol.* **4,** 786–797
22. Haltiwanger, R. S., and Stanley, P. (2002) *Biochim. Biophys. Acta* **1573,** 328–335
23. Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K. D., Haltiwanger, R. S., and Vogt, T. F. (2000) *Nature* **406,** 369–375
24. Chen, J., Moloney, D. J., and Stanley, P. (2001) *Proc. Natl. Acad. Sci.* **98,** 13716–13721
25. Moloney, D. J., and Haltiwanger, R. S. (1999) *Glycobiology* **9,** 679–687
26. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997) *Biochem. J.* **326,** 929–939
27. Heinonen, T. Y., Pasternack, L., Lindfors, K., Breton, C., Gastinel, L. N., Maki, M., and Kainulainen, H. (2003) *Biochem. Biophys. Res. Commun.* **309,** 166–174
28. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* (*Amst.*) **77,** 51–59
29. Klar, A., Baldassare, M., and Jessell, T. M. (1992) *Cell* **69,** 95–110
30. Shao, L., Luo, Y., Moloney, D. J., and Haltiwanger, R. (2002) *Glycobiology* **12,** 763–770
31. Moloney, D. J., Lin, A. I., and Haltiwanger, R. S. (1997) *J. Biol. Chem.* **272,** 19046–19050
32. Krieg, J., Hartmann, S., Vicentini, A., Gläsner, W., Hess, D., and Hofsteenge, J. (1998) *Mol. Biol. Cell* **9,** 301–309
33. Wang, Y., and Spellman, M. W. (1998) *J. Biol. Chem.* **273,** 8112–8118
34. Huwiler, K. G., Vestling, M. M., Annis, D. S., and Mosher, D. F. (2002) *Biochemistry* **41,** 14329–14339
35. Smith, C., Pangburn, M., Vogel, W., and Muller-Eberhard, J. (1984) *J. Biol. Chem.* **259,** 4582–4588
36. Keusch, J. J., Manzella, S. M., Nyame, K. A., Cummings, R. D., and Baenziger, J. U. (2000) *J. Biol. Chem.* **275,** 25308–25314
37. Munro, S., and Freeman, M. (2000) *Curr. Biol.* **10,** 813–820
38. Wiggins, C. A. R., and Munro, S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7945–7950
39. Hammond, C., and Helenius, A. (1994) *J. Cell Biol.* **126,** 41–52
40. Gastinel, L. N., Cambillau, C., and Bourne, Y. (1999) *EMBO J.* **18,** 3546–3557
41. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J. M. (2000) *EMBO J.* **19,** 5269–5280
42. Rampal, R., Li, A. S., Moloney, D. J., Georgiou, S. A., Luther, K. B., Nita-Lazar, A., and Haltiwanger, R. S. (2005) *J. Biol. Chem.* **280,** 42454–42463
43. Rampal, R., Arboleda-Velasquez, J. F., Nita-Lazar, A., Kosik, K. S., and Haltiwanger, R. S. (2005) *J. Biol. Chem.* **280,** 32133–32140
44. Somerville, R. P., Longpre, J. M., Jungers, K. A., Engle, J. M., Ross, M., Evanko, S., Wight, T. N., Leduc, R., and Apte, S. S. (2003) *J. Biol. Chem.* **278,** 9503–9513
45. Blelloch, R., and Kimble, J. (1999) *Nature* **399,** 586–590
46. Menzel, O., Vellai, T., Takacs-Vellai, K., Reymond, A., Mueller, F., Antonarakis, S. E., and Guipponi, M. (2004) *Genomics* **84,** 320–330
47. Dawson, D. W., Pearce, S. F., Zhong, R., Silverstein, R. L., Frazier, W. A., and Bouck, N. P. (1997) *J. Cell Biol.* **138,** 707–717
48. Jimenez, B., Volpert, O. V., Crawford, S. E., Febbraio, M., Silverstein, R. L., and Bouck, N. (2000) *Nat. Med.* **6,** 41–48
49. Asch, A. S., Silbiger, S., Heimer, E., and Nachman, R. L. (1992) *Biochem. Biophys. Res. Comm.* **182,** 1208–1217
50. Simantov, R., Febbraio, M., Crombie, R., Asch, A. S., Nachman, R. L., and Silverstein, R. L. (2001) *J. Clin. Investig.* **107,** 45–52
51. Simantov, R., Febbraio, M., and Silverstein, R. L. (2005) *Matrix Biol.* **24,** 27–34
52. Jacques, C., Baris, O., Prunier-Mirebeau, D., Savagner, F., Rodien, P., Rohmer, V., Franc, B., Guyetant, S., Malthiery, Y., and Reynier, P. (2005) *J. Clin. Endocrinol. Metab.* **90,** 2314–2320
53. Dong, S., Nutt, C. L., Betensky, R. A., Stemmer-Rachamimov, A. O., Denko, N. C., Ligon, K. L., Rowitch, D. H., and Louis, D. N. (2005) *J. Neuropathol. Exp. Neurol.* **64,** 948–955
54. Dunham, A., Matthews, L. H., Burton, J., Ashurst, J. L., Howe, K. L., Ashcroft, K. J., Beare, D. M., Burford, D. C., Hunt, S. E., Griffiths-Jones, S., Jones, M. C., Keenan, S. J., Oliver, K., Scott, C. E., Ainscough, R. *et al.* (2004) *Nature* **428,** 522–528
55. Cheng, Y., Lung, H. L., Wong, P. S., Hao da, C., Man, C. S., Stanbridge, E. J., and Lung, M. L. (2004) *Int. J. Cancer* **109,** 357–362
56. Harris, R. J., and Spellman, M. W. (1993) *Glycobiology* **3,** 219–224
57. Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., and Iwanaga, S. (1989) *J. Biol. Chem.* **264,** 20320–20325
58. 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*, in press

6.3. Characterisation of β1,3-Glucosyltransferase

Introduction

To further characterise the newly identified β3Glc-T, the following questions were addressed:

1. Is there a complex formed between β3Glc-T and POFUT2?
2. Which residues in β3Glc-T are required for catalytic activity?
3. What role do the N-glycans in β3Glc-T play?

Studies on the complex formation between β3Glc-T and POFUT2

There are a number of studies describing the complex formation between proteins involved in a common carbohydrate biosynthetic pathway. In 1996, Nilsson and colleagues demonstrated a physical association between *N*-acetylglucosaminyltransferase I and mannosidase II, two enzymes involved in the synthesis of complex *N*-glycans (Nilsson et al. 1996). This interaction is mediated by the stem region of the *N*-acetylglucosaminyltransferase I and retains the complex in the Golgi. Complex formation between two glycosyltransferases involved in consecutive steps in glycolipid formation has also been demonstrated (Maccioni et al. 2002). Significantly, an *in vivo* complex between the two homologues POMT1 and POMT2 is required for protein *O*-mannosyltransferase activity (Akasaka-Manya et al. 2006). Interestingly, an association between UDP-galactose:ceramide galactosyltransferase and the UDP-galactose transporter changes the intracellular localisation of this transporter from the Golgi to the ER (Sprong et al. 2003).

Most glycosyltransferases contain transmembrane domains that anchor the proteins in the membrane of a particular subcellular compartment. β3Glc-T and POFUT2 are unusual in this respect as they lack transmembrane domains and co-localise in the soluble fraction of the ER (see Chapter 6.2). Whilst the β3Glc-T contains an ER-retrieval sequence at its C-terminus, there is no such localisation signal in POFUT2. This prompted us to speculate on how POFUT2 is retained in the ER. As the majority of mammalian TSRs are fully modified with the disaccharide Glcβ1,3Fuc *in vivo*, it is plausible that the consecutive enzyme reactions are tightly coupled *in vivo* and thus the POFUT2 may be retained in the ER through an association with β3Glc-T. As an initial step to investigate an interaction between these two glycosyltransferases, co-immunoprecipitation experiments were performed from the conditioned medium of transiently transfected mammalian cells.

Identification of the catalytic residues in β3Glc-T

β3Glc-T consists of 498 amino acids with the first 28 amino acids representing the signal sequence. The mature protein can be divided into two approximately equal domains of around 235 residues. In the C-terminal domain there are 12 amino acids and five motifs that are highly conserved amongst the glycosyltransferases in family GT31 of the CAZy database (Heinonen et al. 2003). The most notable motif is the triplet of aspartates (^{349}DDD^{351}; DXD is the more general motif), which corresponds to the catalytic core. The closest relatives to β3Glc-T are Radical Fringe and core 1 β3-galactosyltransferase with 28% and 27% sequence identity, respectively (Heinonen et al. 2003).

The crystal structure of Manic Fringe, a paralogue of Radical Fringe, has recently been solved (Jinek et al. 2006). Manic Fringe shows a mixed α/β-fold belonging to the GT-A superfamily. As *β*3Glc-T and Manic Fringe share 23% sequence identity and both enzymes act on fucosylated substrates, we expected their catalytic domains to display a similar three-dimensional structure. Figure 20 shows the C-terminal sequence of human *β*3Glc-T threaded onto the 3D-structure of murine Manic Fringe, complexed with Mn^{2+} and a modelled UDP-glucose.

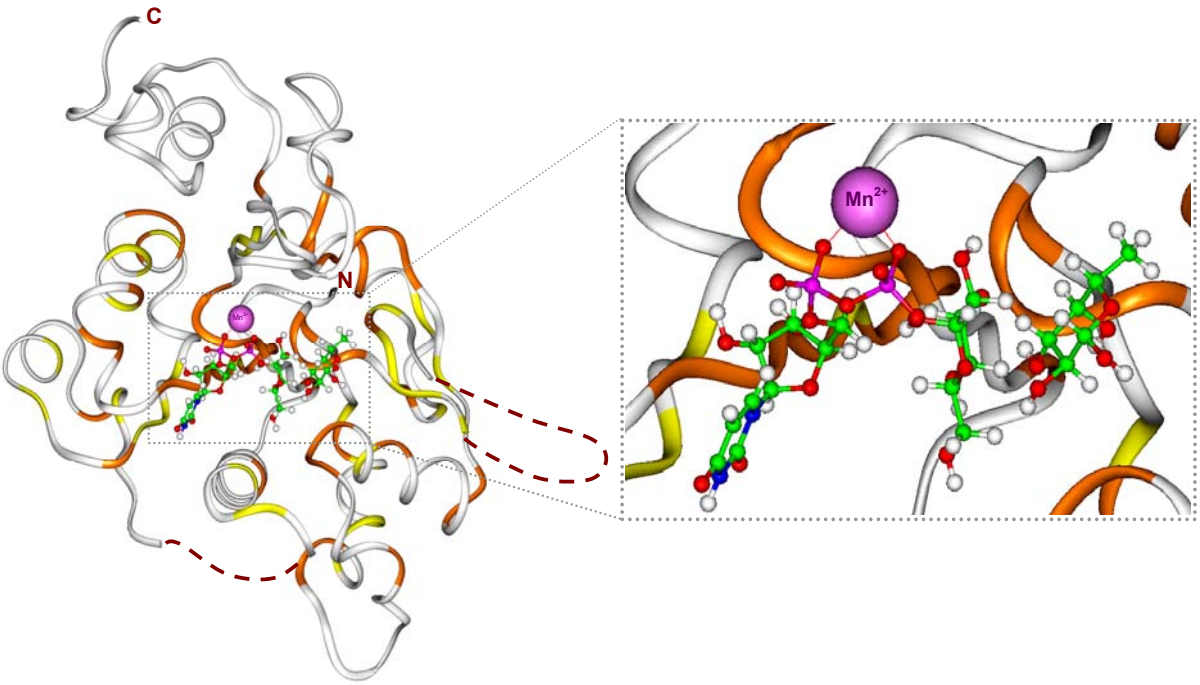


Figure 20. The C-terminal domain of human *β*3Glc-T is modelled on the crystal structure of murine Manic Fringe

The structure of murine Manic Fringe (amino acids 45-321) is shown is white. The C-terminal domain of human *β*3Glc-T (amino acids 264-498) is superimposed on the structure. The conserved residues are shown in orange and residues that are similar are shaded in yellow. In the centre, UDP-glucose is modelled in the catalytic centre. The position of the fucose, which was not experimentally established, has been chosen so that the 3'-OH is close to the C-1 of the glucose. UDP-glucose and the fucosyl residue are shown in ball and stick format and Mn^{2+} as a purple sphere. Disordered loops are indicated with dashed lines. The structure was created by Dr. Jan Hofsteenge using the program Insight II.

The residues surrounding the donor sugar and the divalent cation are conserved. The nucleotide sugar-binding site, ^{142}DDD^{144}, is located in a cleft at the top of the central β-sheet in the core domain of Manic Fringe (Jinek et al. 2006). Asp-144 and the conserved histidine, His-256, coordinate the Mn^{2+}, and thus the α- and β-phosphates of UDP (Figure 21A). Asp-143 contacts two hydroxyl groups of the ribose sugar (Jinek et al. 2006). According to the analysis of the catalytic domain of *β*3Glc-T (Figure 21B) it is likely that its catalytic centre resembles that of Manic Fringe. This is supported by the observation that mutation of Asp-349 and Asp-351 to alanines in the equivalent nucleotide sugar-binding site in *β*3Glc-T, ^{349}DDD^{351}, abolishes enzyme activity (Kozma et al. 2006). In *β*3Glc-T Ala-444 has replaced His-256 in Manic Fringe. The Mn^{2+}-coordination could possibly be taken over by His-456 (Figure 21B).

Studies on the 3D-structures of other inverting GT-A glycosyltransferases indicate that the acceptor sugar binds in a pocket adjacent to the donor sugar-binding site (Pedersen et al. 2000). Asp-232, located in the conserved ^{230}PDD^{232} motif in Manic Fringe, is found in the structurally equivalent position as Asp-191 in SpsA (Charnock et al. 1999) and Glu-281 in glucuronyltransferase (Pedersen et al. 2000), suggesting that this could be the general base for the catalytic mechanism of Fringe (Jinek et al. 2006). In inverting glycosyltransferases, this general base initiates the catalysis by abstracting a proton from the reactive group of the acceptor, allowing the direct nucleophilic attack at the donor sugar *C*1 carbon centre (Unligil et al. 2000). The general base in *β*3Glc-T could be Asp-421, which is in close proximity to the modelled fucose. To determine if Asp-421 in the ^{419}PDD^{421} motif in *β*3Glc-T is important for catalysis, this aspartate residue was changed into alanine and the properties of the expressed mutant enzyme were investigated.

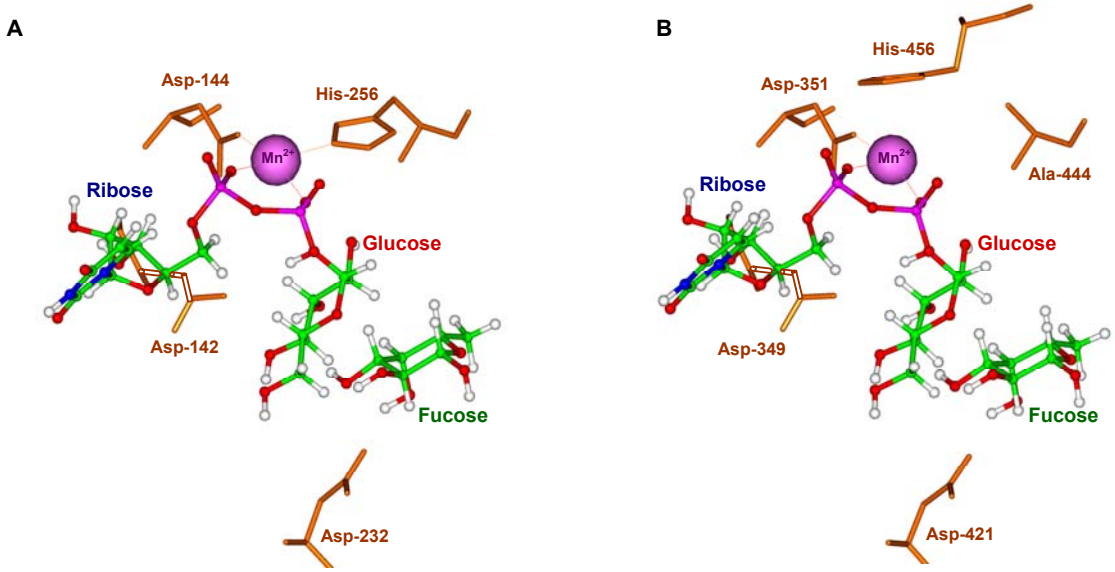


Figure 21. Models of the active sites of murine Manic Fringe and human *β*3Glc-T

(A) Structural model of the crystallised murine Manic Fringe in a complex with Mn^{2+} and UDP. (B) Hypothetical structure of human *β*3Glc-T achieved by the superimposition of its C-terminal sequence onto the structure of murine Manic Fringe. UDP-glucose and a fucosyl residue are modelled in the catalytic centre. Mn^{2+} is shown as purple sphere. The models were done by Dr. Jan Hofsteenge using the program Insight II.

Internal homology between the N- and C-terminal domains of *β*3Glc-T

*β*3Glc-T can be divided into two approximately equal domains, the N-terminal domain containing amino acids approximately 29-263 and the C-terminal catalytic domain with amino acids 264 to 498 (Figure 22).

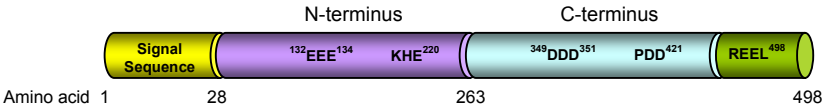


Figure 22. The schematic representation of human *β*3Glc-T

The signal sequence is shaded in yellow, the N-terminal domain in purple and the C-terminal domain in blue. The C-terminal ER-retrieval signal, REEL, is coloured in green. The C-terminus catalytic active site, ^{349}DDD^{351}, and the putative catalytic base, ^{419}PDD^{421}, as well as the corresponding putative catalytic sites in the N-terminal domain are indicated.

Although there is no spatial structural similarity between the N- and C-terminal domains, there is internal sequence similarity (Figure 23). The potential catalytic residues (^{132}EEE^{134} in the N-terminus, ^{349}DDD^{351} in the C-terminus) and the potential catalytic bases (^{218}KHE^{220} in the N-terminus and ^{419}PDD^{421} in the C-terminus) as well as the *N*-glycosylation sites (^{124}N and ^{336}N) are conserved. Eight out of the 12 highly conserved residues in the GT31 family are also found in the N-terminal domain of *β*3Glc-T (indicated with an asterisk).

```
                          *
N-terminus   59  VFVIQSQSNSFHAKRAEQLKK------SILKQAADLTQE-LPSVLLLHQLAKQEGAWTIL  111
                 +FV       FH  R   +K+      S+++  +D T+  +P+V L      +
C-terminus  269  IFVAVKTCKKFHGDRMPIVKQTWESQASLIEYYSDYTENSIPTVDLGIPNTDRGHCGKTF  328

                            *    * *           *           *
N-terminus  112  PLLPHFSVTYSRNSSWIFFCEEETRIQIPKLLETLRRYDPSKEWFLGKALHDEEATIIHH  171
                  +L  F       ++W+   +++T I I +L   L  YD  K  FLG+
C-terminus  329  AILERFLNRSQDKTAWLVIVDDDTLISISRLQHLLSCYDSGKPVFLGE----------R  377

                                    *                       *
N-terminus  172  YAFSENPTVFKYPDFAAGWALSIPLVNKLTKRLKSESLKSDFTIDLKHEIALYIWDKGGG  231
                 Y +      + Y     G   S     +  +RL +   +  ++ D    ++ L +   G G
C-terminus  378  YGYGLGTGGYSYITGGGGMVFS----REAVRRLLASKCRC-YSNDAPDDMVLGMCFSGLG  432

N-terminus  232  PPLTPVPEF  240
                  P+T  P F
C-terminus  433  IPVTHSPLF  441
```

Figure 23. Alignment of the human *β*3Glc-T N- and C-terminal domains

Identical residues are written between the N- and C-terminal alignments. Conserved residues are indicated by a cross. The asterisks show those residues that are also conserved in the family GT31. The potential *N*-glycosylation sites are underlined and the catalytic residues are boxed in grey. The numbers refer to the amino acid position in the full-length *β*3Glc-T molecule. Gaps introduced for the alignment are shown by a dash. The alignment was performed using 'Blast2sequences'.

Modelling the human *β*3Glc-T sequence onto the mouse Manic Fringe structure shows that only the C-terminal domain superimposes (Figure 20). We speculated that if the C-terminal domain is the intrinsic functional glucosyltransferase then a point mutation in the N-terminal domain would not be expected to affect the *β*3Glc-T enzyme activity. To test this hypothesis the glutamic acid in the ^{218}KHE^{220} sequence was mutated to an alanine.

The importance of the *N*-glycosylation sites in *β*3Glc-T

N-linked glycans play important biochemical and physiological roles in the protein they are attached to including thermal stability, solubility, resistance to proteolytic cleavage, biological activity, and *in vivo* clearance rate (Varki 1993). The presence of the Asn-X-Ser/Thr consensus sequence in a protein is necessary but not sufficient for *N*-glycosylation. Occupancy also depends on the location, conformation and residues within and around the acceptor site. When the glycosylation site is occupied, the type of attached glycan may either be a complex, a high-mannose or a hybrid oligosaccharide (Jones et al. 2005).

Most glycosyltransferases are *N*-glycosylated. The lack of *N*-glycosylation site occupancy can result in low enzyme activity, altered subcellular localisation, or decreased enzyme solubility (Martina et al. 1998; Baboval et al. 2000; Christensen et al. 2000; Loriol et al. 2007). However, in other

glycosyltransferases the disruption of their *N*-glycosylation by mutagenesis of the attachment site had no effect on activity (Malissard et al. 1996; Minowa et al. 1998; Jinek et al. 2006).

Human *β*3Glc-T contains two potential *N*-glycosylation sites, Asn-124 and Asn-336, located in the N- and C-terminal domains of the protein, respectively. The two *N*-glycosylation sites are conserved in mammals and bird. In amphibians and teleosteans only the N-terminal glycosylation site is present. In most invertebrates these aligned *N*-glycosylation sites are absent. However, it should be noted that the invertebrates do contain *N*-glycosylation sites outside of the alignment shown in Figure 24, for example there are four in *Drosophila*, three in *Anopheles* and one in *C.elegans*.

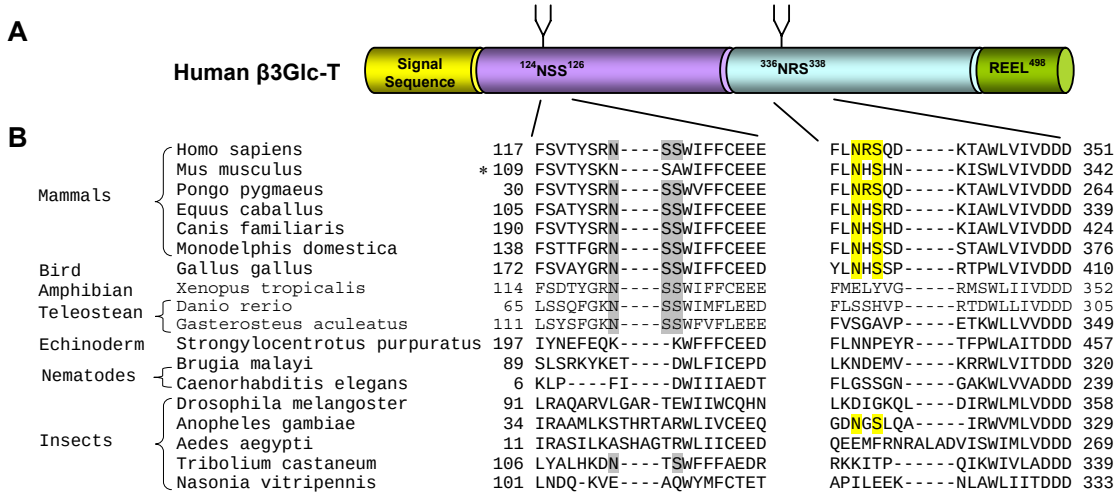


Figure 24. Alignment of the *N*-glycosylation sites in *β*3Glc-T

(A). Schematic representation of human *β*3Glc-T: The signal sequence, N-terminus, C-terminus and the ER-retrieval sequence are represented in yellow, purple, blue and green, respectively. The potential *N*-glycosylation sites (^{124}NSS^{126}, ^{336}NRS^{338}) are also indicated. (B) 18 *β*3Glc-T homologous sequences were selected and aligned using ClustalW. The two *N*-glycosylation sites are shaded in grey and yellow. *Note, that *M. Musculus β*3Glc-T sequence also contains two putative *N*-glycosylation sites, but its N-terminal one, ^{78}NLT^{80}, is not depicted in the figure.

In order to determine if the *N*-glycosylation sites in human *β*3Glc-T are occupied, cell extracts were treated with glycosidases and the *β*3Glc-T monitored by Western blot. The importance of the individual N-glycosylation sites was assessed by site-directed mutagenesis and enzyme activity assays.

Materials and Methods

Constructs

Sequences encoding amino acids 29-498 and 29-494 of human *β*3Glc-T, with N-terminal FLAG tags, were amplified from pcDNA3.1-*β*3Glc-T-Myc-His_6 (see Chapter 6.2) by PCR using the following primer pairs: forward primer (5'-CGGGATCCGATTACAAGGATGACGACGATAAGTCTGAAGATACAAAGAAAGAGGTCAAGC-3'), reverse primer (5'-GCGGATCCCTATAACTCCTCTCGAAAACCTTTCTG-3'); and forward primer (5'-CGGGATCCGATTACAAGGATGACGACGATAAGTCTGAAGATACAAAGAAAGAGGTCAAGC-3'), reverse primer (5'-GCTCTAGACTACTAAAAACCTTTCTGTGTCTCCTGCCTGG-3'), respectively (restriction enzyme sites are underlined). The products were digested with *BamHI* or *BamHI* and *XbaI* and subcloned into the similarly digested pSecTagB to yield pSecTagB-Δ28FLAG*β*3Glc-T and pSecTagB-Δ28FLAG*β*3Glc-TΔREEL. In these constructs, the signal sequence from *β*3Glc-T was replaced with that of Igκ present in the pSecTagB vector. The pcDNA3.1-huPOFUT2-V5-His_6 construct encoding full-length human POFUT2 with C-terminal V5-His_6 tags was provided by Dr. Jeremy Keusch.

Inverse PCR site-directed mutagenesis was performed directly on pcDNA3.1-*β*3Glc-T-Myc-His_6 (see Chapter 6.2) using the following primer pairs for E220A, forward primer (5'-GATTTAAAACATGCTATTGCCCTCTAC-3'), reverse primer (5'-GTAGAGGGCAATAGCATGTTTTAAATC-3'); for D421A, forward primer (5'-GATGCTCCCGATGCTATGGTCCTGGGAATG-3'), reverse primer (5'-CATTCCCAGGACCATAGCATCGGGAGCATC-3'); for N124Q, forward primer (5'-ACATATAGCAGACAGTCATCTTGG-3'), reverse primer (5'-CCAAGATGACTGTCTGCTATATGT-3'); and for N336Q, forward primer (5'-GAAAGATTTCTGCAGCGTAGCCAGGAC-3'), reverse primer (5'-GTCCTGGCTACGCTGCAGAAATCTTTC-3').

Cell culture

HEK293T cells were cultured and transfected as described in Chapter 6.2. Cells and conditioned media were harvested 60 h after transfection. Cell extracts were prepared as described in Chapter 6.2.

Glycosyltransferase activity assays

POFUT and *β*3Glc-T activity assays were performed as described in Canevascini et al. (manuscript in preparation) and Chapter 6.2, respectively.

Co-immunoprecipitation of *β*3Glc-T and POFUT2

Conditioned medium from HEK293T cells, transiently co-transfected with pcDNA3.1-huPOFUT2-V5-His_6 and either pSecTagB-Δ28FLAG*β*3Glc-T or pSecTagB-Δ28FLAG*β*3Glc-TΔREEL, was filtered through a 0.22 µm Express^{TM}Plus filter (Millipore). Complete-EDTA free protease inhibitor (Roche) cocktail was added to the filtered conditioned medium. As a negative control, conditioned medium

from cells transfected with empty pSecTagB cDNA was used. Anti-FLAG M2 affinity gel (Sigma) was equilibrated at 4 °C with TBS (10 mM Tris-HCl, pH 7.4, 150 mM NaCl). The conditioned media or high-speed supernatant was added to the anti-FLAG beads and incubated overnight on a roller at 4 °C. The beads were washed three times with TBS for 30 minutes at 4°C, and the bound recombinant proteins were eluted at room temperature with 0.5 mg/ml 3xFLAG peptide (Sigma).

For the purification of the His_6-tagged protein, Ni-NTA agarose (Qiagen) were equilibrated with buffer A containing 20 mM Tris-HCl, pH 7.4, 500 mM NaCl and 20 mM imidazole. Conditioned media or high speed supernatant containing the His_6-tagged recombinant proteins was bound to the beads and after three wash steps with buffer A, 150 mM EDTA was used for elution.

Deglycosylation reactions

Total protein, from the high-speed supernatant of HEK293T cells transiently transfected with *β*3Glc-T constructs, was deglycosylated with either $EndoH_f$ or PNGaseF (New Englands Biolabs). Digests performed under native conditions used 60 U $EndoH_f$ or 20 U PNGaseF per µg total protein in 10 mM Tris-HCl, pH 7.4, 30 mM NaCl, for 1h at 37˚C. Samples denatured prior to PNGaseF digestion were performed as described in the manufacturer's instructions, with 10 U PNGaseF per µg total protein, for 1h at 37˚C.

Sequence alignments

The amino acid sequences of *β*3Glc-T orthologues were retrieved from the NCBI database using BLAST (http://www.ncbi.nlm.niv.gov/BLAST). GenBank accession numbers of the sequences used in the multiple sequence alignment are: *Homo sapiens* AAO37647.1; *Mus musculus* Q8BHT6.2; Pongo pygmaeus CAH90386.1; *Equus caballus* XP_001495201.1; *Gallus gallus* XP_425633.2; *Canis familiaris* XP_543143.2; *Monodelphis domestica* XP_001377019.1; *Danio rerio* XP_001339799.1; *Gasterosteus aculeatus* ENSGACT00000015186; *Xenopus tropicalis* NP_001072551.1; *Anopheles gambiae* XP_310125.4; *Brugia malayi* EDP35370.1; *Aedes aegypti* XP_001650768.1; *Strongylocentrotus purpuratus* XP_781188.1; *Tribolium castaneum* XP_968234.1; *Caenorhabditis elegans* (zc250.2) NP_504520.1; *Nasonia vitripennis* XP_001606218.1 and *Drosophila melanogaster* NP_608982.2 (CG9109-PA, isoform A). The alignment was performed using ClustalW.

Western blot

Proteins were separated on a 10% SDS-polyacrylamide gel and transferred to nitrocellulose membrane. The primary antibodies used were affinity purified rabbit anti-human *β*3Glc-T (directed against a C-terminal epitope, Agrisera), mouse monoclonal anti-V5 (Invitrogen) and mouse anti-FLAG BioM2 (Sigma). The donkey anti-rabbit IgG-HRP and sheep anti-mouse IgG-HRP secondary antibodies were obtained from GE Healthcare. The extravidin-peroxidase conjugate was diluted 1:2000 (Sigma). All antibodies were diluted 1:5000. The blots were developed using the enhanced chemiluminescence (ECL, Western blotting, GE Healthcare) kit.

Results

Complex formation between β3Glc-T and POFUT2

During the cloning of the FLAG-tagged β3Glc-T constructs, the natural signal sequence (amino acids 1-28) from β3Glc-T was replaced with that of Igκ present in the pSecTagB vector. This exchange of signal sequences did not alter the subcellular localisation of β3Glc-T. When HEK293T cells were transfected with pSecTag-Δ28FLAGβ3Glc-T cDNA (Figure 25A), the overexpressed FLAG-tagged β3Glc-T was still retained in the cellular high-speed supernatant (S2 fraction) and exhibited a similar increase in enzyme activity as the full-length β3Glc-T-Myc-His protein used in Chapter 6.2. In contrast, when HEK293T cells were transfected with pSecTag-Δ28FLAGβ3Glc-TΔREEL cDNA, which lacks the ER-retrieval sequence, the overexpressed FLAG-tagged β3Glc-TΔREEL accumulated in the culture medium (Figure 25C). The secreted β3Glc-TΔREEL protein was immunoprecipitated via its FLAG tag and a greater than 6-fold increase in β3Glc-T activity was detected compared to the controls (Figure 25B). As the secreted form of β3Glc-T protein was readily detectable and active, it could be used to examine if there was a stable physical interaction with POFUT2 in the conditioned medium.

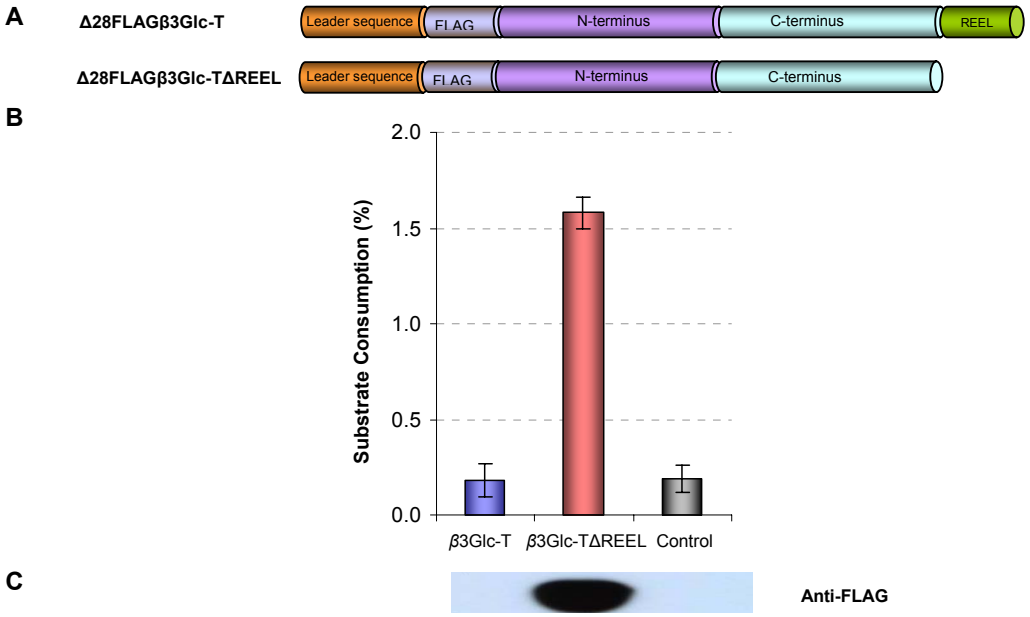


Figure 25. The lack of the ER-retrieval sequence, REEL, results in secretion of β3Glc-T into the culture media

(A) The signal sequence from β3Glc-T was replaced with the Igκ-chain leader sequence. A FLAG-tag was inserted at the N-terminus of β3Glc-T. HEK293T cells were transiently transfected with pSecTag-Δ28FLAGβ3Glc-T (β3Glc-T), pSecTag-Δ28FLAGβ3Glc-TΔREEL (β3Glc-TΔREEL) or empty pSecTag vector (Control). Equal volumes of the conditioned medium were immunoprecipitated using anti-FLAG agarose and assayed for glucosyltransferase activity (B) or for expression of FLAG-tagged β3Glc-T protein by Western blot analysis with anti-FLAG (C). Values represent the average of duplicate samples and the error bars represent standard deviations.

For the complex formation studies, constructs encoding POFUT2-V5-His_6 and FLAG-tagged β3Glc-T (either the secreted or cellular β3Glc-T) were transiently co-transfected into HEK293T cells. The conditioned media was immunoprecipitated with either anti-FLAG agarose or Ni-NTA agarose and the

β3Glc-T and POFUT2 proteins detected by Western blot and their corresponding enzyme activities measured (Figure 26).

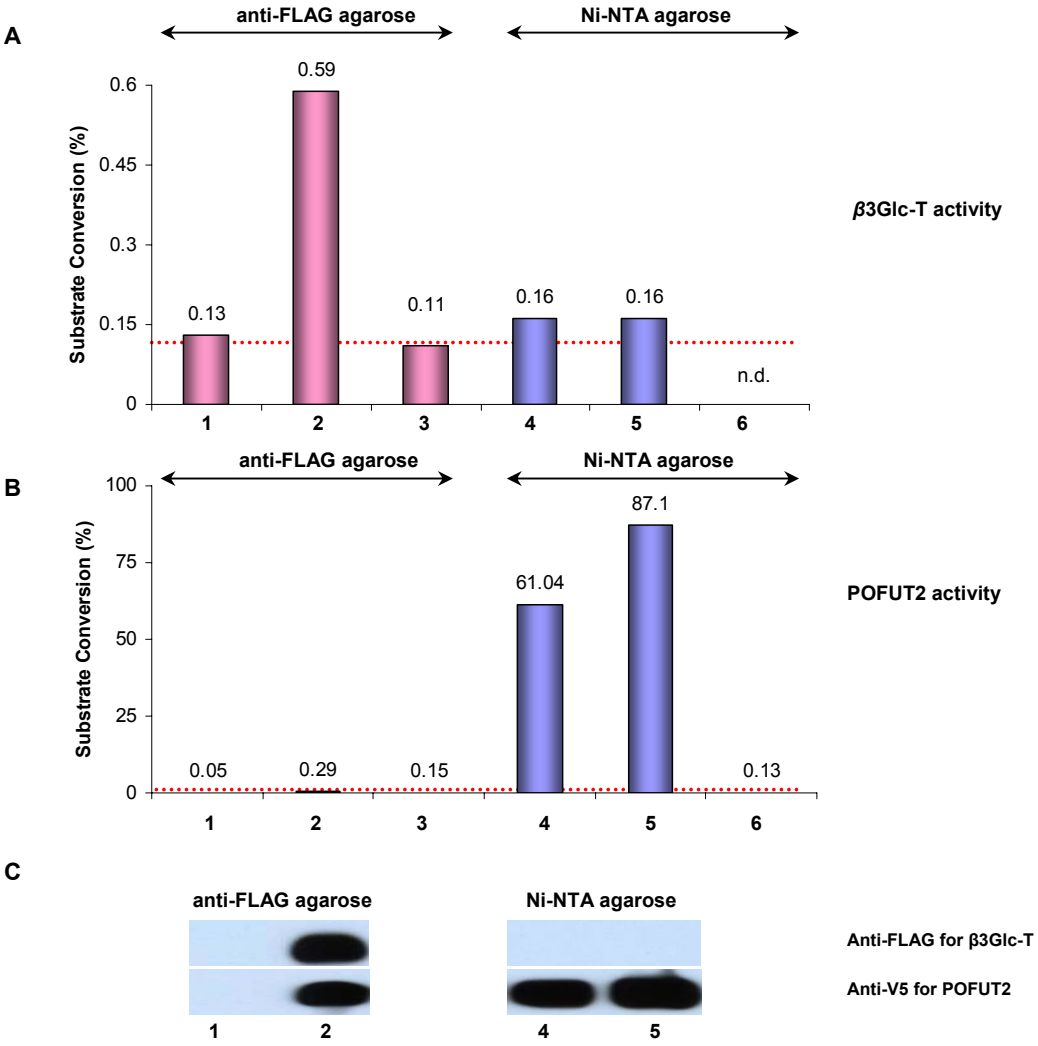


Figure 26. Co-immunoprecipitation of β3Glc-T and POFUT2 from conditioned media

Conditioned media from HEK293T cells co-transfected with pSecTag-Δ28FLAGβ3Glc-T and pcDNA3.1-POFUT2-V5-His_6 (lanes 1, 4), pSecTag-Δ28FLAGβ3Glc-TΔREEL and pcDNA3.1-POFUT2-V5-His_6 (lanes 2, 5), or control empty plasmids (lanes 3, 6) was immunoprecipitated with either anti-FLAG agarose (β3GlcT) or Ni-NTA agarose (POFUT2) as indicated and the corresponding β3Glc-T (A) and POFUT2 (B) activities measured. The substrate consumption values show duplicate measurements in percentile. The background levels from the control samples are shown as a red dotted line. (C) Elutions from the immunoprecipitations were analysed by Western blot. nd, not determined.

Secreted Δ28FLAGβ3Glc-TΔREEL and POFUT2-V5-His_6 proteins could be detected by Western blot using the anti-FLAG and anti-V5 antibodies, respectively (Figure 26C, lanes 2, 4 and 5). As expected, the overexpressed Δ28FLAGβ3Glc-T protein that contains the REEL sequence was retained inside the cells and could not be detected in the medium (Figure 26C, lanes 1 and 4). Interestingly, when the secreted FLAG-tagged Δ28FLAGβ3Glc-TΔREEL was immunoprecipitated with anti-FLAG agarose, the POFUT2-V5-His_6 protein was also detected by Western blot (Figure 26C, lane 2, both panels),

identifying an interaction between the secreted β3Glc-T and POFUT2 proteins. However, the reciprocal immunoprecipitation did not pull down Δ28FLAGβ3Glc-TΔREEL on the Ni-NTA agarose containing POFUT2 (Figure 26C, lane 5, both panels).

To test whether the co-immunoprecipitated material shows a real physical interaction between the two enzymes, bead suspensions were used to measure β3Glc-T and POFUT2 activities (Figure 26A and B). Significant enzyme activities could only be detected in the immunoprecipitations corresponding to the pertinent enzyme i.e. anti-FLAG agarose (β3Glc-T) and Ni-NTA agarose (POFUT2). The anti-FLAG agarose precipitated approximately five-times higher β3Glc-T activity from the Δ28FLAGβ3Glc-TΔREEL co-immunoprecipitations compared to the control sample (Figure 26A). A similar increase in β3Glc-T activity was seen when pSecTagΔ28FLAGβ3Glc-TΔREEL was expressed alone (Figure 25B). The Ni-NTA pull-downs exhibited very high levels of POFUT2 activity, between 470- and 670-fold greater than the control (Figure 26B). Since the POFUT2 activities were measured in the non-linear range the real increase in the experimental versus the control will actually be much greater. β3Glc-T activity was measured in the linear range (less than 10% substrate conversion).

Surprisingly, POFUT2 appears inactive when co-immunopreciptated with β3Glc-T from the conditioned medium. Similar levels of POFUT2-V5-His_6 protein were detected in the co-immunoprecipitations with the different β3Glc-T constructs (Figure 26C lanes 2, 4 and 5). However, there is only significant POFUT2 activity when β3Glc-T protein is absent in these co-immunoprecipitations (compare lanes 4 and 5 in Figure 26B and C with lane 2). Note that β3Glc-T is active when expressed alone or complexed with POFUT2 (Figure 25 and 26A lane 2).

Mutating the putative catalytic base in β3Glc-T affects its activity

Homology modelling between human β3Glc-T and murine Manic Fringe (Figure 21) predicted Asp-421 as the general base in human β3Glc-T. To determine if Asp-421 is catalytically important in β3Glc-T, this conserved residue (located in the ^{419}PDD^{421} motif) was mutated to alanine. The β3Glc-T D421A mutant protein was expressed in the high-speed supernatant of HEK293T cells. Western blot analysis confirmed that the D241 mutant protein migrated at the same apparent molecular weight as the wild type protein, although with a slightly lower level of expression (Figure 27B, lane 4). The D421A mutant showed only ~6% of the wild type β3Glc-T activity (Figure 27A, lane 4) and even exhibited a 60% lower activity than the empty vector transfected sample. The latter finding suggests that the overexpression of the β3Glc-T D421A mutant results in the suppression of the endogenous β3Glc-T activity.

The internal sequence homology in β3Glc-T reveals that ^{132}EEE^{134} and ^{218}KHE^{220} in the N-terminus, aligns with ^{349}DDD^{351} and ^{419}PDD^{421} in the C-terminus, respectively (Figure 23). To investigate whether these glutamic acids in the N-terminus have any affect on β3Glc-T activity, site-directed mutagenesis was performed. Unfortunately, due to technical difficulties, none of the ^{132}EEE^{134} residues were mutated. The E220A mutant was found to be similarly expressed in the high-speed supernatant as the wild type β3Glc-T (Figure 27B, lane 3), yet surprisingly the β3Glc-T activity

was reduced by 8-fold. Again the total β3Glc-T activity in the cell extract overexpressing the E220A mutant was less than that observed for the endogenous β3Glc-T only (empty).

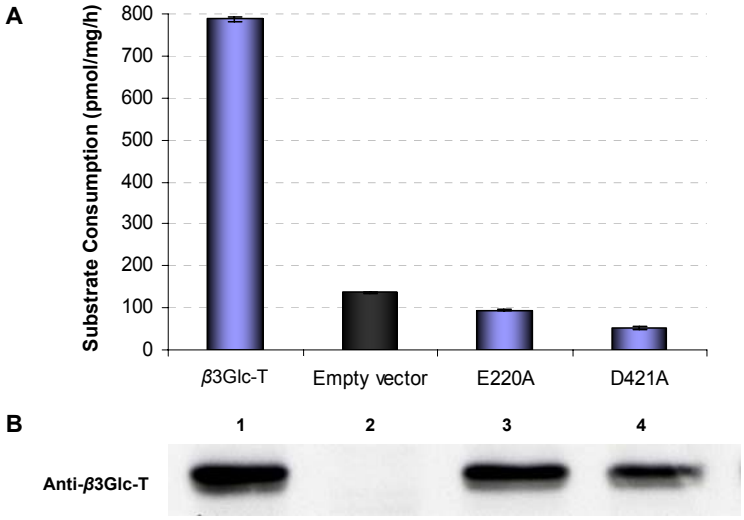


Figure 27. Mutating the putative catalytic bases in β3Glc-T affects its activity

HEK293T cells were transiently transfected with pcDNA3.1-β3Glc-T-Myc-His_6 encoding either the wild type β3Glc-T or harbouring the point mutations E220A or D421A in the ^{218}KHE^{220} and ^{419}PDD^{421} motifs, respectively. (A) Glycosyltransferase activity was measured using 12.5 µg of total protein from the soluble fraction of these transfectants. (B) Western blot analysis was performed with 15 µg of total protein from the soluble fraction and probed with anti-β3Glc-T antibody. Values shown are the average of duplicate samples and the error bars represent standard deviations. Note that the anti-β3GlcT antibody does not detect the endogenous β3Glc-T enzyme in HEK293T cells at this level of total protein.

These results show that a mutation in the C-terminal putative catalytic base ^{419}PDD^{421}, as well as a mutation in the corresponding residue ^{218}KHE^{220}, in the N-terminal domain abolishes the β3Glc-T activity. Thus, despite the lack of structural homology between the N- and C-terminal domains of β3Glc-T, the N-terminal domain also contributes to β3Glc-T activity.

The importance of the *N*-glycosylation sites in β3Glc-T

β3Glc-T has two potential *N*-glycosylation sites at Asn-124 and Asn-336, located in the N- and C-terminal domain, respectively. The calculated molecular mass of the overexpressed, Myc-His_6 tagged β3Glc-T is 60 kDa, whereas the band detected on Western blots migrates at ~65 kDa. In order to determine, how many *N*-glycosylation sites are used in β3Glc-T, the enzyme was treated with glycosidases to remove the *N*-glycans. β3Glc-T is an ER-localised soluble protein, therefore the de-glycosylation experiments were performed using high-speed supernatant of transiently transfected HEK293T cells overexpressing β3Glc-T. The high-speed supernatant was treated with peptide-*N*-glycosidase F (PNGase F), which cleaves all *N*-linked glycoproteins, and analysed by Western blot. Complete deglycosylation was possible under non-denaturing conditions. There was a visible mass shift between the treated and untreated material (Figure 28). To find out the type of the attached *N*-glycans, an endoglycosidase H (Endo H_f) digestion was performed. Endo H_f cleaves only the high-mannose oligosaccharides added in the ER and some hybrid structures, but does not cleave the complex sugar chains synthesised in the Golgi. The β3Glc-T shifted by ~5kDa when digested with either Endo H_f or PNGaseF compared to the control (Figure 28).

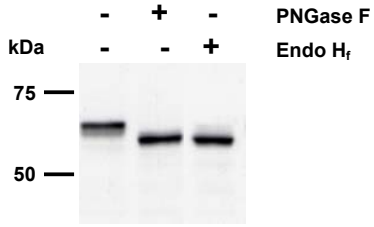


Figure 28. Deglycosylation of β3Glc-T-Myc-His$_6$

Western blot analysis of high-speed supernatant from HEK293T cells transiently transfected with pcDNA3.1-β3Glc-T-Myc-His$_6$ with and without PNGase F and Endo H$_f$ treatment. The glycosidase reactions were performed under native conditions. 12.5 µg of total protein was analysed.

Hence complex *N*-glycans are not present on β3Glc-T. Since β3Glc-T is retained in the ER, the most likely type of *N*-glycans are the high-mannose types ($GlcNAc_2Man_{5-9}$ structures). However, the mass difference observed would correspond to two *N*-glycans of ~2.5kDa, which is more than expected for a typical high mannose ($GlcNAc_2Man_9$) structure (~1.9kDa). A more detailed examination of the *N*-glycans is needed to determine their exact mass and composition.

To determine whether the presence of the N-linked glycans modifies any β3Glc-T properties, *N*-glycosylation site mutants were generated by changing the asparagines at position 124 and 336 to glutamines. Western blot analysis revealed the overexpressed β3Glc-T mutants N124Q and N336Q migrating ~2.5kDa under the wild type β3Glc-T (Figure 29B). The mutant protein bands collapsed to 60kDa after treatment with PNGaseF (data not shown). This indicates that both *N*-glycosylation sites in β3Glc-T are occupied with similar sized glycans. The N124Q was expressed to the same extent as the wild type β3Glc-T, whilst that of N336Q was slightly reduced (Figure 29B).

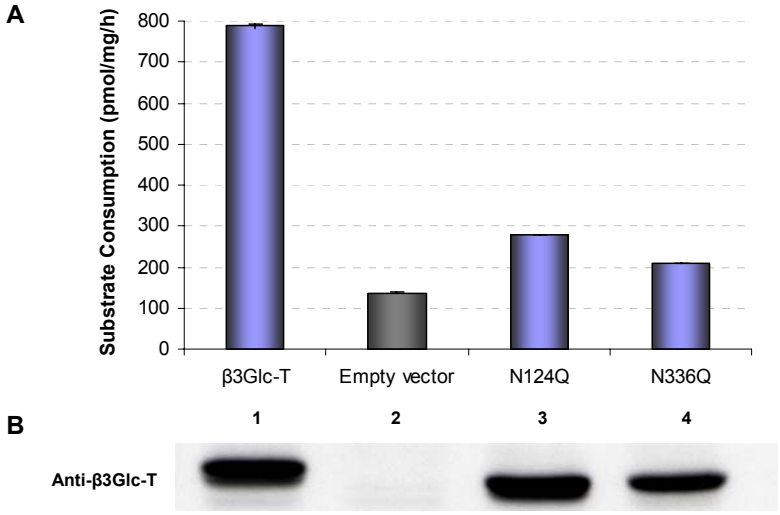


Figure 29. The *N*-glycosylation sites in β3Glc-T are required for its activity

HEK293T cells were transiently transfected with pcDNA3.1-β3Glc-T-Myc-His$_6$ encoding either the wild type β3Glc-T or harbouring the point mutations N124Q and N336Q in the *N*-glycosylation sites. (A) Glucosyltransferase activity was measured using 12.5 µg of total protein from the soluble fraction of these transfectants. (B) Western blot analysis was performed with 15 µg of total protein from the soluble fraction and probed with anti-β3Glc-T antibody. Values shown are the average of duplicate samples and the error bars represent standard deviations. Note that the anti-β3GlcT antibody does not detect the endogenous β3Glc-T enzyme in HEK293T cells at this level of total protein.

To determine whether the *N*-glycosylation sites affects *β*3Glc-T activity, the high-speed supernatant from the transiently transfected cells were tested (Figure 29A). Both the N-terminal and C-terminal *N*-glycosylation site mutants showed greatly diminished *β*3Glc-T activities, with only 35% and 27% of the wild type activity, respectively. As the expression levels of the wild type and N124Q mutant protein was similar (see Figure 29B, lanes 1 and 3), the decrease in activity is due to the *N*-glycosylation site mutation. The *N*-glycan in the C-terminal domain (N336Q) may affect the *β*3Glc-T protein stability to a greater extent than the *N*-glycan in the N-terminal domain. Taken together, the presence of both *N*-glycans is required for full *β*3Glc-T activity.

Discussion

Complex formation between β3Glc-T and POFUT2

A complex between the soluble forms of β3Glc-T and POFUT2 isolated from the conditioned medium was observed. The N-terminal FLAG-tagged β3Glc-T was immunoprecipitated with anti-FLAG agarose and both β3Glc-T and POFUT2 proteins were detected by Western blot. This association inhibited POFUT2 activity but not β3Glc-T activity. However, it was not possible to reproduce this result in the reciprocal pull-down. There may be a trivial technical reason for this that could be resolved by exchanging the tags used or their locations on the expressed protein.

What is more uncertain is the relevance of such an association. Indeed it is not clear if such a complex forms in the ER and transits the secretory pathway into the conditioned media as a stable complex or if the association only occurs outside of the cell. When these experiments on complex formation were initiated, antibodies against the human β3Glc-T and POFUT2 proteins were unavailable. This necessitated the use of tagged proteins in an overexpression system. It would be preferable to investigate complex formation on the endogenous proteins in the ER as this is where the proteins are normally localised. The basal β3Glc-T and POFUT2 activities from HEK293T ER fractions could be detected (Chapter 6.2, Figure 7). Pilot co-immunoprecipitation experiments were performed on the endogenous enzymes from the ER fractions of HEK293T cells using newly generated antibodies against the glycosyltransferases. However, due to the small volumes and low protein concentration in the Nycodenz fractions, neither the anti-β3Glc-T nor the anti-POFUT2 antibodies were able to recognise the endogenous proteins. It may be necessary to screen for a suitable cell line with detectable levels of endogenous β3Glc-T and POFUT2 proteins for future co-immunoprecipitation studies.

Fluorescence energy resonance transfer (FRET) could be used to explore the proximity of endogenous β3Glc-T and POFUT2 proteins in the ER. In this case, fluorophore-tagged antibodies against the two transferases may be used as the donor and acceptor fluorophores and analysed by confocal microscopy or FACS. FRET will only identify close associations but not necessarily direct interactions. The identification of other binding partners involved in complex formation may be addressed using far Western blot (Wu et al. 2007). In this technique, ER fractions containing the endogenous prey protein are separated on SDS or native PAGE, and transferred to a membrane. The proteins are denatured and re-natured on the membrane, and then the membrane is blocked and incubated with purified bait protein (ie POFUT2). Subsequent probing with antibodies to the bait protein will detect the binding partners.

When full-length β3Glc-T, containing the REEL sequence, is overexpressed in HEK293T cells it is still retained in the cellular S2 fraction and does not escape into the culture medium. In contrast, the ER-retention mechanism for POFUT2 is saturable as overexpression of POFUT2 leads to its partial secretion into the culture medium. Even the co-overexpression of the full-length β3Glc-T with POFUT2 does not prevent the secretion of a significant portion of POFUT2 from the cell. Hence, β3Glc-T alone is not capable of retaining POFUT2 in the ER. The SignalP program predicts the signal

sequence for human POFUT2 to be cleaved between amino acids 20 and 21. In our laboratory, the deletion of the first 36 amino acids from POFUT2 results in its bulk secretion from HEK293T cells into the culture medium (Keusch J.J., unpublished observations). This suggests that residues between amino acids 21 and 36 are important for the retention of POFUT2 within the ER.

Identification of catalytic residues in β3Glc-T

As shown in Figure 5A (Chapter 6.2), the mutations in the catalytic core ^{349}DDD^{351} motif, abolish β3Glc-T activity. The modelling of human β3Glc-T onto the catalytic pocket of mouse Manic Fringe crystal structure suggests that not only Asp-349 and Asp-351, but also the highly conserved Asp-421 could be involved in the catalytic mechanism.

Interestingly, the overexpression of β3Glc-T catalytic mutants decreases the endogenous β3Glc-T activity. This had been previously observed with the ^{349}ADD^{351} and ^{349}ADA^{351} mutants (Chapter 6.2) and is reproduced with the D421A mutant and surprisingly with the E220A mutant. This indicates that the mutants could serve a dominant negative function. This inhibition can happen in different ways, for example by homophilic interaction with the active enzyme; association with a presently unknown factor which is responsible for the enzyme activation or by competing for the post-translational modifications required for the enzyme activity. Further experiments are needed to follow up this potential dominant negative effect. To check whether the mutants inhibits the endogenous β3Glc-T *in vivo*, the mixture of mutant and wild type cell extracts should be checked for activity. The mRNA level of the endogenous β3Glc-T also needs to be examined to see whether the decreased activity is due to the down-regulation of the *β3Glc-T* gene.

The results of the structural analysis (Figure 21) and the highly conserved nature of Asp-421 suggest that it might play a direct role in the catalytic mechanism of β3Glc-T as a general base, rather than contributing to the correct folding of an active site. Additional mutational studies, for example replacing the His-456 residue (Figure 21), are needed to gain a broader picture of the catalytic mechanism of β3Glc-T. However, exactly how these residues contribute to β3Glc-T catalytic activity will only be fully understood when the crystal structure of β3Glc-T becomes available.

Substitution of Gln-220 with an alanine, in the N-terminal ^{218}KHE^{220} motif of β3Glc-T abolished β3Glc-T activity. This finding is in contradiction with the idea that the C-terminal domain alone is responsible for the activity of β3Glc-T. To see whether the affinity towards the acceptor or the donor substrate is affected by this mutation, the kinetic parameters of the wild type and the E220A mutant should be examined. It is also conceivable that this point mutation could indirectly affect the enzyme activity through a local conformation change.

The importance of *N*-glycosylation in β3Glc-T

The role of the *N*-linked glycans in glycosyltransferases is protein-dependent. Some glycosyltransferases including bovine POFUT1 (Loriol et al. 2007), α2,6-sialyltransferase (Fast et al. 1993), *Drosophila* Manic Fringe (Jinek et al. 2006) and human core 2 β1,6-*N*-acetylglucosaminyltransferase (Toki et al. 1997), require their *N*-glycosylation sites occupied for full

biological activity. Other glycosyltransferases, like human β1,4-galactosyltransferase (Borsig et al. 1997) and POFUT2 (Chun-I Chen, unpublished observations) retain their catalytic activity in the absence of their *N*-glycans.

Alignments of the *N*-glycosylation sites in β3Glc-T from several species indicated that two sites are conserved in mammals and birds. In human β3Glc-T both *N*-glycosylation sites are occupied. Deletion of either site by site-directed mutagenesis severely diminishes its β3Glc-T activity. The removal of the *N*-linked glycosylation sites can result in the formation of non-native cysteine-bridges leading to a different conformation and consequently reduced enzymatic activity. An alternative approach to study the role of the *N*-glycans in β3Glc-T would be to deglycosylate the high-speed supernatant fraction, under non-denaturing conditions, and then assay for β3Glc-T activity. This would allow the β3Glc-T to achieve its native conformation and thus, the contribution of the *N*-glycans to its activity could be assessed. Preliminary results from such experiments indicate that the *N*-glycans are required for β3Glc-T activity.

The importance of the *N*-glycans in β3Glc-T has several implications. In order to thoroughly study β3Glc-T we would like to purify this enzyme to homogeneity and eventually to crystallise it. The choice of a particular expression system determines the protein's glycosylation. For instance, the expression of a secreted form of β3Glc-T will generate complex type *N*-glycans instead of the high-mannose structures found on the β3Glc-T in the ER. The impact these different glycoforms may have on β3Glc-T activity needs to be studied further. It is possible to produce secreted glycoproteins with a more homogeneous structure by using mutant cell lines deficient in GlcNAcT1 (the committed step for complex *N*-glycan synthesis) or by expressing in the presence of *N*-glycosylation processing inhibitors such as kifunensine (Chang et al. 2007).

6.4. Expression of β1,3-Glucosyltransferase Domains with Reference to Peters Plus Syndrome

Introduction

Recently, the identification of biallelic mutations in the gene encoding β3Glc-T was shown to cause the developmental disorder Peters Plus syndrome (PPS) (Lesnik Oberstein et al. 2006). We investigated whether constructs encoding these mutations expressed the truncated protein products. In addition the expression of the β3Glc-T N- and C-terminal domains were studied. Finally, the first steps to purify β3Glc-T were performed.

The connection between Peters Plus syndrome and β3Glc-T

PPS (MIM 261540) is a rare autosomal recessive disorder characterized by anterior eye-chamber defects, disproportionate short-stature, developmental delay, typical craniofacial features, broad hands and feet, and cleft lip and/or palate. The most frequently occurring eye-chamber defect is Peters anomaly consisting of central corneal opacity, thinning of the posterior aspect of the cornea, iridocorneal adhesions and may include cataracts and glaucoma (Maillette de Buy Wenniger-Prick et al. 2002; Lesnik Oberstein et al. 2006). The eye involvement is usually bilateral.

Lesnik Oberstein and colleagues performed a genome wide array-based comparative genomic hybridization analysis on six patients with PPS and initially identified a deletion of ~1.5 Mb in chromosome 13 (q12.3q13.1) in two brothers (Lesnik Oberstein et al. 2006). Further sequencing of the six genes in this region identified biallelic truncating mutations in a β1,3-galactosyltransferase-like gene (*B3GALTL)* in a total of 20 PPS patients examined. Although no enzyme activity had been demonstrated for *B3GALTL* gene product, it was proposed to be β3-glycosyltransferase-like based on its sequence similarity with other members of the GT31 family in the CAZY database (Heinonen et al. 2003). We and others have since determined that the *B3GALTL* gene encodes the β3Glc-T that synthesizes the unusual disaccharide Glcβ1,3Fuc-*O*- on TSRs (Kozma et al. 2006; Sato et al. 2006). Recent work performed in our laboratory has identified defective glucosylation of TSRs in patients with PPS, which clearly establishes PPS as a new congenital disorder of glycosylation (Hess et al. 2008).

The gene encoding human β3Glc-T is located on chromosome 13 at 13q12.3 and contains 15 exons divided by 14 introns spanning 132 kb (Figure 30). It is expressed in a wide range of tissues with differential expression of 3.4 kb and 4.2 kb transcripts arising from the use of alternative 3'UTRs (Heinonen et al. 2003). The expression is especially high in the brain, heart and kidney. β3Glc-T has orthologues in *C. elegans* through to mammals suggesting that it plays an important role in the tissues of multicellular organisms. In the mouse there are two transcripts of 5 kb and 2 kb expressed in a similar pattern to that observed in man (Heinonen et al. 2006). Curiously, in some mouse tissues the expression of β3Glc-T mRNA is developmentally regulated. For example, *in situ* hybridisation revealed that the expression in the foetal kidney was high in the cortex and weak in the medulla,

whereas in the adult kidney the pattern was reversed. The strong β3Glc-T mRNA signal observed in the retina of a one-day old mouse is downregulated in the adult mouse (Heinonen et al. 2006).

In all the 20 patients with PPS examined there was a hot-spot mutation, present in either one (4/20) or two copies (16/20), in the donor splice site of exon 8 (c.660+1G→A). In two cases the c.660+1G→A mutation was combined with another point mutation in intron 5 (c.347+5G→A). In the other two cases, the c.660+1G→A point mutation was found together with a deletion of ~1.5 Mb in chromosome 13 (q12.3q13.1) (Lesnik Oberstein et al. 2006). Nested RT-PCR performed on RNA isolated from lymphocytes or fibroblasts of patients with PPS, showed that the point mutations c.347+5G→A and c.660+1G→A lead to altered splicing, resulting in the skipping of exons 5 and 8, respectively (Lesnik Oberstein et al. 2006) (Figure 30).

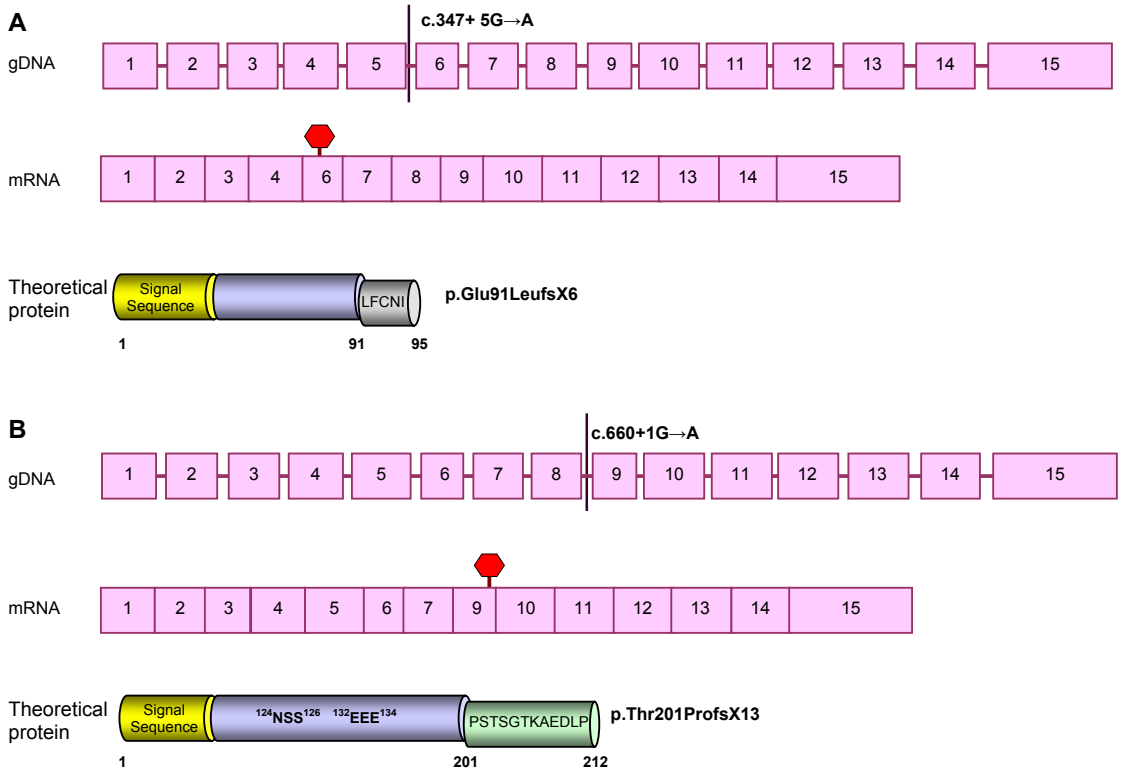


Figure 30. Patients with PPS have truncating mutations in the gene encoding β3Glc-T.

The 15 exons of the gene encoding β3Glc-T are shown as pink boxes (top line in (A) and (B)). The location of the splicing mutations is marked by a vertical black line in the genomic DNA (gDNA). The premature stop codons are represented by red hexagons on the mRNA (middle line in (A) and (B)).

The 347+5G→A and 660+1G→A mutations are predicted to lead to the formation of severely truncated proteins consisting of 95 amino acids (p.Glu91LeufsX6) and 212 amino acids (p.Thr201ProfsX13), respectively. These proteins would lack the catalytic C-terminal domain including the ER-retrieval sequence, which will probably alter their subcellular localisation. Hence, the aberrant splicing of β3Glc-T is likely to result in deficient β3Glc-T activity in patients with PPS. In addition, there may be either a loss- or gain-of-function with the truncated N-terminal domains which could have pathological effects in PPS. We wanted to investigate whether the two predicted truncated proteins are expressed. Cell lines derived from patients with PPS were unavailable at the time, therefore the cDNAs encoding

the putative p.Glu91LeufsX6 and p.Thr201ProfsX13 proteins were generated, cloned into p3xFLAG-CMV-14 expression vector and transiently transfected into HEK293T cells.

Purification of full-length and the N- and C-terminal domains of β3Glc-T

The full-length β3Glc-T has internal sequence homology (Figure 23). Despite the sequence similarity there is no apparent spatial structural similarity. In Chapter 6.3, point mutations of potential catalytic residues in the C- *and* N-terminal domains of β3Glc-T affected the enzyme activity. It is unclear whether the N-terminal domain is also a glycosyltransferase or fulfills another role. The expression of the N- and C-terminal domains of wild-type β3Glc-T was initiated to see if it was possible to express these domains separately. Constructs encoding the N- (amino acids 1-263) and C-terminal (amino acids 264-494) β3Glc-T domains were designed with different secretion signals and purification tags. The further aim was to try and identify ligands for the domains using isothermal titration calorimetry.

The large-scale expression and purification of FLAG-tagged full-length β3Glc-T was started in order to better study the enzyme kinetics of β3Glc-T, and as an initial step towards its crystallization.

Materials and Methods

Constructs

Sequences encoding amino acids 29-263 or 1-263 (N-terminal constructs) of human *β*3Glc-T were amplified from pcDNA3.1-*β*3Glc-T-Myc-His_6 (Chapter 6.2) by PCR using the following primer pairs: forward primer (5'-GGGGTACCACCACCATGCGGCCGCCCGCCTGCTG-3') and reverse primer *(*5'-GCGGATCCTGGCTTTCTACAAAGCGGTAGAAAAG-3'); forward primer (5'-TGCTCTAGAACCAC CATGCGGCCGCCCGCCTGCTG-3') and reverse primer (5'-GCGGATCCTGGCTTTCTACAAAG CGGTAGAAAAG-3'); forward primer (5'-GGGGTACCACCACCATGCGGCCGCCCGCCTGCTG-3') and reverse primer (5'-GCGGATCCTGGCTTTCTACAAAGCGGTAGAAAAG-3'), respectively (restriction sites are underlined). The products were digested with *Kpn*I and *BamH*I, *Xba*I and *BamH*I, or *Kpn*I and *BamH*I, and ligated into similarly digested pSecTagB, pcDNA3.1 (-) and p3xFLAG-CMV-14 vectors, respectively. This yielded the differently tagged pSecTagB-Δ28FLAG*β*3Glc-T26, pcDNA3.1-*β*3Glc-T263-Myc-His_6 and pCMV14-*β*3Glc-T263-3xFLAG constructs, respectively. Sequences encoding amino acids 264-494 (C-terminal constructs) of human β3Glc-T were amplified from pcDNA3.1-β3Glc-T-Myc-His_6 using the following primers: forward primer (5'-CGGGATCCGAT TACAAGGATGACGACGATAAGGTGAAGAAGAAGGATATTTTTGTTGC-3') and reverse primer (5'-CGAGATCTAAAACCTTTCTGTGTCTCCTGCCTGG-3'); forward primer (5'-CGGGATCCGTGA AGAAGAAGGATATTTTTGTTGC-3') and reverse primer (5'-CGAGATCTAAAACCTTTCTGTGT CTCCTGCCTGG-3') (restriction sites are underlined). The products were digested with *BamH*I and *Xba*I, then ligated into similarly digested pSecTagB to yield the C-terminal constructs, pSecTagB-Δ263FLAG*β*3Glc-TΔREEL and pSecTagB-Δ263*β*3Glc-TΔREEL-Myc-His_6.

To create DNA sequences encoding the two theoretically truncated *β*3Glc-T proteins from patients with PPS (p.Glu91LeufsX6 and pThr201ProfsX13), pcDNA3.1-*β*3Glc-T-Myc-His_6 template was amplified by PCR using the following primer pairs: forward primer (5'-GGGGTACCA CCACCATGCGGCCGCCCGCCTGCTG-3') and reverse primer (5'-GCGGATCCTATGTTACAGAAA AGCTGTGTAAG-3'); forward primer (5'-GGGGTACCACCACCATGCGGCCGCCC GCCTGCTG 3') and reverse primer (5'-GCGGATCCTCAGGGGAGGTCCTCCGCCTTTGTCCCAGATGTAG AGGGCAATTTGTTTACAAGTGGAATACTTA-3'), respectively. The PCR products were digested with *Kpn*I and *BamH*I and ligated in-frame into similarly digested p3xFLAG-CMV-14 to yield pCMV14-ΔEx5*β*3Glc-T-3xFLAG and pCMV14-ΔEx8*β*3Glc-T-3xFLAG.

Cell culture

Transfections and preparations of cell extracts were performed as described in Chapters 6.2 and 6.3. The work was done in collaboration with Dr. Jeremy Keusch. To generate stable clones, p343X cDNA (that confers resistance to hygromycin B), was co-transfected with either the pSecTagB-Δ28FLAG*β*3Glc-T263 construct encoding the N-terminus, or pSecTagB-Δ28FLAG*β*3Glc-TΔREEL construct encoding the full-length secreted *β*3Glc-T. Selection of stable clones was started 48 hours after transfection in medium containing 300 µg/ml hygromycin B. Single resistant clones were picked,

expanded and the conditioned medium screened for the highest expression of the recombinant FLAG-fusion proteins (see Chapter 6.3).

Large-scale purification of full-length and N-terminal domain of β3Glc-T

Conditioned media from HEK293T cells stably expressing Δ28FLAGβ3Glc-TΔREEL clone C3 (1 litre) and Δ28FLAGβ3Glc-T263 (N-terminal) clone A42-8 (0.9 litre) constructs were collected. Anti-FLAG beads were equilibrated in ice-cold TBS-PI (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, with Complete Protease Inhibitor® (Roche)). The culture media was divided into 50 ml aliquots and 270 µl of equilibrated anti-FLAG agarose (50% suspension) was added per 50 ml aliquot. The bead-media suspensions were rotated overnight at 4 °C. The anti-FLAG agarose were recovered by centrifugation at 4000 rpm for 0.5 min, and washed three times with 50 ml of TBS-PI by rotating for 1 h at 4 °C. The washed beads were stored as a 50% suspension in TBS-PI at 4 °C.

Ion-exchange chromatography of affinity-purified full-length β3Glc-T

The 50% suspension of anti-FLAG agarose with full-length β3Glc-T bound was spun down at 500 rpm for 10 min 4 °C and washed twice with one bead volume of 10 mM Tris-HCl, pH 7.5, 3 M NaCl solution for 15 min at 4 °C with rotation. The salt solution was removed by applying three washes of TBS-PI buffer for 15 min at 4 C with rotation. The anti-FLAG beads were eluted with one bead volume of 0.5 mg/ml FLAG peptide for 1 h at 4 °C with rotation. Half of the eluate (~2 ml) was dialysed against 2 litres 20 mM Tris-HCl, pH 8.5, at 4 C, using a 10 kDa MWCO Slide-A-Lyzer.

Anion-exchange chromatography was done using a 1ml Mono Q 5/50 fast protein liquid chromatography column (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 8.5. The sample was loaded onto the column at a flow rate of 1 ml/min and proteins were eluted using a 30 min linear gradient of 0-500 mM NaCl in 20 mM Tris-HCl, pH 8.5 at a flow rate of 1 ml/min. The fractions were collected and 10 µl aliquots analysed by SDS-PAGE and Western blot as described in Chapter 6.3.

Co-immunoprecipitation of the N- and C-terminal domains of β3Glc-T

Co-immunoprecipitation of conditioned media from HEK293T cells co-transfected with the N-terminal pSec-tagB-Δ28FLAGβ3Glc-T263 and the C-terminal pcDNA3.1-Δ263β3Glc-TΔREEL-Myc-His_6 was performed with either anti-FLAG agarose or Ni-NTA-agarose (see Chapter 6.3). A fraction of the eluates was analysed by Western blot and assayed for β3Glc-T activity (see Chapters 6.2 and 6.3).

Results

Expression of truncated β3Glc-T proteins

Lesnik Oberstein and colleagues identified mutations in the gene encoding β3Glc-T in patients with PPS (Lesnik Oberstein et al. 2006). These mutations generate out-of-frame mRNA transcripts that predict severely truncated proteins lacking the β3Glc-T catalytic domain. Further, patients with PPS exhibit defective glucosylation of their TSRs in properdin (Hess et al. 2008). These reports suggest that β3Glc-T activity is absent in the patients with PPS. Neither the stability of the mutated transcripts nor the existence of the truncated β3Glc-T translation products is known. Since the only available antibody against β3Glc-T would not detect these shortened proteins, constructs encoding the predicted truncated β3Glc-T proteins with a 3xFLAG tag were generated and transiently transfected into HEK293T cells.

Interestingly, proteins were expressed from both constructs encoding the truncated β3Glc-T proteins, p.Glu91LeufsX6 and p.Thr201ProfsX13, with similar levels detected inside the cell as well as in the conditioned media (Figure 31). The bands observed on the Western blot were ~2kDa greater than their expected size. In the case of the p.Thr201ProfsX13 protein this increase in molecular weight may be due to the addition of an *N*-linked glycan at Asn-124. As there is no apparent ER-retention signal present in these truncated proteins it is surprising that approximately half of the expressed protein is retained inside the cells.

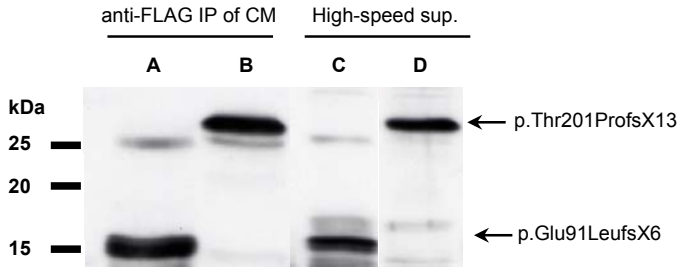


Figure 31. Expression of truncated β3Glc-T proteins

DNAs encoding the truncated β3Glc-T proteins (pGlu91LeufsX6 and p.Thr201ProfsX13), predicted from mutations identified in patients with PPS, were transiently transfected in HEK293T cells (lanes A, C and B,D, respectively). Conditioned media (CM) was immunoprecipitated with anti-FLAG agarose and the eluates (~3.8% of total; lanes A, B) or 25 µg of total protein from the high-speed supernatant (~1.4% of total; lanes C, D) was analysed by Western blot using anti-FLAG.

Expression studies of the N- and C-terminal domains of β3Glc-T

As discussed in Chapter 6.3, β3Glc-T consists of two approximately equal N- and C-terminal domains. Differentially-tagged constructs encoding secreted forms of these β3Glc-T domains were generated. The N-terminal domain (amino acids 1-263) constructs incorporated either the natural signal sequence from β3Glc-T or the one from IgK. For the C-terminal domain (amino acids 264-294) constructs the signal sequence from IgK was used and the ER-retrieval like sequence REEL was deleted from the sequence to facilitate protein expression into the culture medium.

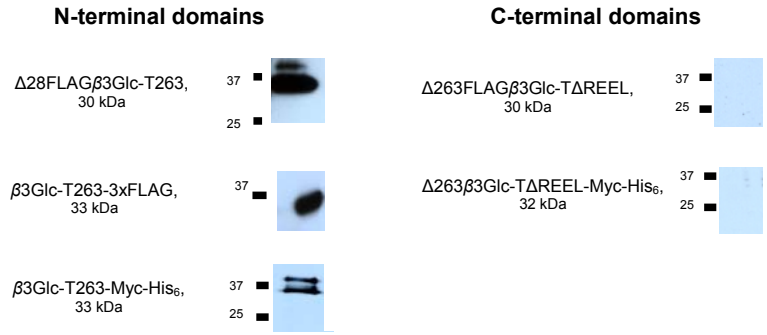


Figure 32. Expression of β3Glc-T N- and C-terminal domains in HEK293T cells

The differentially tagged β3Glc-T N- and C-terminal domain constructs were designed to allow protein expression into the culture media. Conditioned medium from the transiently transfected HEK293T cells was immunoprecipitated using either anti-FLAG agarose or Ni-NTA agarose. The eluates were analysed by Western blot using either anti-FLAG or anti-Myc antibodies for the detection of the N-terminal domains, or with the anti-β3Glc-T antibody for the detection of the C-terminal domain. The expected molecular weights of the proteins are indicated.

The constructs were transiently transfected into HEK293T cells and the conditioned media was immunoprecipitated using either anti-FLAG agarose or Ni-NTA agarose, and the eluates analysed by Western blot (Figure 32). Proteins expressed from all the N-terminal domain constructs could be detected by Western blot and were stable. In all cases the observed molecular weights were higher than expected, which was due to an *N*-glycan on the N-terminal *N*-glycosylation site. As expected, there was no β3Glc-T activity observed for the N-terminal domain of β3Glc-T (Figure 33A, lane 2).

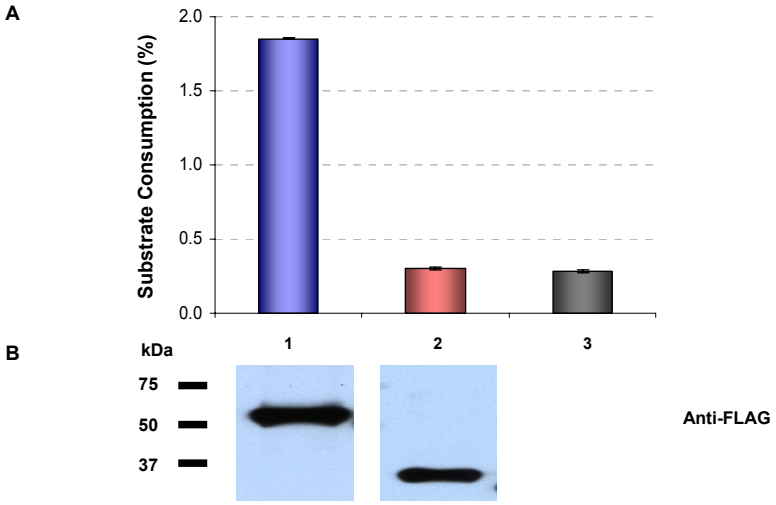


Figure 33. Immunoprecipitation of the full-length and N-terminal β3Glc-T from culture media

Conditioned media from HEK293T cells transfected with pSecTag-Δ28FLAGβ3Glc-TΔREEL (lane 1), pSecTag-Δ28FLAGβ3Glc-TΔ263 (lane 2) or control empty plasmids (lane 3) was immunoprecipitated with anti-FLAG beads and the β3Glc-T activity was measured (A). The substrate consumption values show duplicate measurements in percentile. (B) Elutions from the immunoprecipitations were analysed by Western blot.

Interestingly, the N-terminal domain construct encoding β3Glc-T263-3xFLAG was not only expressed in the conditioned media but also in the high-speed supernatant (data not shown). This high-speed supernatant material also lacked β3Glc-T activity (data not shown). A similar pattern of expression

was observed with the truncated β3Glc-T proteins (Figure 31). In contrast, the N-terminal domain construct encoding Δ28FLAGβ3Glc-T263 that has the IgK signal sequence in place of the β3Glc-T signal sequence, was only detected in the conditioned medium. It appears that the natural signal sequence of β3Glc-T not only directs the proteins into the secretory pathway but also influences the retention of the shortened forms of β3Glc-T that lack an ER-retrieval sequence.

HEK293T cells did not express any of the C-terminal domain proteins, either inside the cell (data not shown) or secreted into the culture medium (Figure 32). However, when the N- and C-terminal domain constructs were co-transfected into HEK293T cells both the N- *and* the C-terminal domain proteins could be immunoprecipitated from the conditioned media. The anti-FLAG immunoprecipitated the β3Glc-T263-3xFLAG N-terminal domain (Figure 34B, top panel, lane 1) and the Ni-NTA pulled-down the Δ263β3Glc-TΔREEL-Myc-His_6 C-terminal domain (Figure 34B, bottom panel, lane 3). Moreover, the N-terminal domain was detected in the C-terminal domain pull-down (Figure 34B, top panel, lane 3) and the C-terminal domain was detected in the N-terminal domain immunoprecipitation (Figure 34B, bottom panel, lane 1). These results indicate that the N- and C-terminal domains are non-covalently interacting forming complex. Furthermore, the expression of the C-terminal domain requires the co-expression of the N-terminal domain.

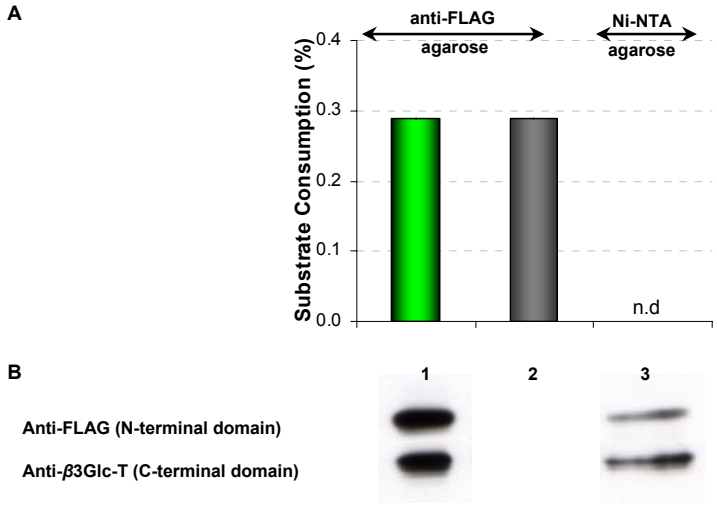


Figure 34. Co-immunoprecipitation of N- and C-terminal domains

HEK293T cells were transiently co-transfected with constructs encoding β3Glc-T263-3xFLAG and Δ263β3Glc-TΔREEL-Myc-His_6 (lanes 1 and 3) or with empty vector alone (lane 2). The conditioned media was immunoprecipitated using anti-FLAG agarose (lanes 1 and2) or Ni-NTA agarose (lane 3) and assayed for β3Glc-T activity (A) or analysed by Western blot (B). n.d., not determined.

The recovery of the C-terminal domain expression by the co-expression of the N- and C-terminal domains did not result in any β3Glc-T activity above background (Figure 34A, lanes 1 and 2). This unambiguously shows that for β3Glc-T activity, the N- and C-terminal domains need to be expressed within the same molecule, i.e. in the intact β3Glc-T protein.

Initial purification of the full-length β3Glc-T

Pilot experiments (Figure 25A) showed that active recombinant β3Glc-T, Δ28FLAGβ3Glc-TΔREEL, was successfully secreted into the culture media. Conditioned media from HEK293T cells stably expressing Δ28FLAGβ3Glc-TΔREEL was used in an initial enzyme purification trial. The FLAG-tagged β3Glc-T was purified from 1 litre of conditioned media using anti-FLAG agarose. The protein was eluted from the affinity column using 3xFLAG peptide and analysed by SDS-PAGE. After staining the gel with Coomassie blue, β3Glc-T could be detected although there were also many contaminating proteins present (Figure 35A, lane 1). A 3 M NaCl salt-wash of the anti-FLAG agarose was performed before elution with the 3xFLAG peptide in order to reduce the non-specifically bound contaminants (Figure 35A, lane 2). The β3Glc-T activity in the eluates before and after salt-wash was found to convert 0.93% and 0.86% of the substrate input, respectively. Hence the salt-wash did not result in a significant decrease in β3Glc-T activity but did reduce the complexity of the elution fraction (Figure 35A, lane 2). From the BSA protein standard curve, the amount of β3Glc-T protein was estimated as ~53 ng/µl and ~110 ng/µl for the eluates with and without salt-wash, respectively (Figure 35). Almost a 2-fold increase in specific enzyme activity for the salt-washed eluate (5.4 nmol mg^{-1} h^{-1}) compared to the non-salt-washed eluate (2.8 nmol mg^{-1} h^{-1}) was observed. The total amount of isolated β3Glc-T protein from the 1 litre of conditioned medium is estimated at 120 µg.

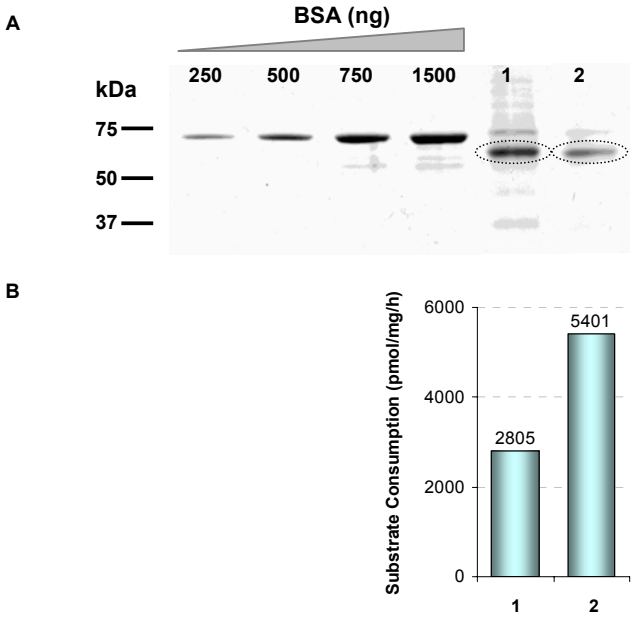


Figure 35. Quantification of β3Glc-T eluted from anti-FLAG agarose

β3Glc-T protein was purified, from 1 litre of conditioned media from HEK293T cells stably expressing the Δ28FLAGβ3Glc-TΔREEL, using anti-FLAG agarose. (A) 10 µl of the eluates before (lane 1) and after a high-salt wash (lane 2), were analysed by SDS-PAGE. A BSA protein standard curve was used for quantification. The gel was stained with Coomassie blue. (B) 7.5 µl of the eluates before (lane 1) and after a high-salt wash (lane 2), were assayed for β3Glc-T activity.

The eluate from the salt-washed beads was subjected to dialysis to remove the 3xFLAG-peptide. Using the protein analysis software GPMAW, an overall net charge of β3Glc-T was calculated as -3 with the pH between pH 7-9. Thus, anion-exchange chromatography, at pH 8.5, was

used to purify β3Glc-T from the dialysed salt-washed eluate. The dialysed material bound to the column as no protein was detected in the flow through. Proteins were eluted from the column by increasing the NaCl concentration (Figure 36A). Fractions where significant A_{280} was detected were analysed by SDS-PAGE and Western blot. Initially, the peak at 27.1 min was expected to be the recombinant β3Glc-T protein, but only traces of protein were detected by Western blot or silver-stained SDS-PAGE (Figure 36B).

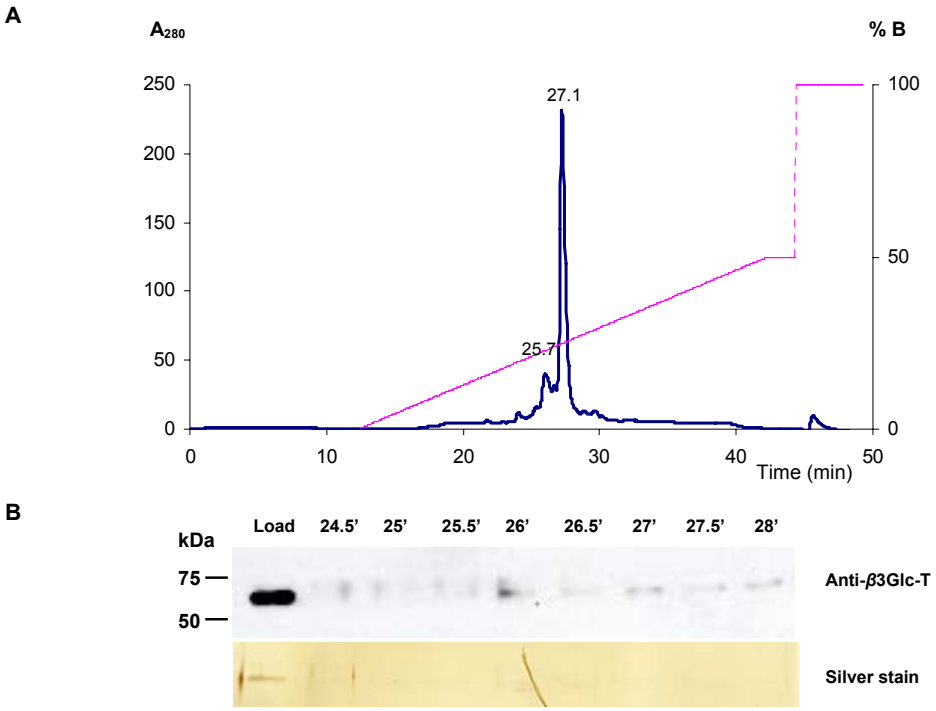


Figure 36. Purification of affinity-purified β3Glc-T by anion-exchange chromatography

(A) Anion-exchange chromatography was performed using a 1ml MonoQ 5/50 FPLC column. The dialysed sample was loaded onto the column in 20 mM Tris-HCl, pH 8.5. Proteins (A_{280} trace, blue line) were eluted from the column using a linear gradient of 0-500 mM NaCl in 20 mM Tris-HCl, pH 8.5 (pink line) at a flow rate of 1ml/min. (B) 10 µl (2%) from the fractions that eluted between 24.5 and 28.5 minutes were separated on a 10% SDS-PAGE and either analysed by Western blot or silver-stained. The load represents 0.25% of the dialysed sample.

Subsequent studies in our laboratory have revealed that the 3xFLAG peptide is not efficiently dialysed out and that the absorbance peak at 27.1 min on the Mono Q corresponds to the 3xFLAG peptide. It appears that much of the β3Glc-T protein was lost during the dialysis step. The load band (Figure 36B) should have represented >250 ng protein but this dialysed material was poorly visible on the silver-stained gel and produced only a moderate signal on the Western blot.

Discussion

Truncated β3Glc-T proteins in Peters Plus syndrome

PPS is caused by truncating mutations in the gene encoding β3Glc-T (Lesnik Oberstein et al. 2006). The deficient glucosylation of TSRs observed in patients with PPS (Hess et al. 2008) is thought to be due to the loss of β3Glc-T activity. The two splicing mutations described in PPS are predicted to yield short β3Glc-T proteins that contain partial N-terminal domains. Constructs encoding these truncated β3Glc-T proteins (p.Glu91LeufsX6 and pThr201ProfsX13) were efficiently expressed in HEK293T cells. The expressed truncated proteins were stable and detected in the conditioned media as well as in the high-speed supernatant. These results lead one to speculate on the existence and possible roles of such truncated β3Glc-T proteins in patients with PPS.

The truncated proteins were not assayed for β3Glc-T activity but are likely to be inactive as observed for the β3Glc-T N-terminal domain (amino acids 1-263). In the previous chapter an association between β3Glc-T and POFUT2 was shown that inhibited POFUT2 activity. In PPS, it is conceivable that complex formation would not be possible between the truncated forms of β3Glc-T and POFUT2 resulting in elevated POFUT2 activity.

Until the function of the N-terminal domain of β3Glc-T is determined it is difficult to speculate on the roles of the truncated forms of β3Glc-T. It is also unclear if the truncated β3Glc-T proteins functionally represent shortened N-terminal domain proteins. Considering that the truncated β3Glc-T proteins lack disulphide bridges and contain unrelated β3Glc-T sequences at their C-terminus arising from the frameshifts, these truncated proteins may actually reveal some novel functions. Since the truncated proteins are stable they could be used in immunoprecipitation experiments as a way to identify potential binding partners.

The β3Glc-T signal sequence appeared to enhance the retention of various full-length and partial β3Glc-T N-terminal domain proteins inside the cell (high-speed supernatant fraction). It is unclear if this is due to self-association with the endogenous full-length β3Glc-T or if there is an additional role attributable to the natural β3Glc-T signal sequence. There is growing evidence that signal peptides contain a significant amount of functional information apart from directing the protein into the secretory pathway. For example, the signal peptide can function as a trans-acting factor to promote Lassa virus GP-C proteolytic processing (Eichler et al. 2003). Signal sequences have also been shown to promote protein oligomerisation (Benach et al. 2003) and prevent protein misfolding (Szabady et al. 2005).

Expression of β3Glc-T N- and C-terminal domains

Internal sequence homology exists between the β3Glc-T N- and C-terminal domains (Figure 23), yet only the C-terminal domain aligns with the 3D-structure of Manic Fringe. Recently, the Global Trace Graph (GTG-DEEP; http://ekhidna.biocenter.helsinki.fi/gtg/start) algorithm became available that performs searches of very distant homologues to the query protein sequence and identifies protein folds (Heger et al. 2007). Using this algorithm, the β3Glc-T N-terminal domain aligned to the PDB

structures of several galectins, a family of compact globular galactose-binding proteins. This alignment of amino acids 96-240 from β3Glc-T with the galectins gave a relatively high score (925) despite the large 50 amino acid insertion in the β3Glc-T sequence. The purification of the β3Glc-T N-terminal domain and subsequent analysis by isothermal calorimetry should clarify whether this domain has intrinsic sugar-binding activity.

The catalytic C-terminal domain of β3Glc-T could not be expressed independently of its N-terminal domain. Co-transfection of β3Glc-T N- and C-terminal domain constructs in HEK293T cells did rescue the expression of the C-terminal domain but not its β3Glc-T activity (Figure 34). This suggests that in the full-length β3Glc-T molecule, the folding of the N- and C-terminal domains is inter-dependent to maintain β3Glc-T activity, possibly through interdomain disulphide-bond formation.

It was not possible to express the *C. elegans* homologue of β3Glc-T, *zc250.2,* in mammalian HEK293T cells (data not shown). The *C. elegans* β3Glc-T protein lacks the first 106 amino acids from its N-terminal region compared to the human β3Glc-T. It could be that the absence of this portion of the N-terminal domain prevents the expression of the *C. elegans* β3Glc-T in mammalian cells. Another possibility is that either a specific co-factor or a chaperone molecule in *C. elegans* needs to be present to facilitate the protein expression in an orthologous expression system. This inter-species discrepancy has been observed with core 1 β1,3-galactosyltransferase (T-synthase). The mammalian T-synthase can only be expressed in mammalian expression system, where the specific molecular chaperone, Cosmc, is present (Ju et al. 2002).

The β3Glc-T N-terminal domain can be stably expressed and is able to support the expression of the C-terminal domain when co-transfected, suggesting that the N-terminal domain is acting as an internal chaperone. A similar phenomenon was observed with the Golgi β1,4-galactosyltransferase, where the N-terminal stem region showed a positive effect on the *in vitro* folding of the C-terminal catalytic domain (Boeggeman et al. 2003).

Purification of full-length β3Glc-T

When β3Glc-T was expressed in the soluble fraction (high-speed supernatant), the activity was 5-times higher than that measured for the endogenous enzyme. An increase in β3Glc-T specific activity was expected when β3Glc-T was secreted and immunoprecipitated from the conditioned media. However, despite the relatively high amount of recombinant protein (Δ28FLAGβ3Glc-TΔREEL; ~0.12 mg/l) immunoprecipitated from the conditioned media, the specific β3Glc-T activity was surprisingly low, with still only a 5-fold increase over the endogenous activity. Assuming that the substrate concentrations in the activity assay are saturating, the calculated turnover number (k_{cat}) of the enzyme would be 10^{-4} sec^{-1}, which is 1000-fold less active than that calculated for POFUT2 (Dominique Klein, unpublished observations). There can be several reasons for the low specific activity of the secreted Δ28FLAGβ3Glc-TΔREEL protein.

It has been shown for several membrane-bound Golgi glycosyltransferases that dimerisation plays a role in regulating their activities. Dimers of α1,3-fucosyltransferase VI (Borsig et al. 1998) and the Galβ1,3-glucuronosyltransferase (Ouzzine et al. 2000) are catalytically active. In contrast, dimerisation of the sialyltransferase ST6Gal1 results in a conformational change, which leads to

decreased affinity for its nucleotide sugar donor and consequently to diminished activity (Ma et al. 1996). My preliminary experiments indicate that the secreted β3Glc-T dimerises and that the N-terminal domain alone can support dimerisation (data not shown). It would be worthwhile to investigate the ratio between the dimeric and monomeric β3Glc-T, determine how and where the dimerisation occurs and establish if the catalytic activity is preferentially associated with a particular form.

Another explanation for the low specific activity of the secreted β3Glc-T may be that it cannot fully function when it is removed from the ER, its natural working environment. It is possible that β3Glc-T requires additional component(s) for its full activity. Thus, overexpression of β3Glc-T in any cellular systems would not be optimal if the co-factor(s) would be rate-limiting for the production of active β3Glc-T. To determine the existence of the additional factor(s), the enzyme purification should be performed from a cell line expressing endogenous β3Glc-T at a level which can be detected by the anti-β3Glc-T antibody. Affinity columns containing TSR-fucose or UDP could be employed for the purification. The semi-purified material could be analysed by MS to identify β3Glc-T binding partners, which might be the co-factor(s) necessary for its full activity.

7. SUMMARY AND OUTLOOK

Any one of the three main players in glycobiology (Figure1) is an entry point for research into this fascinating field. In 2001, a unique Glc*β*1,3Fuc disaccharide was discovered on the TSRs of TSP-1 (Hofsteenge et al. 2001). Five years later, the glycosyltransferases involved in this *O*-fucosylation pathway were revealed, starting with POFUT2 (Luo et al. 2006b) and now part of the work in this thesis has identified the *β*3Glc-T that completes the glycan structure (Kozma et al. 2006; Sato et al. 2006). Severe truncating mutations in *β*3Glc-T, disrupt the synthesis of Glc*β*1,3Fuc and cause the developmental disorder Peters Plus syndrome (PPS) (Lesnik Oberstein et al. 2006).

Searching for the missing pieces in the glycobiology puzzle…

It is likely that there is a third player, a lectin, involved in TSR glycobiology. Some ER glycosyltransferases that recognise correctly folded substrates may play a role in the protein folding quality control mechanism (Okajima et al. 2005). If *β*3Glc-T is assisting in the supra-domain assembly of multiple TSRs, then perhaps the addition of the Glc*β*1,3Fuc disaccharide could act as a specific export signal for an ER lectin. Alternatively, lectins in the blood or on the surface of cells may recognise this sugar modification on secreted glycoproteins. We now have the tools available to generate different TSR glycoforms that could be used in the search for a Glc*β*1,3Fuc disaccharide-specific lectin. The *in vitro* production currently uses the bacterially expressed TSR and builds up the *O*-fucosylated glycan in a step-wise manner with the purified recombinant glycosyltransferases.

It is also feasible to do similar experiments in cell culture systems where the additional modifications of TSRs, notably *C*-mannosylation would also be produced. Characterised TSR glycoforms will help to elucidate the biological significance of the *O*-glycosylation modification. It would be interesting to determine if particular TSP-1 glycoforms affect its activation of TGFβ or alters its binding to CD36 and ability to activate apoptosis in endothelial cells.

Abnormal *O*-glycosylation was detected in properdin TSRs from patients with PPS (Hess et al. 2008), but not in unrelated *O*- and *N*-linked glycosylation pathways (Lesnik Oberstein et al. 2006). Only a modest decrease (1.3-1.7-fold) in the serum properdin levels from patients with PPS was observed which is unlikely to be of any significance for the disease (Hess et al. 2008). Whilst properdin remains a useful reporter for the glycosylation status in PPS it is expected that many of the other 140 human TSR-containing proteins will also lack glucose and that this deficiency may affect the biological activity of some key glycoproteins in PPS. Inhibiting the fucosylation (and thus the glucosylation) in the TSRs of ADAMTS13 (Ricketts et al. 2007) and ADAMTS-like 1 (Wang et al. 2007), decreased their secretion in cell culture. It is unknown if the lack of the glucose alone would have any consequence on the secretion of TSR-containing proteins.

Clearly there is a need to be able to monitor the Glc*β*1,3Fuc disaccharide on more than a few proteins. Although we know this modification occurs in TSRs it may exist elsewhere. Sometimes so-called “rare” forms of protein glycosylation are only thought to be rare because the tools required for their detection are unavailable, or they make up a minor component in an abundant source. An antibody against the Glc*β*1,3Fuc disaccharide would be a useful reagent to develop. The antigen

could be prepared *in vitro* using the purified recombinant glycosyltransferases and either coupled to a carrier for immunisation or used to screen phage display libraries.

Another open approach to identify novel *O*-fucosylated proteins may be possible through the adaptation of some recently available glycan tagging methods. Here chemically modified sugars are taken up by cells and incorporated into glycoproteins and glycolipids via the normal biosynthetic pathways. The modified glycoconjugates are then detected using a specific chemically coupled probe, allowing enrichment of the tagged glycoproteins and their subsequent analysis by mass spectrometry (Laughlin et al. 2006). There are a variety of azide-specific chemistries available but the least cytotoxic for probing fucose incorporates an 6-alkyne fucose analogue and labels by Cu(I) catalysed azide–alkyne condensation, so called “click chemistry” (Hsu et al. 2007). The ability of POFUT2 to utilize GDP-fucose analogues, with C6 modifications on the sugar ring, could be verified through the expression of a TSR module in cells labelled with the alkyne fucose and analysis by methods established in our laboratory. The judicious use of an appropriate cell line could enhance the incorporation of the fucose into *O*-fucosylated glycans rather than *N*-linked structures.

Mutations in the gene encoding β3Glc-T (B3GALTL)

As discussed in Chapters 6.3 and 6.4, the further characterisation of β3Glc-T in control cells and healthy organisms will help our understanding of TSR glycobiology in disease states. In particular, we need to determine the function of the β3Glc-T N-terminal domain and that of the truncated β3Glc-T proteins predicted to exist in PPS.

The diagnosis of PPS is difficult as the spectrum of symptoms is varied and there are no formal diagnostic criteria. Patients can now be screened to identify mutations in the gene encoding β3Glc-T (B3GALTL). The majority of patients with PPS are homozygous for a hot-spot mutation in intron 8 (c.660 +1G>A) but additional mutations are described in GeneReviews (http://www.geneclinics.org) and in the Leiden Open Variation Database (http://chromium.liacs.nl/lovd/index.php?select_db=B3GALTL). Interestingly, in two clinically confirmed cases of PPS, one patient was homozygous for a truncating mutation in exon 13 (p.Tyr366X) (Table 4), while the other patient carried the c.1207G>A (p.Val403Ile) point mutation on one allele. As more patients suspected of having PPS are screened, the number of allelic variations in the gene encoding β3Glc-T is increasing.

DNA change (reported cases)	Expected protein	Reference
c.347+5G→A/ c.660+1G→A (2)	p.Glu91LeufsX6	(Lesnik Oberstein et al. 2006)
c.660+1G→A/ c.660+1G→A (36)	p.Thr201ProfsX13	(Lesnik Oberstein et al. 2006)
c.1098T→A/ c.1098T→A(1)	p.Tyr366X	GeneReviews
deletion of one of the alleles, with a *c.660+1G→A* mutation on the other allele (2)	p.Thr201ProfsX13	(Lesnik Oberstein et al. 2006)
c.1207G→A (1)	p.Val403Ile	Unpublished data

Table 4. *B3GALTL* allelic variations found in PPS.

Out of the other 24 unconfirmed cases of PPS, the majority have mutations on a single allele in the non-coding region where the impact on the protein is difficult to predict. Three patients from this group have point mutations that would lead to the following substitutions, Val245Met; Gly369Ser; and Lys370Glu. It would be interesting to determine what impact these mutations have on the β3Glc-T activity as haploinsufficiency or variations in β3Glc-T may increase glaucoma susceptibility in individuals that do not have PPS (Lesnik Oberstein et al. 2006).

Recently, a group in the US screened seven patients with PPS for the two major splice site mutations (c.347+5G>A and c.660+1G>A) and only found two patients homozygous for c.660+1G>A, and one patient heterozygous for the same mutation (Reis et al. 2007). A complete screening of the entire *β3Glc-T* gene is currently in progress for these seven patients. In addition, another group has identified a patient with PPS that does not carry mutations in the *β3Glc-T* gene (Dr. E. Morova, University of Nijmengen, the Netherlands, personal communications) and shows no apparent defect in the TSR glycosylation (J. Hofsteenge and D. Hess, unpublished observations). These later results suggest that there may be other components, apart from β3Glc-T, associated with PPS.

Alternative splicing of β3Glc-T

RNA splicing is carried out by spliceosomes, ribonucleoprotein complexes that recognise the exon-intron junctions and catalyse the precise removal of introns and the joining of exons. The pre-mRNA can yield many different mRNA leading to the translation of different protein isoforms with varied and sometimes conflicting functions. It is estimated that 74% of all human genes are alternatively spliced (Johnson et al. 2003). A search of the Alternative Splicing and Transcript Diversity 1.1 database (http://www.ebi.ac.uk/astd/main.html) revealed four different β3Glc-T protein isoforms, suggesting the presence of additional properties for β3Glc-T. Maintaining the normal ratio of alternative splice isoforms is important in preventing the onset of disease (Garcia-Blanco et al. 2004). The ability to monitor the levels of β3Glc-T isoforms in healthy people and truncated forms in PPS may be informative.

Greater than 15% of disease-causing mutations result in splicing defects (reviewed in (Faustino et al. 2003)). These mutations may lie in the recognition sequence for splicing (donor/acceptor splice sites, enhancer/silencer elements) or in the splicing factors that associate with these splice sequences. The family of SR (Ser/ Arg-rich) proteins are major splice factors that bind to regulatory elements in the pre-mRNA and recruit splicing machinery to splicing signals. Titration of these SR proteins affects the use of alternative splice sites. In Drosophila, the splicing factor SRp55/B52 controls eye size through the production of two alternative splice forms of the master control gene, *eyeless* (Fic et al. 2007). It is noteworthy that *Pax-6*, the human homologue of *eyeless*, also produces two alternative spliced isoforms that differ in their DNA-binding activity (Epstein et al. 1994).

In PPS the most prevalent mutations in the gene encoding β3Glc-T result in a lost splicing function by destroying a 5' splice site sequence. Although this is a particularly strong mutation, some splicing mutations are occasionally leaky and produce both the mutant and wild type transcripts from the same allele. Depending on the level of missed splicing a certain level of normal protein may be produced that could have a dramatic effect on the disease phenotype. In addition, since splicing

outcomes can be very cell-type specific, it would be worthwhile checking several purified cell types from patients with PPS for levels of wild-type β3Glc-T protein and activity.

There are a growing number of strategies aimed at altering splicing in human disease-causing genes, including antisense oligonucleotides (Hua et al. 2008), RNA reprogramming (Garcia-Blanco et al. 2004) and low molecular weight compounds (Yeo 2005). In a small compound screen the plant cytokinin kinetin was identified as a strong splicing enhancer for the *IKBKAP* gene that is involved in the splicing disorder, familial dysautonomia (Hims et al. 2007). The kinetin activity required a CAA element at the end of the affected *IKBKAP* exon 20. Interestingly, this element is also conserved in exon 8 of the *β3Glc-T* gene, the commonly omitted exon in PPS. Hence it would be interesting to test similar compounds in PPS cells for the enhanced inclusion of the β3Glc-T exon 8.

Developing an animal model for PPS

Patients with PPS may share the same mutation in the gene encoding β3Glc-T but the symptoms and their degree of severity varies. Treating the symptoms in PPS usually involves corneal transplantation to preserve the vision and some patients respond to growth hormone. The extent of mental retardation is particularly varied in PPS and is progressive, which suggests a potential therapeutic window (http://www.geneclinics.org).

How can a single genetic mutation in *β3Glc-T* (*B3GALTL*) be responsible for the various clinical manifestations observed in PPS? As discussed above, if β3Glc-T's crucial biological role is through its enzymatic activity then the pleiotropic symptoms found in PPS may be due to the altered properties of many key glycosylated TSR-containing proteins. It could be that the time and location of β3Glc-T activity is required for very specific developmental events. PPS may result from abnormal neural crest development (Maillette de Buy Wenniger-Prick et al. 2002). One of the major symptoms described in PPS is the anterior eye-chamber disorder, Peters' anomaly. The neural crest contributes many structures in the eye, including the stroma and endothelium of the eye and the stroma of the iris, which are all disrupted in PPS. In order to understand if β3Glc-T is expressed in neural crest-derived cells, its gene and protein expression pattern in the developing eye is required. If β3Glc-T is found to be expressed during the appropriate time for eye development, then it is technically possible to selectively inactivate it in neural crest stem cells (Ittner et al. 2005). This approach would hopefully provide a model to study Peters' anomaly, if not PPS.

Deletion of the gene encoding β3Glc-T in the mouse may reveal the enzyme's biological role in an organism. An alternative approach to try and gain some insight into the role of β3Glc-T in disease would be to try and make a mouse model of PPS. Some very elegant approaches in developing mouse models of CDG include engineering animals harbouring the equivalent hot-spot mutation observed in CDG patients (Wang et al. 2001; Cromphout et al. 2006; Thiel et al. 2006). This is important as most CDGs result from hypomorphic alleles rather than complete loss of function. Thus for a mouse model of PPS, donor splice mutations in the gene encoding β3Glc-T that result in an exon 5 or exon 8 skip with the equivalent frameshifts observed in the PPS patients, could be envisioned. These mice could be investigated for the multiple symptoms seen in PPS.

8. List of Figures and Tables

Figures

Tables

9. REFERENCES

Acar, M., H. Jafar-Nejad, et al. (2008). "Rumi Is a CAP10 Domain Glycosyltransferase that Modifies Notch and Is Required for Notch Signaling." *Cell* **132**(2): 247-58.

Adamowicz, M. and E. Pronicka (1996). "Carbohydrate deficient glycoprotein syndrome--like transferrin isoelectric focusing pattern in untreated fructosaemia." *Eur J Pediatr* **155**(4): 347-8.

Adams, J. C. and R. P. Tucker (2000). "The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development." *Dev Dyn* **218**(2): 280-99.

Aebi, M., A. Helenius, et al. (1999). "Carbohydrate-deficient glycoprotein syndromes become congenital disorders of glycosylation: an updated nomenclature for CDG. First International Workshop on CDGS." *Glycoconj J* **16**(11): 669-71.

Akasaka-Manya, K., H. Manya, et al. (2006). "Physical and functional association of human protein O-mannosyltransferases 1 and 2." *J Biol Chem* **281**(28): 19339-45.

Ambrosi, M., N. R. Cameron, et al. (2005). "Lectins: tools for the molecular understanding of the glycocode." *Org Biomol Chem* **3**(9): 1593-608.

Apweiler, R., H. Hermjakob, et al. (1999). "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database." *Biochim Biophys Acta* **1473**(1): 4-8.

Asch, A. S., S. Silbiger, et al. (1992). "Thrombospondin sequence motif (CSVTCG) is responsible for CD36 binding." *Biochem Biophys Res Commun* **182**(3): 1208-17.

Baboval, T., O. Koul, et al. (2000). "N-glycosylation site occupancy of rat alpha-1,3-fucosyltransferase IV and the effect of glycosylation on enzymatic activity." *Biochim Biophys Acta* **1475**(3): 383-9.

Benach, J., Y. T. Chou, et al. (2003). "Phospholipid-induced monomerization and signal-peptide-induced oligomerization of SecA." *J Biol Chem* **278**(6): 3628-38.

Bessette, P. H., F. Aslund, et al. (1999). "Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm." *Proc Natl Acad Sci U S A* **96**(24): 13703-8.

Bjoern, S., D. C. Foster, et al. (1991). "Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine." *J Biol Chem* **266**(17): 11051-7.

Boeggeman, E. E., B. Ramakrishnan, et al. (2003). "The N-terminal stem region of bovine and human beta1,4-galactosyltransferase I increases the in vitro folding efficiency of their catalytic domain from inclusion bodies." *Protein Expr Purif* **30**(2): 219-29.

Bolos, V., J. Grego-Bessa, et al. (2007). "Notch signaling in development and cancer." *Endocr Rev* **28**(3): 339-63.

Borsig, L., E. G. Berger, et al. (1997). "Expression and purification of His-tagged beta-1,4-galactosyltransferase in yeast and in COS cells." *Biochem Biophys Res Commun* **240**(3): 586-9.

Borsig, L., A. G. Katopodis, et al. (1998). "Trafficking and localization studies of recombinant alpha1, 3-fucosyltransferase VI stably expressed in CHO cells." *Glycobiology* **8**(3): 259-68.

Brew, K. and R. L. Hill (1975). "Lactose biosynthesis." *Rev Physiol Biochem Pharmacol* **72**: 105-58.

Bruckner, K., L. Perez, et al. (2000). "Glycosyltransferase activity of Fringe modulates Notch-Delta interactions." *Nature* **406**(6794): 411-5.

Buko, A. M., E. J. Kentzer, et al. (1991). "Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator." *Proc Natl Acad Sci U S A* **88**(9): 3992-6.

Burstyn-Cohen, T., A. Frumkin, et al. (1998). "Accumulation of F-spondin in injured peripheral nerve promotes the outgrowth of sensory axons." *J Neurosci* **18**(21): 8875-85.

Burstyn-Cohen, T., V. Tzarfaty, et al. (1999). "F-Spondin is required for accurate pathfinding of commissural axons at the floor plate." *Neuron* **23**(2): 233-46.

Butters, T. D. (2007). "Pharmacotherapeutic strategies using small molecules for the treatment of glycolipid lysosomal storage disorders." *Expert Opin Pharmacother* **8**(4): 427-35.

Campbell, J. A., G. J. Davies, et al. (1997). "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities." *Biochem J* **326 (Pt 3)**: 929-39.

Caramelo, J. J., O. A. Castro, et al. (2004). "The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates." *J Biol Chem* **279**(44): 46280-5.

Chakravarti, S., T. Magnuson, et al. (1998). "Lumican regulates collagen fibril assembly: skin fragility and corneal opacity in the absence of lumican." *J Cell Biol* **141**(5): 1277-86.

Chang, J. Y. (1994). "Controlling the speed of hirudin folding." *Biochem J* **300 (Pt 3)**: 643-50.

Chang, V. T., M. Crispin, et al. (2007). "Glycoprotein structural genomics: solving the glycosylation problem." *Structure* **15**(3): 267-73.

Charlwood, J., P. Clayton, et al. (1998). "Defective galactosylation of serum transferrin in galactosemia." *Glycobiology* **8**(4): 351-7.

Charnock, S. J. and G. J. Davies (1999). "Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms." *Biochemistry* **38**(20): 6380-5.

Chen, J., L. Lu, et al. (2006). "Expression of Notch signaling pathway genes in mouse embryos lacking beta4galactosyltransferase-1." *Gene Expr Patterns* **6**(4): 376-82.

Chen, J., D. J. Moloney, et al. (2001). "Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1." *Proc Natl Acad Sci U S A* **98**(24): 13716-21.

Cheng, X., R. N. Cole, et al. (2000). "Alternative O-glycosylation/O-phosphorylation of the murine estrogen receptor beta." *Biochemistry* **39**(38): 11609-20.

Chigira, Y., T. Oka, et al. (2008). "Engineering of a mammalian O-glycosylation pathway in the yeast Saccharomyces cerevisiae: production of O-fucosylated epidermal growth factor domains." *Glycobiology*.

Chiu, C. P., A. G. Watts, et al. (2004). "Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog." *Nat Struct Mol Biol* **11**(2): 163-70.

Christensen, L. L., U. B. Jensen, et al. (2000). "The C-terminal N-glycosylation sites of the human alpha1,3/4-fucosyltransferase III, -V, and -VI (hFucTIII, -V, adn -VI) are necessary for the expression of full enzyme activity." *Glycobiology* **10**(9): 931-9.

Clark, R. J., P. M. McDonough, et al. (2003). "Diabetes and the accompanying hyperglycemia impairs cardiomyocyte calcium cycling through increased nuclear O-GlcNAcylation." *J Biol Chem* **278**(45): 44230-7.

Coutinho, P. M., E. Deleury, et al. (2003). "An evolving hierarchical family classification for glycosyltransferases." *J Mol Biol* **328**(2): 307-17.

Cromphout, K., W. Vleugels, et al. (2006). "The normal phenotype of Pmm1-deficient mice suggests that Pmm1 is not essential for normal mouse development." *Mol Cell Biol* **26**(15): 5621-35.

Davies, G. J., T. M. Gloster, et al. (2005). "Recent structural insights into the expanding world of carbohydrate-active enzymes." *Curr Opin Struct Biol* **15**(6): 637-45.

Dawson, D. W., S. F. Pearce, et al. (1997). "CD36 mediates the In vitro inhibitory effects of thrombospondin-1 on endothelial cells." *J Cell Biol* **138**(3): 707-17.

de Beer, T., J. F. Vliegenthart, et al. (1995). "The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase Us is alpha-mannopyranose." *Biochemistry* **34**(37): 11785-9.

de Marco, A. and V. De Marco (2004). "Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning." *J Biotechnol* **109**(1-2): 45-52.

de Marco, A., E. Deuerling, et al. (2007). "Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli." *BMC Biotechnol* **7**: 32.

del Castillo Busto, M. E., M. Montes-Bayon, et al. (2005). "Strategies to study human serum transferrin isoforms using integrated liquid chromatography ICPMS, MALDI-TOF, and ESI-Q-TOF detection: application to chronic alcohol abuse." *Anal Chem* **77**(17): 5615-21.

Dennis, J. W., S. Laferte, et al. (1987). "Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis." *Science* **236**(4801): 582-5.

Doucey, M. A., D. Hess, et al. (1999). "Recombinant human interleukin-12 is the second example of a C-mannosylated protein." *Glycobiology* **9**(5): 435-41.

Doucey, M. A., D. Hess, et al. (1998). "Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor." *Mol Biol Cell* **9**(2): 291-300.

Du, J., H. Tekeuchi, et al. (2007). *O-Fucosylation of Thrombospondin Type 1 Repeats Is Essential for Normal Gastrulation in Mouse Embryos*. Annual Conference of the Society for Glycobiology, Boston, MA, Glycobiology.

Du, X. L., D. Edelstein, et al. (2000). "Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing Sp1 glycosylation." *Proc Natl Acad Sci U S A* **97**(22): 12222-6.

Eichler, R., O. Lenz, et al. (2003). "Identification of Lassa virus glycoprotein signal peptide as a trans-acting maturation factor." *EMBO Rep* **4**(11): 1084-8.

Ellgaard, L., M. Molinari, et al. (1999). "Setting the standards: quality control in the secretory pathway." *Science* **286**(5446): 1882-8.

Elling, L., A. Zervosen, et al. (1999). "UDP-N-Acetyl-alpha-D-glucosamine as acceptor substrate of beta-1,4-galactosyltransferase. Enzymatic synthesis of UDP-N-acetyllactosamine." *Glycoconj J* **16**(7): 327-36.

Epstein, J. A., T. Glaser, et al. (1994). "Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing." *Genes Dev* **8**(17): 2022-34.

Evrard, Y. A., Y. Lun, et al. (1998). "lunatic fringe is an essential mediator of somite segmentation and patterning." *Nature* **394**(6691): 377-81.

Fan, J. Q. (2008). "A counterintuitive approach to treat enzyme deficiencies: use of enzyme inhibitors for restoring mutant enzyme activity." *Biol Chem* **389**(1): 1-11.

Fast, D. G., J. C. Jamieson, et al. (1993). "The role of the carbohydrate chains of Gal beta-1,4-GlcNAc alpha 2,6-sialyltransferase for enzyme activity." *Biochim Biophys Acta* **1202**(2): 325-30.

Faustino, N. A. and T. A. Cooper (2003). "Pre-mRNA splicing and human disease." *Genes Dev* **17**(4): 419-37.

Fic, W., F. Juge, et al. (2007). "Eye development under the control of SRp55/B52-mediated alternative splicing of eyeless." *PLoS ONE* **2**(2): e253.

Finn, R. D., J. Mistry, et al. (2006). "Pfam: clans, web tools and services." *Nucleic Acids Res* **34**(Database issue): D247-51.

Freeze, H. H. (2006). "Genetic defects in the human glycome." *Nat Rev Genet* **7**(7): 537-51.

Freeze, H. H. (2007). "Congenital Disorders of Glycosylation: CDG-I, CDG-II, and beyond." *Curr Mol Med* **7**(4): 389-96.

Furmanek, A., D. Hess, et al. (2003). "The WSAWS motif is C-hexosylated in a soluble form of the erythropoietin receptor." *Biochemistry* **42**(28): 8452-8.

Futerman, A. H. and G. van Meer (2004). "The cell biology of lysosomal storage disorders." *Nat Rev Mol Cell Biol* **5**(7): 554-65.

Gade, G., R. Kellner, et al. (1992). "A tryptophan-substituted member of the AKH/RPCH family isolated from a stick insect corpus cardiacum." *Biochem Biophys Res Commun* **189**(3): 1303-9.

Garcia-Blanco, M. A., A. P. Baraniak, et al. (2004). "Alternative splicing in disease and therapy." *Nat Biotechnol* **22**(5): 535-46.

Ge, C. and P. Stanley (2008). "The O-fucose glycan in the ligand-binding domain of Notch1 regulates embryogenesis and T cell development." *Proc Natl Acad Sci U S A* **105**(5): 1539-44.

Gonzalez de Peredo, A., D. Klein, et al. (2002). "C-mannosylation and o-fucosylation of thrombospondin type 1 repeats." *Mol Cell Proteomics* **1**(1): 11-8.

Grunewald, S. (2007). "Congenital disorders of glycosylation: Rapidly enlarging group of (neuro)metabolic disorders." *Early Hum Dev*.

Haines, N. and K. D. Irvine (2003). "Glycosylation regulates Notch signalling." *Nat Rev Mol Cell Biol* **4**(10): 786-97.

Haines, N. and K. D. Irvine (2005). "Functional analysis of Drosophila beta1,4-N-acetlygalactosaminyltransferases." *Glycobiology* **15**(4): 335-46.

Hallgren, P., A. Lundblad, et al. (1975). "A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine." *J Biol Chem* **250**(14): 5312-4.

Harris, R. J. and M. W. Spellman (1993). "O-linked fucose and other post-translational modifications unique to EGF modules." *Glycobiology* **3**(3): 219-24.

Harris, R. J., H. van Halbeek, et al. (1993). "Identification and structural analysis of the tetrasaccharide NeuAc alpha(2-->6)Gal beta(1-->4)GlcNAc beta(1-->3)Fuc alpha 1-->O-linked to serine 61 of human factor IX." *Biochemistry* **32**(26): 6539-47.

Hart, G. W., M. P. Housley, et al. (2007). "Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins." *Nature* **446**(7139): 1017-22.

Hartmann, S. and J. Hofsteenge (2000). "Properdin, the positive regulator of complement, is highly C-mannosylated." *J Biol Chem* **275**(37): 28569-74.

Hase, S., S. Kawabata, et al. (1988). "A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factors VII and IX." *J Biochem* **104**(6): 867-8.

Hase, S., H. Nishimura, et al. (1990). "The structure of (xylose)2glucose-O-serine 53 found in the first epidermal growth factor-like domain of bovine blood clotting factor IX." *J Biol Chem* **265**(4): 1858-61.

Hauri, H., C. Appenzeller, et al. (2000). "Lectins and traffic in the secretory pathway." *FEBS Lett* **476**(1-2): 32-7.

Hebert, D. N. and M. Molinari (2007). "In and out of the ER: protein folding, quality control, degradation, and related human diseases." *Physiol Rev* **87**(4): 1377-408.

Heger, A., S. Mallick, et al. (2007). "The global trace graph, a novel paradigm for searching protein sequence databases." *Bioinformatics* **23**(18): 2361-7.

Heinonen, T. Y., L. Pasternack, et al. (2003). "A novel human glycosyltransferase: primary structure and characterization of the gene and transcripts." *Biochem Biophys Res Commun* **309**(1): 166-74.

Heinonen, T. Y., M. Pelto-Huikko, et al. (2006). "Murine ortholog of the novel glycosyltransferase, B3GTL: primary structure, characterization of the gene and transcripts, and expression in tissues." *DNA Cell Biol* **25**(8): 465-74.

Helenius, A., T. Marquardt, et al. (1992). "The endoplasmic reticulum as a protein-folding compartment." *Trends Cell Biol* **2**(8): 227-31.

Helmus, Y., J. Denecke, et al. (2006). "Leukocyte adhesion deficiency II patients with a dual defect of the GDP-fucose transporter." *Blood* **107**(10): 3959-66.

Hess, D., J. J. Keusch, et al. (2008). "Peters plus syndrome is a new congenital disorder of glycosylation and involves defective O-glycosylation of thrombospondin type 1 repeats." *J Biol Chem*.

Hims, M. M., C. Ibrahim el, et al. (2007). "Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia." *J Mol Med* **85**(2): 149-61.

Hirschberg, C. B., P. W. Robbins, et al. (1998). "Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus." *Annu Rev Biochem* **67**: 49-69.

Ho, A. and T. C. Sudhof (2004). "Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage." *Proc Natl Acad Sci U S A* **101**(8): 2548-53.

Hoe, H. S., D. Wessner, et al. (2005). "F-spondin interaction with the apolipoprotein E receptor ApoEr2 affects processing of amyloid precursor protein." *Mol Cell Biol* **25**(21): 9259-68.

Hofsteenge, J., M. Blommers, et al. (1999). "The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues." *J Biol Chem* **274**(46): 32786-94.

Hofsteenge, J., K. G. Huwiler, et al. (2001). "C-mannosylation and O-fucosylation of the thrombospondin type 1 module." *J Biol Chem* **276**(9): 6485-98.

Hofsteenge, J., D. R. Muller, et al. (1994). "New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us." *Biochemistry* **33**(46): 13524-30.

Hsu, T. L., S. R. Hanson, et al. (2007). "Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells." *Proc Natl Acad Sci U S A* **104**(8): 2614-9.

Hu, Y. and S. Walker (2002). "Remarkable structural similarities between diverse glycosyltransferases." *Chem Biol* **9**(12): 1287-96.

Hua, Y., T. A. Vickers, et al. (2008). "Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice." *Am J Hum Genet* **82**(4): 834-48.

Hwang, C., H. F. Lodish, et al. (1995). "Measurement of glutathione redox state in cytosol and secretory pathway of cultured cells." *Methods Enzymol* **251**: 212-21.

Ittner, L. M., H. Wurdak, et al. (2005). "Compound developmental eye disorders following inactivation of TGFbeta signaling in neural-crest stem cells." *J Biol* **4**(3): 11.

Jaeken, J. and G. Matthijs (2001). "Congenital disorders of glycosylation." *Annu Rev Genomics Hum Genet* **2**: 129-51.

Jaeken, J. and G. Matthijs (2007). "Congenital disorders of glycosylation: a rapidly expanding disease family." *Annu Rev Genomics Hum Genet* **8**: 261-78.

Jaeken, J., M. Pirard, et al. (1996). "Inhibition of phosphomannose isomerase by fructose 1-phosphate: an explanation for defective N-glycosylation in hereditary fructose intolerance." *Pediatr Res* **40**(5): 764-6.

Jimenez, B., O. V. Volpert, et al. (2001). "c-Jun N-terminal kinase activation is required for the inhibition of neovascularization by thrombospondin-1." *Oncogene* **20**(26): 3443-8.

Jinek, M., Y. W. Chen, et al. (2006). "Structural insights into the Notch-modifying glycosyltransferase Fringe." *Nat Struct Mol Biol* **13**(10): 945-6.

Johnson, J. M., J. Castle, et al. (2003). "Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays." *Science* **302**(5653): 2141-4.

Johnston, S. H., C. Rauskolb, et al. (1997). "A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway." *Development* **124**(11): 2245-54.

Jones, J., S. S. Krag, et al. (2005). "Controlling N-linked glycan site occupancy." *Biochim Biophys Acta* **1726**(2): 121-37.

Joutel, A., C. Corpechot, et al. (1996). "Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia." *Nature* **383**(6602): 707-10.

Ju, T. and R. D. Cummings (2002). "A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase." *Proc Natl Acad Sci U S A* **99**(26): 16613-8.

Kamemura, K. and G. W. Hart (2003). "Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription." *Prog Nucleic Acid Res Mol Biol* **73**: 107-36.

Kansas, G. S. (1996). "Selectins and their ligands: current concepts and controversies." *Blood* **88**(9): 3259-87.

Kaplan, A., D. T. Achord, et al. (1977). "Phosphohexosyl components of a lysosomal enzyme are recognized by pinocytosis receptors on human fibroblasts." *Proc Natl Acad Sci U S A* **74**(5): 2026-30.

Kellenberger, C., C. Boudier, et al. (1995). "Serine protease inhibition by insect peptides containing a cysteine knot and a triple-stranded beta-sheet." *J Biol Chem* **270**(43): 25514-9.

Kitagawa, H., T. Uyama, et al. (2001). "Molecular cloning and expression of a human chondroitin synthase." *J Biol Chem* **276**(42): 38721-6.

Klar, A., M. Baldassare, et al. (1992). "F-spondin: a gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension." *Cell* **69**(1): 95-110.

Kornfeld, R. and S. Kornfeld (1985). "Assembly of asparagine-linked oligosaccharides." *Annu Rev Biochem* **54**: 631-64.

Kozma, K., J. J. Keusch, et al. (2006). "Identification and characterization of abeta1,3-glucosyltransferase that synthesizes the Glc-beta1,3-Fuc disaccharide on thrombospondin type 1 repeats." *J Biol Chem* **281**(48): 36742-51.

Kranz, C., C. Jungeblut, et al. (2007). "A defect in dolichol phosphate biosynthesis causes a new inherited disorder with death in early infancy." *Am J Hum Genet* **80**(3): 433-40.

Krieg, J., W. Glasner, et al. (1997). "C-Mannosylation of human RNase 2 is an intracellular process performed by a variety of cultured cells." *J Biol Chem* **272**(42): 26687-92.

Krieg, J., S. Hartmann, et al. (1998). "Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp." *Mol Biol Cell* **9**(2): 301-9.

Kukowska-Latallo, J. F., R. D. Larsen, et al. (1990). "A cloned human cDNA determines expression of a mouse stage-specific embryonic antigen and the Lewis blood group alpha(1,3/1,4)fucosyltransferase." *Genes Dev* **4**(8): 1288-303.

Labaied, M., N. Camargo, et al. (2007). "Depletion of the Plasmodium berghei thrombospondin-related sporozoite protein reveals a role in host cell entry by sporozoites." *Mol Biochem Parasitol* **153**(2): 158-66.

Lairson, L. L. and S. G. Withers (2004). "Mechanistic analogies amongst carbohydrate modifying enzymes." *Chem Commun (Camb)*(20): 2243-8.

Laughlin, S. T., N. J. Agard, et al. (2006). "Metabolic labeling of glycans with azido sugars for visualization and glycoproteomics." *Methods Enzymol* **415**: 230-50.

Lawler, J. (2000). "The functions of thrombospondin-1 and-2." *Curr Opin Cell Biol* **12**(5): 634-40.

Lawler, J. and R. O. Hynes (1986). "The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins." *J Cell Biol* **103**(5): 1635-48.

Lei, L., A. Xu, et al. (2003). "An O-fucose site in the ligand binding domain inhibits Notch activation." *Development* **130**(26): 6411-21.

Leroy, J. G. (2006). "Congenital disorders of N-glycosylation including diseases associated with O- as well as N-glycosylation defects." *Pediatr Res* **60**(6): 643-56.

Lesk, A. M. (1995). "Systematic representation of protein folding patterns." *J Mol Graph* **13**(3): 159-64.

Lesnik Oberstein, S. A., M. Kriek, et al. (2006). "Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase." *Am J Hum Genet* **79**(3): 562-6.

Levy, R., R. Weiss, et al. (2001). "Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones." *Protein Expr Purif* **23**(2): 338-47.

Li, L., I. D. Krantz, et al. (1997). "Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1." *Nat Genet* **16**(3): 243-51.

Li, Q. M., H. van den Heuvel, et al. (1991). "Mass spectral characterization of C-glycosidic flavonoids isolated from a medicinal plant (Passiflora incarnata)." *J Chromatogr* **562**(1-2): 435-46.

Liu, K., A. J. Paterson, et al. (2000). "Glucose stimulates protein modification by O-linked GlcNAc in pancreatic beta cells: linkage of O-linked GlcNAc to beta cell death." *Proc Natl Acad Sci U S A* **97**(6): 2820-5.

Löffler, A., M. A. Doucey, et al. (1996). "Spectroscopic and protein chemical analyses demonstrate the presence of C-mannosylated tryptophan in intact human RNase 2 and its isoforms." *Biochemistry* **35**(37): 12005-14.

Lord, J. M., J. Davey, et al. (2000). "Endoplasmic reticulum-associated protein degradation." *Semin Cell Dev Biol* **11**(3): 159-64.

Loriol, C., A. Audfray, et al. (2007). "The two N-glycans present on bovine Pofut1 are differently involved in its solubility and activity." *Febs J* **274**(5): 1202-11.

Lu, L. and P. Stanley (2006). "Roles of O-fucose glycans in notch signaling revealed by mutant mice." *Methods Enzymol* **417**: 127-36.

Ludemann, N., A. Clement, et al. (2005). "O-glycosylation of the tail domain of neurofilament protein M in human neurons and in spinal cord tissue of a rat model of amyotrophic lateral sclerosis (ALS)." *J Biol Chem* **280**(36): 31648-58.

Luo, Y. and R. S. Haltiwanger (2005). "O-fucosylation of notch occurs in the endoplasmic reticulum." *J Biol Chem* **280**(12): 11289-94.

Luo, Y., K. Koles, et al. (2006b). "Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats." *J Biol Chem* **281**(14): 9393-9.

Luo, Y. a., A. Nita-Lazar, et al. (2006a). "Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats." *J Biol Chem* **281**(14): 9385-92.

Ma, J. and K. J. Colley (1996). "A disulfide-bonded dimer of the Golgi beta-galactoside alpha2,6-sialyltransferase is catalytically inactive yet still retains the ability to bind galactose." *J Biol Chem* **271**(13): 7758-66.

Maccioni, H. J., C. G. Giraudo, et al. (2002). "Understanding the stepwise synthesis of glycolipids." *Neurochem Res* **27**(7-8): 629-36.

Maillette de Buy Wenniger-Prick, L. J. and R. C. Hennekam (2002). "The Peters' plus syndrome: a review." *Ann Genet* **45**(2): 97-103.

Malissard, M., L. Borsig, et al. (1996). "Recombinant soluble beta-1,4-galactosyltransferases expressed in Saccharomyces cerevisiae. Purification, characterization and comparison with human enzyme." *Eur J Biochem* **239**(2): 340-8.

Marek, K. W., I. K. Vijay, et al. (1999). "A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality." *Glycobiology* **9**(11): 1263-71.

Martina, J. A., J. L. Daniotti, et al. (1998). "Influence of N-glycosylation and N-glycan trimming on the activity and intracellular traffic of GD3 synthase." *J Biol Chem* **273**(6): 3725-31.

Martinez-Duncker, I., R. Mollicone, et al. (2003). "The nucleotide-sugar transporter family: a phylogenetic approach." *Biochimie* **85**(3-4): 245-60.

Matuschewski, K., A. C. Nunes, et al. (2002). "Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system." *Embo J* **21**(7): 1597-606.

Medina, L., K. Grove, et al. (1998). "SV40 large T antigen is modified with O-linked N-acetylglucosamine but not with other forms of glycosylation." *Glycobiology* **8**(4): 383-91.

Mer, G., H. Hietter, et al. (1996). "Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors." *J Mol Biol* **258**(1): 158-71.

Mer, G., H. Hietter, et al. (1996). "Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated proteinase inhibitor." *Nat Struct Biol* **3**(1): 45-53.

Meyer, K. (1958). "Chemical structure of hyaluronic acid." *Fed Proc* **17**(4): 1075-7.

Miao, W. M., W. L. Seng, et al. (2001). "Thrombospondin-1 type 1 repeat recombinant proteins inhibit tumor growth through transforming growth factor-beta-dependent and -independent mechanisms." *Cancer Res* **61**(21): 7830-9.

Minowa, M. T., S. Oguri, et al. (1998). "cDNA cloning and expression of bovine UDP-N-acetylglucosamine: alpha1, 3-D-mannoside beta1,4-N-acetylglucosaminyltransferase IV." *J Biol Chem* **273**(19): 11556-62.

Missiakas, D., C. Georgopoulos, et al. (1993). "Identification and characterization of the Escherichia coli gene dsbB, whose product is involved in the formation of disulfide bonds in vivo." *Proc Natl Acad Sci U S A* **90**(15): 7084-8.

Moloney, D. J. and R. S. Haltiwanger (1999). "The O-linked fucose glycosylation pathway: identification and characterization of a uridine diphosphoglucose: fucose-beta1,3-glucosyltransferase activity from Chinese hamster ovary cells." *Glycobiology* **9**(7): 679-87.

Moloney, D. J., A. I. Lin, et al. (1997). "The O-linked fucose glycosylation pathway. Evidence for protein-specific elongation of o-linked fucose in Chinese hamster ovary cells." *J Biol Chem* **272**(30): 19046-50.

Moloney, D. J., V. M. Panin, et al. (2000). "Fringe is a glycosyltransferase that modifies Notch." *Nature* **406**(6794): 369-75.

Moloney, D. J., L. H. Shair, et al. (2000). "Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules." *J Biol Chem* **275**(13): 9604-11.

Moran, J. L., J. M. Levorse, et al. (1999). "Limbs move beyond the radical fringe." *Nature* **399**(6738): 742-3.

Munte, C. E., G. Gade, et al. (2008). "C-mannosylation in the hypertrehalosaemic hormone from the stick insect Carausius morosus." *Febs J*.

Nakakura, N., H. Hietter, et al. (1992). "Isolation and structural determination of three peptides from the insect Locusta migratoria. Identification of a deoxyhexose-linked peptide." *Eur J Biochem* **204**(1): 147-53.

Nilsson, T., C. Rabouille, et al. (1996). "The role of the membrane-spanning domain and stalk region of N-acetylglucosaminyltransferase I in retention, kin recognition and structural maintenance of the Golgi apparatus in HeLa cells." *J Cell Sci* **109 (Pt 7)**: 1975-89.

Nishimura, H., S. Kawabata, et al. (1989). "Identification of a disaccharide (Xyl-Glc) and a trisaccharide (Xyl2-Glc) O-glycosidically linked to a serine residue in the first epidermal growth factor-like domain of human factors VII and IX and protein Z and bovine protein Z." *J Biol Chem* **264**(34): 20320-5.

Nishimura, H., S. Yamashita, et al. (1992). "Evidence for the existence of O-linked sugar chains consisting of glucose and xylose in bovine thrombospondin." *J Biochem* **111**(4): 460-4.

Oda, T., A. G. Elkahloun, et al. (1997). "Identification and cloning of the human homolog (JAG1) of the rat Jagged1 gene from the Alagille syndrome critical region at 20p12." *Genomics* **43**(3): 376-9.

Ohno, K., I. Yuasa, et al. (1992). "The carbohydrate deficient glycoprotein syndrome in three Japanese children." *Brain Dev* **14**(1): 30-5.

Ohtsubo, K. and J. D. Marth (2006). "Glycosylation in cellular mechanisms of health and disease." *Cell* **126**(5): 855-67.

Okajima, T. and K. D. Irvine (2002). "Regulation of notch signaling by o-linked fucose." *Cell* **111**(6): 893-904.

Okajima, T., A. Matsuura, et al. (2008). "Biological functions of glycosyltransferase genes involved in O-fucose glycan synthesis." *J Biochem*.

Okajima, T., B. Reddy, et al. (2008). "Contributions of chaperone and glycosyltransferase activities of O-fucosyltransferase 1 to Notch signaling." *BMC Biol* **6**: 1.

Okajima, T., A. Xu, et al. (2003). "Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe." *J Biol Chem* **278**(43): 42340-5.

Okajima, T., A. Xu, et al. (2005). "Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding." *Science* **307**(5715): 1599-603.

Ornitz, D. M., A. B. Herr, et al. (1995). "FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides." *Science* **268**(5209): 432-6.

Orntoft, T. F. and E. M. Vestergaard (1999). "Clinical aspects of altered glycosylation of glycoproteins in cancer." *Electrophoresis* **20**(2): 362-71.

Ouzzine, M., S. Gulberti, et al. (2000). "Structure/function of the human Ga1beta1,3-glucuronosyltransferase. Dimerization and functional activity are mediated by two crucial cysteine residues." *J Biol Chem* **275**(36): 28254-60.

Paakkonen, K., H. Tossavainen, et al. (2006). "Solution structures of the first and fourth TSR domains of F-spondin." *Proteins* **64**(3): 665-72.

Panin, V. M., V. Papayannopoulos, et al. (1997). "Fringe modulates Notch-ligand interactions." *Nature* **387**(6636): 908-12.

Panin, V. M., L. Shao, et al. (2002). "Notch ligands are substrates for protein O-fucosyltransferase-1 and Fringe." *J Biol Chem* **277**(33): 29945-52.

Paulson, J. C. and K. J. Colley (1989). "Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation." *J Biol Chem* **264**(30): 17615-8.

Pedersen, L. C., K. Tsuchida, et al. (2000). "Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I." *J Biol Chem* **275**(44): 34580-5.

Pelham, H. R. (1988). "Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment." *Embo J* **7**(4): 913-8.

Perez-Vilar, J., S. H. Randell, et al. (2004). "C-Mannosylation of MUC5AC and MUC5B Cys subdomains." *Glycobiology* **14**(4): 325-37.

Pidoux, A. L. and J. Armstrong (1992). "Analysis of the BiP gene and identification of an ER retention signal in Schizosaccharomyces pombe." *Embo J* **11**(4): 1583-91.

Qiu, X., H. Xu, et al. (2004). "Sequence and embryonic expression of three zebrafish fringe genes: lunatic fringe, radical fringe, and manic fringe." *Dev Dyn* **231**(3): 621-30.

Rabbani, S. A., A. P. Mazar, et al. (1992). "Structural requirements for the growth factor activity of the amino-terminal domain of urokinase." *J Biol Chem* **267**(20): 14151-6.

Raman, R., S. Raguram, et al. (2005). "Glycomics: an integrated systems approach to structure-function relationships of glycans." *Nat Methods* **2**(11): 817-24.

Rampal, R., A. S. Li, et al. (2005). "Lunatic fringe, manic fringe, and radical fringe recognize similar specificity determinants in O-fucosylated epidermal growth factor-like repeats." *J Biol Chem* **280**(51): 42454-63.

Rathore, D., J. B. Sacci, et al. (2002). "Binding and invasion of liver cells by Plasmodium falciparum sporozoites. Essential involvement of the amino terminus of circumsporozoite protein." *J Biol Chem* **277**(9): 7092-8.

Reis, L., J. Toonen, et al. (2007). Peters-Plus syndrome: a review of 7 cases and results of B3GALTL screening. *Annual Clinical Genetic Meeting*. Nashville, TN, USA.

Ricketts, L. M., M. Dlugosz, et al. (2007). "O-fucosylation is required for ADAMTS13 secretion." *J Biol Chem* **282**(23): 17014-23.

Saito, T., E. Miyoshi, et al. (2002). "A secreted type of beta 1,6-N-acetylglucosaminyltransferase V (GnT-V) induces tumor angiogenesis without mediation of glycosylation: a novel function of GnT-V distinct from the original glycosyltransferase activity." *J Biol Chem* **277**(19): 17002-8.

Sasaki, N., T. Sasamura, et al. (2007). "Polarized exocytosis and transcytosis of Notch during its apical localization in Drosophila epithelial cells." *Genes Cells* **12**(1): 89-103.

Sasamura, T., H. O. Ishikawa, et al. (2007). "The O-fucosyltransferase O-fut1 is an extracellular component that is essential for the constitutive endocytic trafficking of Notch in Drosophila." *Development* **134**(7): 1347-56.

Sato, T., M. Sato, et al. (2006). "Molecular cloning and characterization of a novel human beta1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain." *Glycobiology* **16**(12): 1194-206.

Schiffer, S. G., S. Foley, et al. (2001). "Fucosylation of Cripto is required for its ability to facilitate nodal signaling." *J Biol Chem* **276**(41): 37769-78.

Scott, R. H., V. J. Gorton, et al. (1997). "Inhibition of neuronal high voltage-activated calcium channels by insect peptides: a comparison with the actions of omega-conotoxin GVIA." *Neuropharmacology* **36**(2): 195-208.

Shafi, R., S. P. Iyer, et al. (2000). "The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny." *Proc Natl Acad Sci U S A* **97**(11): 5735-9.

Shao, L., Y. Luo, et al. (2002). "O-glycosylation of EGF repeats: identification and initial characterization of a UDP-glucose: protein O-glucosyltransferase." *Glycobiology* **12**(11): 763-70.

Sherman, L. S. and S. A. Back (2008). "A 'GAG' reflex prevents repair of the damaged CNS." *Trends Neurosci* **31**(1): 44-52.

Shi, S., C. Ge, et al. (2007). "The threonine that carries fucose, but not fucose, is required for Cripto to facilitate Nodal signaling." *J Biol Chem* **282**(28): 20133-41.

Shimizu, K., S. Chiba, et al. (2001). "Manic fringe and lunatic fringe modify different sites of the Notch2 extracellular region, resulting in different signaling modulation." *J Biol Chem* **276**(28): 25753-8.

Slawson, C., M. P. Housley, et al. (2006). "O-GlcNAc cycling: how a single sugar post-translational modification is changing the way we think about signaling networks." *J Cell Biochem* **97**(1): 71-83.

Soejima, K. and T. Nakagaki (2005). "Interplay between ADAMTS13 and von Willebrand factor in inherited and acquired thrombotic microangiopathies." *Semin Hematol* **42**(1): 56-62.

Sparrow, D. B., G. Chapman, et al. (2006). "Mutation of the LUNATIC FRINGE gene in humans causes spondylocostal dysostosis with a severe vertebral phenotype." *Am J Hum Genet* **78**(1): 28-37.

Spiro, R. G. (2002). "Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds." *Glycobiology* **12**(4): 43R-56R.

Sprong, H., S. Degroote, et al. (2003). "Association of the Golgi UDP-galactose transporter with UDP-galactose:ceramide galactosyltransferase allows UDP-galactose import in the endoplasmic reticulum." *Mol Biol Cell* **14**(8): 3482-93.

Stahl, M., K. Uemura, et al. (2008). "Roles of Pofut1 and O-fucose in mammalian notch signaling." *J Biol Chem*.

Sturiale, L., R. Barone, et al. (2005). "Hypoglycosylation with increased fucosylation and branching of serum transferrin N-glycans in untreated galactosemia." *Glycobiology* **15**(12): 1268-76.

Sultan, A. A., V. Thathy, et al. (1997). "TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites." *Cell* **90**(3): 511-22.

Szabady, R. L., J. H. Peterson, et al. (2005). "An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter." *Proc Natl Acad Sci U S A* **102**(1): 221-6.

Tan, K., M. Duquette, et al. (2002). "Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication." *J Cell Biol* **159**(2): 373-82.

Taylor, M. E. and K. Drickamer (2003). "Introduction to Glycobiology."

Ten Hagen, K. G., T. A. Fritz, et al. (2003). "All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases." *Glycobiology* **13**(1): 1R-16R.

Thiel, C., T. Lubke, et al. (2006). "Targeted disruption of the mouse phosphomannomutase 2 gene causes early embryonic lethality." *Mol Cell Biol* **26**(15): 5615-20.

Toki, D., M. Sarkar, et al. (1997). "Expression of stable human O-glycan core 2 beta-1,6-N-acetylglucosaminyltransferase in Sf9 insect cells." *Biochem J* **325 (Pt 1)**: 63-9.

Torres, C. R. and G. W. Hart (1984). "Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc." *J Biol Chem* **259**(5): 3308-17.

Tortorella, M., M. Pratta, et al. (2000). "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for aggrecan substrate recognition and cleavage." *J Biol Chem* **275**(33): 25791-7.

Tossavainen, H., T. Pihlajamaa, et al. (2006). "The layered fold of the TSR domain of P. falciparum TRAP contains a heparin binding site." *Protein Sci* **15**(7): 1760-8.

Tucker, R. P. (2004). "The thrombospondin type 1 repeat superfamily." *Int J Biochem Cell Biol* **36**(6): 969-74.

Tzarfati-Majar, V., T. Burstyn-Cohen, et al. (2001). "F-spondin is a contact-repellent molecule for embryonic motor neurons." *Proc Natl Acad Sci U S A* **98**(8): 4722-7.

Unligil, U. M. and J. M. Rini (2000). "Glycosyltransferase structure and mechanism." *Curr Opin Struct Biol* **10**(5): 510-7.

Varki, A. (1993). "Biological roles of oligosaccharides: all of the theories are correct." *Glycobiology* **3**(2): 97-130.

Varki, A., R. Cummings, et al. (1999). "Essentials of Glycobiology " *Cold Spring Habor Laboratory Press*.

Vellodi, A. (2005). "Lysosomal storage disorders." *Br J Haematol* **128**(4): 413-31.

Wada, Y. (2006). "Mass spectrometry for congenital disorders of glycosylation, CDG." *Journal of Chromatography B* **838**(1): 3-8.

Wandall, H. H., F. Irazoqui, et al. (2007). "The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation." *Glycobiology* **17**(4): 374-87.

Wang, L. W., M. Dlugosz, et al. (2007). "O-fucosylation of thrombospondin type 1 repeats in ADAMTS-like-1/punctin-1 regulates secretion: implications for the ADAMTS superfamily." *J Biol Chem* **282**(23): 17024-31.

Wang, X., J. Gu, et al. (2006). "Core fucosylation regulates epidermal growth factor receptor-mediated intracellular signaling." *J Biol Chem* **281**(5): 2572-7.

Wang, Y., L. Shao, et al. (2001). "Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase." *J Biol Chem* **276**(43): 40338-45.

Wang, Y. and M. W. Spellman (1998). "Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells." *J Biol Chem* **273**(14): 8112-8.

Wang, Y., J. Tan, et al. (2001). "Modeling human congenital disorder of glycosylation type IIa in the mouse: conservation of asparagine-linked glycan-dependent functions in mammalian physiology and insights into disease pathogenesis." *Glycobiology* **11**(12): 1051-70.

Wells, L., K. Vosseller, et al. (2003). "A role for N-acetylglucosamine as a nutrient sensor and mediator of insulin resistance." *Cell Mol Life Sci* **60**(2): 222-8.

Wopereis, S., D. J. Lefeber, et al. (2006). "Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review." *Clin Chem* **52**(4): 574-600.

Wu, Y., Q. Li, et al. (2007). "Detecting protein-protein interactions by Far western blotting." *Nat Protoc* **2**(12): 3278-84.

Xu, A., N. Haines, et al. (2007). "In vitro reconstitution of the modulation of Drosophila Notch-ligand binding by Fringe." *J Biol Chem* **282**(48): 35153-62.

Yan, Y. T., J. J. Liu, et al. (2002). "Dual roles of Cripto as a ligand and coreceptor in the nodal signaling pathway." *Mol Cell Biol* **22**(13): 4439-49.

Yang, L. T., J. T. Nichols, et al. (2005). "Fringe glycosyltransferases differentially modulate Notch1 proteolysis induced by Delta1 and Jagged1." *Mol Biol Cell* **16**(2): 927-42.

Yang, X., P. P. Ongusaha, et al. (2008). "Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance." *Nature* **451**(7181): 964-9.

Yeo, G. W. (2005). "Splicing regulators: targets and drugs." *Genome Biol* **6**(12): 240.

Young, G. D. and J. E. Murphy-Ullrich (2004a). "Molecular interactions that confer latency to transforming growth factor-beta." *J Biol Chem* **279**(36): 38032-9.

Young, G. D. and J. E. Murphy-Ullrich (2004b). "The tryptophan-rich motifs of the thrombospondin type 1 repeats bind VLAL motifs in the latent transforming growth factor-beta complex." *J Biol Chem* **279**(46): 47633-42.

Young, W. W., Jr., D. R. Holcomb, et al. (2003). "Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isoforms in murine tissues determined by real-time PCR: a new view of a large family." *Glycobiology* **13**(7): 549-57.

Zachara, N. E. and G. W. Hart (2004). "O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress." *Biochim Biophys Acta* **1673**(1-2): 13-28.

Zachara, N. E. and G. W. Hart (2006). "Cell signaling, the essential role of O-GlcNAc!" *Biochim Biophys Acta* **1761**(5-6): 599-617.

Zanetta, J. P., A. Pons, et al. (2004). "Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins." *Anal Biochem* **329**(2): 199-206.

Zhang, J. X., I. Braakman, et al. (1997). "Quality control in the secretory pathway: the role of calreticulin, calnexin and BiP in the retention of glycoproteins with C-terminal truncations." *Mol Biol Cell* **8**(10): 1943-54.

Zhang, N. and T. Gridley (1998). "Defects in somite formation in lunatic fringe-deficient mice." *Nature* **394**(6691): 374-7.

Zhen, L., H. Baumann, et al. (1993). "The signal for retention of the egasyn-glucuronidase complex within the endoplasmic reticulum." *Arch Biochem Biophys* **304**(2): 402-14.

Zisman, S., K. Marom, et al. (2007). "Proteolysis and membrane capture of F-spondin generates combinatorial guidance cues from a single molecule." *J Cell Biol* **178**(7): 1237-49.

10. CURRICULUM VITAE

PERSONAL DATA

Name	Krisztina Keusch (*née* Kozma)
Address	General Guisan-strasse 30, 4054 Basel
Mobile	+41 79 442 32 51
E-Mail	krisztinakeusch@gmail.com
Date of Birth	31st October 1979
Nationality	Hungarian

EDUCATION

Sept. 2003 – present **Expected graduation date** **Jun. 2008**	**PhD.** in Biochemistry, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. Identification and characterisation of a uridine diphosphoglucose: fucose-β1,3-glucosyltransferase
Sept. 1998 – Jul. 2003	**M.Sc**. in Bioengineer in Industrial Biotechnology, Faculty of Chemical Engineering, (specialised in Industrial Biotechnology), Budapest University of Technology and Economics, Budapest, Hungary. The utilisation of industrial by-products and wastes as substrates for cellulase enzyme fermentation. (Distinction "excellent")
Sept. 1994 – Jul. 1998	István Széchenyi Secondary Grammar School, Dunaújváros, Hungary

PRACTICAL EXPERIENCE

Jan. 2003 – Apr. 2003	**Erasmus Internship, Lund, Sweden.** University of Lund, Division of Biotechnology. Cloning and expression of a putative maltogenic amylase in *Saccharomyces cerevisiae*
Aug. 2002 – Sept. 2002	**International Association for the Exchange of Students for Technical Experience (IAESTE) traineeship, Lija, Malta.** Ministry for Agriculture and Fisheries**,** Plant Biotechnology Centre. Propagation of plants by tissue culture

TEACHING

Nov. 2005-Feb 2006	University of Basel, Switzerland. Tutorials for first year biology students

FURTHER TRAINING

Apr. 2007-Apr. 2008	University of Basel and Novartis, Switzerland. Participation in the Women Into Industry (WIN) program
Sept. 2003	University of Basel, Switzerland. EMBnet Course: Introduction to Bioinformatics

EVENT ORGANISATION

2004	Member of the International Association for the Exchange of Students for Technical Experience (IAESTE), Basel, Switzerland. Responsible for organising events and activities
May 2005	Member of the organising committee of the "Career Guidance Conference in Life Sciences" under the auspices of Novartis, Basel, Switzerland. Publicity team
Dec. 2005	Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. Graphical assistance in developing a student homepage

LABORATORY SKILLS

Cell biology	Tissue culture; transfections; generating stable cell lines; protein expression and purification; siRNA; FACS
Molecular Biology	Primer design; PCR; cloning,
Biochemistry	ELISA; Immunoprecipitation; SDS-PAGE; Western blot; radiochemical assays
Chromatography	Ion-exchange; gel-filtration; affinity purification; FPLC (ÄKTA); HPLC (Agilent 1100)
Mass Spectrometry	LC-MS, MALDI-TOF-TOF

OTHER SKILLS

Languages	Hungarian (native speaker), English (fluent written and spoken), German (basic knowledge)
IT Literacy	Apple, Microsoft Windows, MS Office, VectorNTI, GPMAW, WinMDI, Photo and video editing, Internet Explorer

HOBBIES

Sport	Swimming (Member of a professional Hungarian water polo team from 1994 to 1998 and WSV Basel water polo team in 2005); aerobics; hiking
Music	Played the harp from 1992 to 1997
Others	Photography; making videos; handicrafts; cooking

REFERENCES

PD. Dr. Jan Hofsteenge
Friedrich Miescher Institute for Biomedical Research,
CH-4058, Basel, Maulbeerstrasse 66,
Phone:+41 61 697 4722
E-mail: jan.hofsteenge@fmi.ch

Dr. Kati Réczey, Associate Professor
BUTE, Dept. of Agricultural Chemical Technology,
H-1521 Budapest, Szt. Gellért tér 4.
Phone: +36 1 463 2843
E-mail: kati_reczey@mkt.bme.hu

PUBLICATIONS

2008 Canevascini, S., Klein, D., Grob, M., Keusch, K., Hess, D., Althaus, J., Chiquet-Ehrismann, R. and Hofsteenge, J.
The O-fucosyltransferase PAD-2/POFUT2 regulates distal tip cell migration in *C. elegans*, manuscript in preparation

Dec.2006 Kozma, K., Keusch, J. J., Hegemann, B., Luther, K. B., Klein, D., Hess, D., Haltiwanger, R. S. and Hofsteenge J.
Identification and Characterization of a β-1,3-Glucosyltransferase That Synthesizes the Glc-β1,3-Fuc Disaccharide on Thrombospondin Type 1 Repeats, *J. Biol. Chem*., Dec 2006; 281: 36742 – 36751

CONFERENCES ATTENDED AND ORAL PRESENTATIONS

18th Nov. 2007 **Initiating a collaboration** with the Center for Human and Clinical Genetics, Leiden University Medical Center, The Netherlands.
*β*3Glc-T=B3GALTL-*oral presentation*

25-28th Jun. 2006 **GlycoT2006 5th International Symposium on Glycosyltransferase** Tsukuba, Japan.
Identification of a UDP-Glucose glucosyltransferase that modifies fucosylated thrombospondin type 1 repeats-*oral and poster presentation*

15th Apr. 2003 **National Students' Research Conference**, Eötvös Lóránt University, Biochemistry and Biotechnology, Budapest, Hungary.
The utilisation of industrial by-products and wastes as substrates for cellulase enzyme fermentation-*oral presentation*

12th Nov. 2002 **Students' Research Conference at Budapest University of Technology and Economics, 2nd Prize,** Biochemistry and Biotechnology**,** Budapest, Hungary.
The utilisation of industrial by-products and wastes as substrates for cellulase enzyme fermentation-*oral presentation*

3-5th Jul. 2002 **8th International Students' Research Conference on Environmental Protection**, Mezőtúr, Hungary.
Cellulase enzyme fermentation utilising old corrugated cardboard as a fermentation substrate-*oral presentation*