Structure of Human POFUT2: Insights into Thrombospondin Type 1 Repeat Fold and *O*-Fucosylation

Inauguraldissertation

Zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Chun-I Chen

aus Taipei, Taiwan

Basel 2012

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Jan Hofsteenge

Prof. Dr. Thierry Hennet

Dr. Heinz Gut

Dr. Nicolas Thoma

Basel, den 24.04.2012

Prof. Dr. Martin Spiess

Dekan

Table of Contents

Table of Contents

1. Acknowledgements

First and foremost, I would like to thank Dr. Jan Hofsteenge and Dr. Heinz Gut as my supervisors. Jan gave me the opportunity to work on this fascinating project and supported me whenever I had a problem in science or in life. Heinz offered me the great working environment at Protein Structure Facility and mentored me through my PhD. I really appreciate that when I made mistakes or the project was not working as it should, I received only encouragement and support that kept me moving and looking at the bright side. I am really thankful to both of you.

I also want to acknowledge the thesis committee members, Dr. Nicolas Thoma and Dr. Thierry Hennet for their advice, encouragement and time for thesis committee meetings.

A big thank to Dr. Jeremy Keusch, who helped me so much in my project from discussing the initial ideas, sharing the knowledge and enthusiasm of science to guiding me the scientific skills. Without your support and help I would never have learned as much about science as I have.

I also would like to thank the former members of the Hofsteenge lab and Protein Analysis Facility: Stefano Canevascini, Constanze Heinrich, Daniel Hess, Claudia Keller, Carsten Krantz, Dominique Klein, Krisztina Keusch, Reto Portmann, Matthias Scharenberg and Ragna Sack. Dominique was always so kind, helpful and friendly. I will always remember that my first batch of POFUT2 was purified with you and I really love the cookies from you. I want to thank Daniel and Ragna for helping so much on my project and offering me the great chance to learn and use mass spectrometry.

Finally, I want to thank my family. I am really grateful to my sister for supporting the family members and me during my study. Thanks particularly to my parents for encouraging me to pursue my dream although this dream took me away from you for five years on the other side of earth. Most importantly, I want to thank my wife, Fan-Yi. Your love and support that cheered me up in these years especially when I was recovering from the accident.

2. Abbreviation

ADAM	a disintegrin and metalloproteinase	LLO	Lipid-linked oligosaccharide
ADAMTS	A disintegrin and metalloproteinase with thrombospondin type 1 repeats	Kcat	Turnover number
BSA	Bovine serum albumin	LC	Liquid chromatography
CAZy	Carbohydrate-Active enZYmes database	m/z	Mass to charge ratio
CDG	Congenital disorders of glycosylation	mAb	Monoclonal antibody
Cer	Ceramide	MALDI	Matrix associated laser desorption ionization
DMEM	Dulbecco's modified eagle's medium	MPI	Phosphomannose-isomerase
ECL	Enhanced chemiluminescence	MRM	Multiple reaction monitoring
ECM	Extracellular matrix	MS	Mass spectrometry
EDTA	Ethylenediaminetetraacetic acid	OGT	*O*-GlcNAc transferase
EGF	Epidermal growth factor	OST	Oligosaccharyltransferase
EndoHf	Endoglycosidase Hf	PAGE	Polyacrylamide gel electrophoresis
ER	Endoplasmic reticulum	PBS	Phosphate buffered saline
ERAD	ER-associated degradation	pI	Isoelectric point
ESI	Electron spray ionization	PIP	Phosphatidylinositolphosphate
FCS	Fetal calf serum	PNGase F	Peptide *N*-glycosidase F
FPLC	Fast protein liquid chromatography	POFUT1	Protein *O*-fucosyltransferase 1
FT	Fourier transformation	POFUT2	Protein *O*-fucosyltransferase 2
GAG	Glycosaminoglycan	PPS	Peters Plus syndrome
GDP	Guanosine diphosphate	RT	Room temperature
GHs	Glycosidases	SDS	Sodium dodecyl sulfate
GnT	GlcNAc transferase	SEC	Size-Exclusion Chromatography
GPI	Glycosylphosphatidylinositol	Tf	Transferrin
GTs	Glycosyltransferases	TIC	Total ion counts
HEK293T	Human embryonic kidney cells with SV40 large T antigen	TOF	Time of flight
HM	High-mannose type	TSR4	4th TSR from F-spondin
HPLC	High performance liquid chromatography	TSP-1	Thrombospondin-1
IEF	Isoelectric focusing	TRAP	Thrombospondin-related anonymous protein
IMAC	Immobilized metal affinity chromatography	UDP	Uridine diphosphate
IP	Immunoprecipitation	VEGF	Vascular endothelial growth factor

Abbreviation of amino acids

Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
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

Symbolic representation and abbreviation of sugars

●	Glucose	Glc
■	*N*-acetylglucosamine	GlcNAc
●	Galactose	Gal
■	*N*-acetylgalactosamine	GalNAc
●	Mannose	Man
▲	Fucose	Fuc
◆	*N*-acetylneuraminic acid	Neu5Ac
★	Xylose	Xyl
◆	Glucuronic acid	GlcA
◇	Iduronic acid	IdoA

3. Abstract

Protein *O*-fucosylation is a post-translational modification found on epidermal growth factor-like (EGF) repeats and thrombospondin type 1 repeats (TSR). The EGF modification was shown to regulate embryonic development by controlling the ligand specificity of Notch (Chapter 5.1) but the role of protein *O*-fucosyltransferase 2 (POFUT2) -dependent fucosylation of TSRs is less understood. Nevertheless, recent work on *Pofut2* knockout mice found that *O*-fucosylation of TSRs is essential for restricting epithelial to mesenchymal transition, correct patterning of mesoderm and localization of endoderm, suggesting a critical role of POFUT2 in the early development of the mouse embryo (Du et al, 2010). More than 50 human proteins, located either on the membrane or being secreted, contain the TSR consensus sequence for POFUT2-dependent fucosylation. Together, these proteins regulate a wide variety of biological processes such as anti-angiogenesis, growth factor release or innate immunity response (Chapter 5.4).

In order to better understand *O*-fucosylation on TSRs we carried out a structural and functional analysis of human POFUT2. Crystal structures of native POFUT2 reveals that the protein folds into GT-B superfamily of glycosyltransferase (Chapter 5.3 and Chapter 6.2). HPLC mass-spectrometry enzyme activity measurements of wild-type POFUT2 show that the full activity of POFUT2 is divalent metal ion dependent (Chapter 6.2) and that K_M value of wild-type POFUT2 of GDP-fucose and TSR4 are 10 μM and 30 μM, respectively. Based on the structure of POFUT2 and activity assays of POFUT2 mutants, we proposed that E54 is the catalytic base. The structure of GDP-fucose in complex with E54A-POFUT2 shows that GDP-fucose locates in a conserved minor cleft on the C-terminal domain. Mutations of conserved residues involved in GDP-fucose interaction abolish POFUT2 activity (Chapter 6.2). The *in vivo* fucosylation assays of TSR4 mutants (Chapter 6.3) and *in vitro* fucosylation assays of mini-TSR provide a model of how POFUT2 recognizes a properly folded TSR and transfers fucose on it (Chapter 6.1 and 6.3).

Our experimental findings give insight into the catalytic mechanism and the special structural requirements of POFUT2 modifying a completely folded protein sugar acceptor that transiently forms a protein-protein interaction.

4. Research Objective

In order to better understand the mechanism of *O*-fucosylation on TSRs, we aimed at characterizing the catalytic mechanism and substrate specificity of POFUT2 by solving the structure of POFUT2 and performing enzyme kinetic assays. To achieve the goals we divided the work into different steps:

1. Expression and purification of properly folded TSR4 for the POFUT2 activity assay (Chapter 6.1)
2. Expression, purification and crystallization of wild-type POFUT2 (Chapter 6.2)
3. Expression and characterization of POFUT2 mutants (Chapter 6.2)
4. Crystallization of the GDP-fucose-E54A-POFUT2 binary complex (Chapter 6.2)
5. *In vitro* and *in vivo* fucosylation assays of TSR mutants and mini-TSR (Chapter 6.1 and 6.2)

5. General Introduction

5.1 Protein Glycosylation

Glycan is one of the most common molecules in all cell types. In the bacterial cell wall, glycan polymerizes with amino acids to form peptidoglycan that forms a mesh-like layer outside the plasma membrane (Typas et al, 2012). In plants, the energy produced by photosynthesis is stored as starch, which is a polymer of glucose. Glycans can exist alone or attached to lipids, proteins, and DNA to create a variety of glycoconjugates. The most common way to classify glycans in eukaryotic cells is according to the nature of the linkage to the aglycone (Varki & Sharon, 2009). For example, lipids and proteins modified with glycans are called glycosphingolipids and glycoproteins, respectively (Fig. 5.1.1). Attachment of glycans to glycoproteins can be *via* N- and O-linkages; therefore, the glycans on proteins can be further divided into N-glycans and O-glycans. Other types of glycans also exist in cells (Fig. 5.1.1). This chapter will provide only a brief overview of glycoproteins in animal cells and will focus on the structure and function of unusual glycosylations of the epidermal growth factor (EGF) repeat and thrombospondin repeat (TSR).
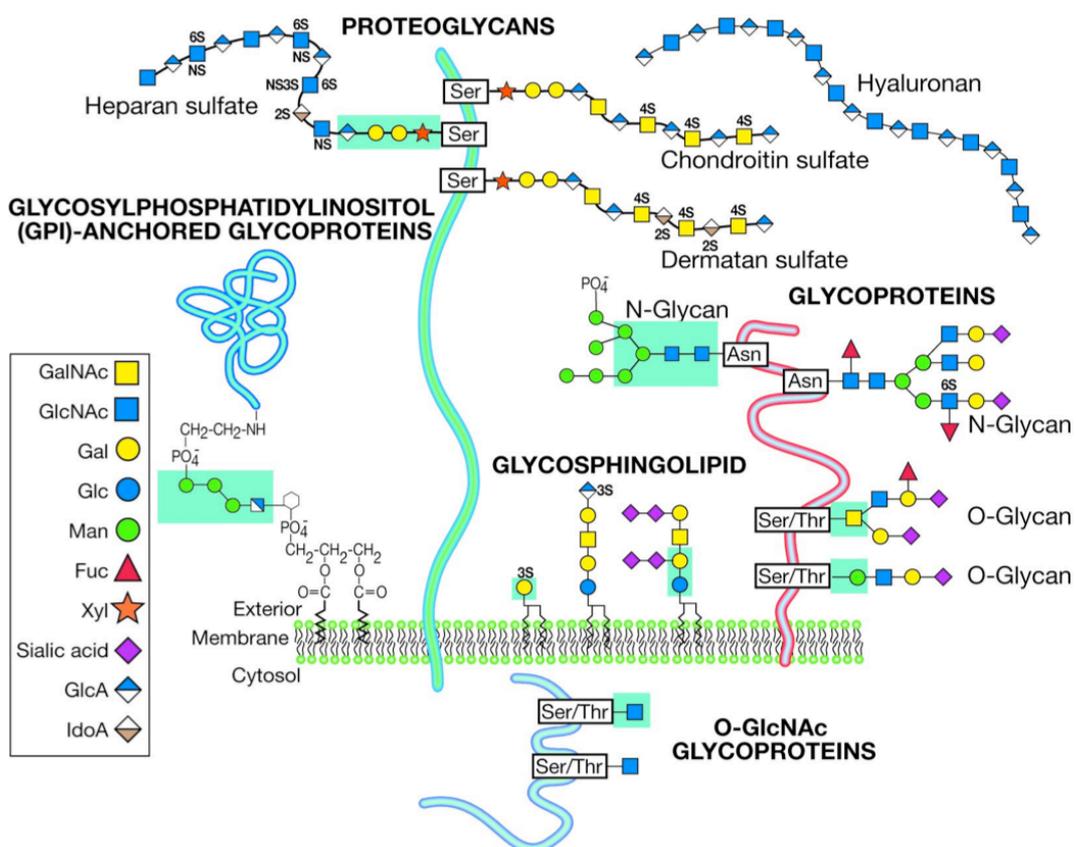
Figure 5.1.1 Common classes of animal glycans

Glycans can be found on glycoproteins with *N*- or *O*-linkages on Asn and Ser/Thr, respectively. Different types of glycans can be covalently linked to lipids to form glycosphingolipids. GPI-anchored glycoproteins are a group of proteins anchored to the cell surface by lipids *via* four monosaccharides. Proteoglycans consist of a protein core and the covalently attached glycosaminoglycan chains on Ser. Unlike other glycoconjugates, *O*-GlcNAc glycoproteins appear in the cytosol and nucleus. The core structures of different types of glycosylation are highlighted with a green background.

N-Glycans

An *N*-linked glycan consists of an oligosaccharide covalently attached to an asparagine residue with the consensus sequence Asn-X-Ser/Thr. All *N*-glycans harbor the same pentasaccharide core structure: three mannose groups on GlcNAc-GlcNAc (highlighted in green background in Fig. 5.1.1). High-mannose-type *N*-glycans have an additional two to six mannose groups on the pentasaccharide core. In the complex-type glycans, two mannose groups on the core are modified by GlcNAc, and different glycans can link to this GlcNAc. In the hybrid-type *N*-glycans, the mannose groups are attached on the Manα1-6 arm of the core, and are extended with GlcNAc on the Manα1-3 arm (Fig. 5.1.2).
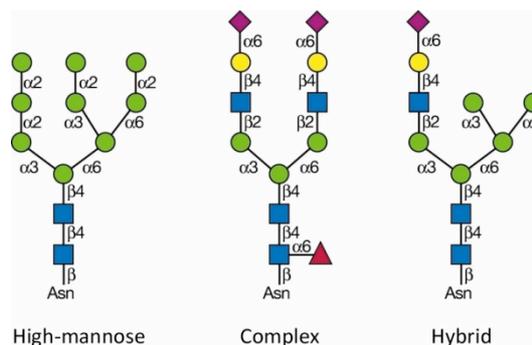
Figure 5.1.2 Types of *N*-glycans

N-glycans added to the amide group of Asn of the conserved Asn-X-Ser/Thr motif are classified into three different types, which share the same Man$_3$GlcNAc$_2$ core structure.

The transfer of *N*-glycans takes place in the endoplasmic reticulum (ER) by oligosaccharyltransferase (OST), which is a protein complex with nine subunits in yeast, that transfers the preformed dolichol-P-P-GlcNAc$_2$Man$_9$GlcNAc$_2$ to nascent proteins with the conserved sequon (Schwarz & Aebi, 2011). From the ER to the Golgi apparatus, many glycosyltransferases and glycosidases may be involved in processing the *N*-glycan structure on the protein. The process of adding/removing glycans is not template-driven; therefore, the same protein expressed in different cell types or even the same cell type but in different physical conditions may have a

different *N*-glycan structure, which depends on many factors such as the nutritional condition and the types of glycosyltransferases and glycosidases in the cell's secretory pathway.

The functions of *N*-glycans include quality control of proteins, maintenance of half-life of proteins in the cell, and regulation of cell-cell and cell-matrix interactions (Lowe & Marth, 2003; Ruddock & Molinari, 2006). In humans, a growing number of diseases with impairments in the *N*-glycosylation pathway have been identified, and the importance of a group of diseases called congenital disorders of glycosylation (CDG) is emerging (Chapter 5.3).

O-GalNAc modification

O-glycosylation is a common protein modification on the hydroxyl side chain of serine and threonine. The most abundant type of protein *O*-glycosylation is *O*-linked *N*-acetylgalactosamine (GalNAc) *via* an α-linked *O*-glycosidic bond to proteins. Many mucins (a type of glycoprotein in mucus) are highly glycosylated with *O*-linked GalNAc; therefore, *O*-GalNAc glycans are also called mucin *O*-glycans. The synthesis of mucin type *O*-glycans is initiated by transferring GalNAc from UDP-GalNAc to Ser/Thr on proteins catalyzed by the enzyme *N*-acetylgalactosaminyltransferase (ppGalNAcT). The α-linked GalNAc on the protein is called the Tn antigen, and all mucin *O*-glycans have this core structure (Table 5.1.1). PpGalNAcT is a type II membrane protein with a short transmembrane domain and a catalytic domain facing the lumen of the Golgi (Fritz et al, 2006). The consensus sequence on the protein for recognition by ppGalNAcT has not been determined, but it is known that a proline next to the Ser/Thr may increase the occupancy of *O*-GalNAc, which may be because proline can cause a β-turn on the protein to expose the hydroxyl group of Ser/Thr to ppGalNAcT.

Table 5.1.1 Core structures of mucin *O*-glycans

O-Glycan	Structure
Core	
Tn antigen	GalNAcαSer/Thr
Sialyl-Tn antigen	Siaα2-6GalNAcαSer/Thr
Core 1 or T antigen	Galβ1-3GalNAcαSer/Thr
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr
Core 3	GlcNAcβ1-3GalNAcαSer/Thr
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr
Core 5	GalNAcα1-3GalNAcαSer/Thr
Core 6	GlcNAcβ1-6GalNAcαSer/Thr
Core 7	GalNAcα1-6GalNAcαSer/Thr
Core 8	Galα1-3GalNAcαSer/Thr

Adapted from (Brockhausen et al, 2009)

The GlcNAc of the Tn antigen can be extended at 3- and 6-hydroxul group, which results in different core structures (Table 5.1.1). Synthesis of the core 1 structure is catalyzed by core 1 β1,3 galactosyltransferase (C1GalT-1), which transfers a galactose to the 3-position of GalNAc and produces core 1 Galβ1-3GalNAc dissacharide (Fig. 5.1.3). The core 1 structure can be branched at the 6-position of GalNAc by a GlcNAc residue, and the trisaccharide product is the core 2 structure. This reaction is catalyzed by C2 β1-6 *N*-acetylglucosaminyltransferase (C2GnT).
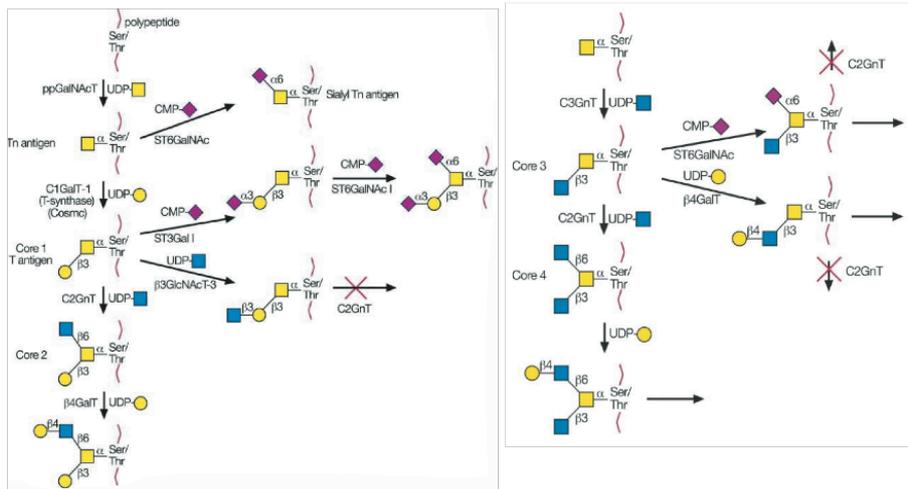
Adapted from (Brockhausen et al, 2009)

Figure 5.1.3 Biosynthesis of core 1 to core 4 *O*-GalNAc structures

The core structures of mucin *O*-glycans are synthesized by different glycosyltransferases. The final structure of the *O*-glycan may be determined by the activities of different glycosyltransferases. For example, if the activity of sialyltransferase is high, the chains will normally be short and terminated with sialic acid.

The core 3 structure is synthesized by elongation of the Tn antigen with GlcNAc on the 3-position. An additional GlcNAc linked to the 6-position GalNAc on the core 3 structure produces the core 4 structure (Fig. 5.1.3). The glycosyltransferases responsible for synthesizing the core 3 and core 4 structures are β1,3 and β1,6 *N*-acetylglucosaminyltransferase, respectively. Core 5 to Core 8 structures are more tissue specific, and the enzymes that synthesize these structures remain to be characterized.

Different glycan epitopes such as ABO blood group antigens and Lewis-type antigens can be linked to different core structures; fucose, sialic acid, and sulfate that are common on other glycoconjugates can also be found on mucin *O*-glycans. Overall, the different core structures and diverse glycan structures that can extend from the core structure make mucin *O*-glycans complicated molecules for analysis.

O-GalNAc glycans on mucins are important for the function of mucins. These glycans are usually hydrophilic with a negatively charged sialic acid that promotes the binding of water and salt on mucins and contributes to the viscosity and adhesiveness of mucus. These properties provide the physical barrier between the lumen and epithelium. *O*-GalNAc glycans are also involved in the interaction between leukocytes and endothelial cells; hence, they regulate leukocyte homing and inflammation. The interaction is mediated by the selectins expressed on the endothelial cell and sialyl Lewis[X] antigen on core 2 *O*-glycans. Mice lacking C2GnT-1 activity, the key enzyme for core 2 *O*-glycan synthesis, have severe deficiencies in their immune system, such as reducing neutrophil rolling on substrates bearing E-, L-, and P-selectins (Ellies et al, 1998). Changes in the biosynthesis of glycan structures have been observed in cancers (Chapter 5.3). A recent study showed that increased expression of C2GnT-1 and the resulting up-regulation of the sialyl Lewis[X] epitope on core 2 O-glycans mediated the adhesive and invasive behaviors of human carcinomas, which may influence their metastatic potential (St. Hill et al, 2011).

O-GlcNAc modification

O-GlcNAc is a unique type of protein glycosylation compared to *N*-glycans and mucin *O*-glycans. (1) *O*-GlcNAc exists as a single sugar modification, and is not extended by other glycans. (2) The *O*-GlcNAc modification is reversible; the *O*-GlcNAc groups can be attached and removed multiple times from proteins. (3) The "classic" *O*-GlcNAc reaction occurs in the nuclear and cytoplasmic compartments, while other glycosylations occur in the ER or Golgi in secretory pathways. Although a rare *O*-linked GlcNAc has been found on the EGF repeat, it is catalyzed by a distinct glycosyltransferase that will be discussed later.

The enzyme that catalyzes *O*-GlcNAcylation is *O*-GlcNAc transferase (OGT), which is a cytosolic and nuclear protein with two distinct regions: the tetratricopeptide repeat and a multidomain catalytic region (Lazarus et al, 2011). Unlike other glycosylations, which are generally static, the cycling of *O*-GlcNAcylation has been observed, which is coordinated by a cytosolic and nuclear β-*N*-acetylglucosaminidase (*O*-GlcNAcase) (Fig. 5.1.4). The cycling of *O*-GlcNAc on regulatory proteins responds rapidly to metabolic and environmental signals at rates comparable to phosphorylation (Hart et al, 2007).
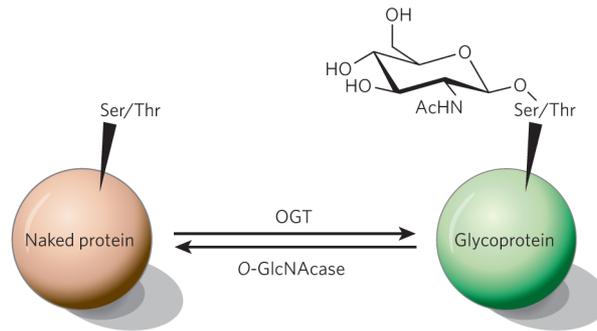
Figure 5.1.4 Cycling of *O*-linked GlcNAc

O-GlcNAcylation on serine or threonine is controlled by two highly conserved enzymes: OGT and *O*-GlcNAcase. The cycling rate of *O*-GlcNAcylation is similar to that of *O*-phosphorylation in response to various stimulations.

The activity of OGT is controlled by the concentration of its sugar-donor UDP-GlcNAc, which is sensitive to metabolic molecules such as glucose, fatty acids, and nucleotides, making it an ideal sensor for nutritional status (Hart & Akimoto, 2009). In the cell, approximately 2%–5% of intracellular glucose enters the hexosamine biosynthesis pathway and produces GlcNAc. Hyperglycemic cells have a higher concentration of GlcNAc and UDP-GlcNAc, and therefore up-regulate the activity of OGT. Thus, the extent of protein *O*-GlcNAcylation is often considered to be sensitive to nutrient availability.

The importance of *O*-GlcNAc is partially contributed by the fact that *O*-GlcNAc is ubiquitous in proteins belonging to different functional classes, ranging from proteins involved in transcription/translation to protein metabolism in response to stress and signaling (Hart & Akimoto, 2009). Furthermore, *O*-GlcNAc can interplay with *O*-phosphorylation (Fig.5.1.5). *O*-GlcNAc and *O*-phosphate can compete for the same site, and the site occupied by one modification excludes the other modification or there can be two adjacent available modification sites, and the *O*-linked GlcNAc is required for *O*-linked phosphorylation on the adjacent site (Hart et al, 2007).
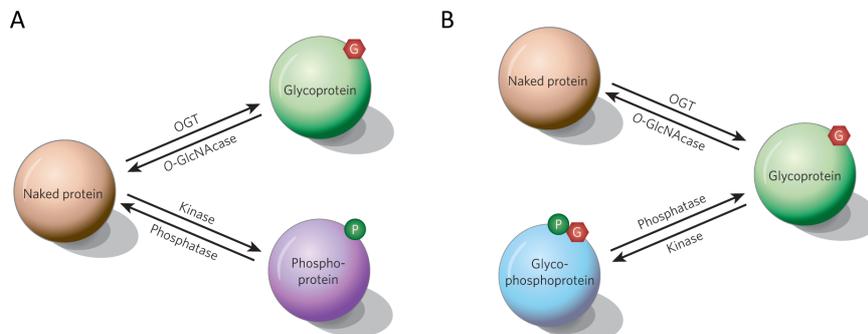
Figure 5.1.5 Dynamic interplay between *O*-GlcNAc and *O*-phosphate

(A) In some proteins, *O*-GlcNAc and *O*-phosphate compete dynamically for the same modification site. (B) In other proteins, *O*-GlcNAc and *O*-phosphate can occur next to each other. The presence of *O*-GlcNAc can increase the occupancy of the phosphate site.

The ubiquitous existence of *O*-linked GlcNAc on proteins in various functional classes may reflect the diverse function of *O*-GlcNAc glycans. Similar to protein phosphorylation, the specific function of *O*-GlcNAc is related to the *O*-GlcNacylated protein. For example, if *O*-GlcNAc is on the transcription factor Sp1, the modification will interfere with the binding of Sp1 to other transcription factors, and thus, it will down-regulate gene expression controlled by Sp1 (Lim & Chang, 2009). OGT can also sense the nutritional status and respond by up- or down-regulating its activity accordingly. This regulation is mediated by the concentration of UDP-GlcNAc. In cells with high concentration of glucose, the concentration of UDP-GlcNAc is also high, which increases the activity of OGT. Studies have shown that increased *O*-GlcNAc on proteins mimics hyperglycemia, which caused insulin resistance in a mouse model, and mice overexpressing OGT displayed features of type II diabetes (McClain et al, 2002; Vosseller et al, 2002b). These results indicate that *O*-GlcNAc is increased in response to high glucose, and elevated levels of *O*-GlcNAc on proteins inhibits insulin signaling (Vosseller et al, 2002a).

Glycosylation of epidermal growth factor repeat

Notch receptors are transmembrane proteins of around 300 kDa that play an important role in developmental pathways (Rana & Haltiwanger, 2011; Stanley, 2007). The dysregulation of Notch signaling pathways can cause developmental diseases and some types of cancer (Bolos et al, 2007). *Drosophila* possess only one Notch receptor, while there are four Notch receptors in mammals (Kopan & Ilagan, 2009).

The extracellular domain of Notch receptors consists primarily of 26–36 EGF-like repeats; some of these EGF repeats mediate interactions with Notch ligands. EGF repeats are small protein motifs with six conserved cysteines that form three disulfide bonds, which contribute to the unique three-strand three-dimensional structure of the small EGF protein (Hiruma-Shimizu et al, 2010). On the EGF repeat of Notch proteins, in addition to the conventional *N*-glycans, three other rare *O*-linked glycans have been observed: *O*-fucose, *O*-glucose, and *O*-GlcNAc (Matsuura et al, 2008; Moloney et al, 2000b). Recent studies of these rare *O*-linked glycosylations have shown the importance of glycosylation on regulating Notch signaling.

11

In the Notch signaling pathway in Drosophila, Delta which is the Notch ligand on the signal-sending cell binds to the extracellular domain of Notch receptors (NECD) on signal-receiving cells. The receptor-ligand binding activates three sequential proteolytic cleavages. The first cleavage is at the S2 site (Figs. 5.1.6 and 5.1.7) catalyzed by the ADAM10 metalloprotease. This cleavage generates a membrane-bound Notch extracellular truncation (NEXT) and releases NECD (Fig. 5.1.6). Sequential cleavages then occur at the S3 and S4 sites catalyzed by the γ-secretase complex, which releases the Notch intracellular domain (NICD). The NICD is translocated to the nucleus, and then complexes with the CSL family of transcription factor (CBF1 in vertebrates, Su in flies and Lag1 in worms), recruits the co-activator Mastermind (Maml), and activates Notch target genes, including transcriptional regulators such as Hey genes (Jarriault et al, 1995).
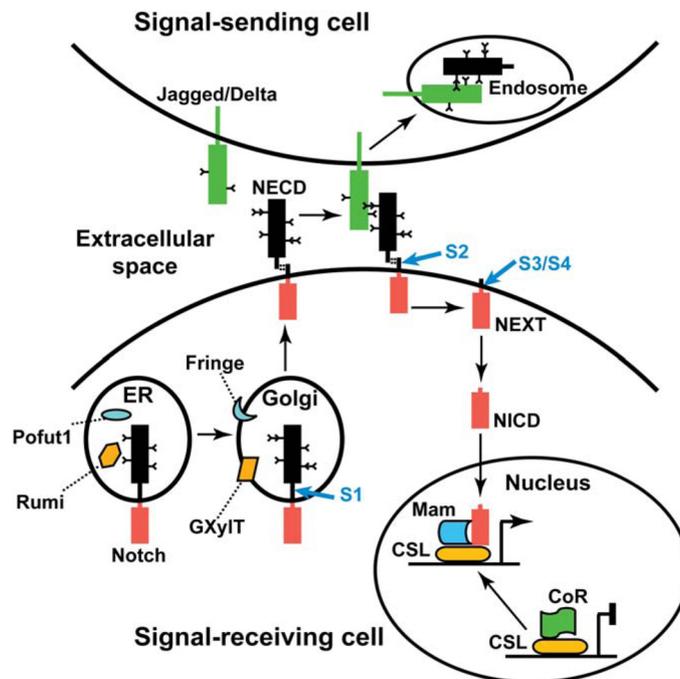
Adapted from (Jafar-Nejad et al, 2010)

Figure 5.1.6 Notch signaling pathway

Notch synthesis begins in the ER where it is glycosylated by Pofut1 and Rumi, which generate *O*-linked fucose and glucose, respectively. In the Golgi, GlcNAc and xylose are added to fucose and glucose by Fringe and GXYLT, respectively. The S1 cleavage generates a heterodimeric form of Notch, which is composed of the extracellular (NECD, black box) and intracellular (NICD, red box) domains of Notch. Binding of ligands to the NECD results in the S2 cleavage of Notch. Notch is further cleaved at the S3/S4 sites. NICD is then translocated into the nucleus, and activates target genes with other components.

O-Fucosylation on the EGF repeat was first described on pro-urokinase (Kentzer et al, 1990), and this modification has since been found on the other proteins with EGF repeats such as the Notch receptor and coagulation factors with the consensus

sequence $C^2XXXX(\underline{S/T})C^3$ (Haines & Irvine, 2003; Moloney et al, 2000b). EGF repeats with this consensus sequence are substrates for the protein *O*-fucosyltransferase 1 (POFUT1). POFUT1 is a soluble ER-resident enzyme that recognizes properly folded EGF repeats and transfers fucose from GDP-fucose to the consensus site (Luo & Haltiwanger, 2005). The importance of *Pofut1* was revealed by mice lacking POFUT1 activity; *Pofut1$^{-/-}$* mice died at midgestation (E10) with severe defects in somitogenesis, vasculogenesis, cardiogenesis, and neurogenesis, similar to mice with blocked Notch signaling pathways due to global inactivation of Notch receptor signaling (Shi & Stanley, 2003). Down-regulation of OFUT1 (the POFUT1 homolog in *Drosophila*) by RNAi demonstrated that OFUT1 regulates Notch signaling *via* Fringe-independent and Fringe-dependent pathways (Okajima & Irvine, 2002). In the Fringe-independent pathway, OFUT1 acts as a molecular chaperone and promotes the folding and secretion of Notch to the cell surface. Notch accumulated in the ER in OFUT1 mutant cells and the phenotype can be rescued by an enzymatically inactive R254A OFUT1. These data suggest that the enzyme activity is indispensible (Okajima et al, 2005). However, the chaperone-like activity of OFUT1 that aids in the secretion of the Notch receptor to the cell surface is limited to *Drosophila*. In mammalian cell systems, murine *Pofut1$^{-/-}$* cells showed the same expression level of the Notch receptor compared to wild-type cells, although deficient Notch signaling in *Pofut1$^{-/-}$* cells was observed (Stahl et al, 2008). These data suggest that differences exist between OFUT1 in *Drosophila* and POFUT1 in mammals with respect to the regulation of Notch receptor secretion.
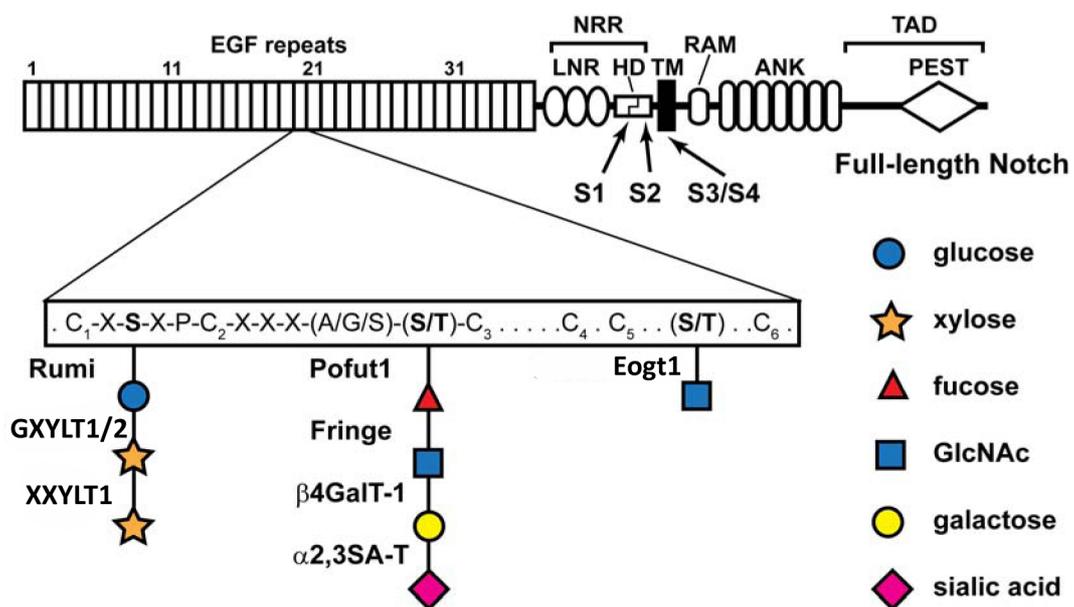

Modified from (Jafar-Nejad et al, 2010)

Figure 5.1.7 *O*-Glycosylation on EGF repeats

A schematic representation of full-length *Drosophila* Notch is shown. Thirty-six EGF repeats followed by the negative regulatory region (NRP) constitute the extracellular domain. Four Notch cleavage sites (S1–S4) are indicated with arrows before the trans-membrane (TM) domain. The intracellular domain contains an RBP-J association module (RAM), two nuclear localization signals (NLS), seven Ankyrin repeats (ANK), and a PEST domain. The consensus sequences for *O*-linked glucose, fucose, and GlcNAc on EGF repeats are highlighted in boxes. Glycosyltransferases involved in the glycosylation of the three *O*-glycosylation pathways are given next to the monosaccharides represented in cartoons.

Interestingly, Fringe was known to play an important role in the regulation of dorsal-ventral boundary formation in *Drosophila* wings, presumably *via* the Notch signaling pathway, before it was identified as a glycosyltransferase (Cohen et al, 1997; Irvine & Wieschaus, 1994; Panin et al, 1997). Later, Fringe was shown to be the glycosyltransferase responsible for transferring a GlcNAc specifically to the *O*-linked fucose on the EGF repeat. More importantly, the glycosyltransferase activity of Fringe regulates Notch signaling (Bruckner et al, 2000; Moloney et al, 2000a). How does glycosylation regulate Notch signaling? Recombinant Notch expressed on the cell surface with Fringe was shown to bind significantly better to cells expressing Delta than to cells expressing Serrate. In contrast, Notch was shown to bind better to Serrate than to Delta when Notch was expressed in the absence of Fringe. These results suggest that Fringe determines the ligand specificity of Notch (Okajima et al, 2003). Another experiment using purified components in the Notch signaling pathway demonstrated that the addition of GlcNAc to *O*-fucose *in vitro* was sufficient to both enhance Notch binding to the Delta ligand and to inhibit Notch binding to the Serrate ligand (Xu et al, 2007). These observations suggest that GlcNAc-Fuc-*O* disaccharide produced by POFUT1 and Fringe works as a boundary-specific cell-signaling molecule that is responsible for dorsal-ventral cell interactions during wing development (Fig. 5.1.8).
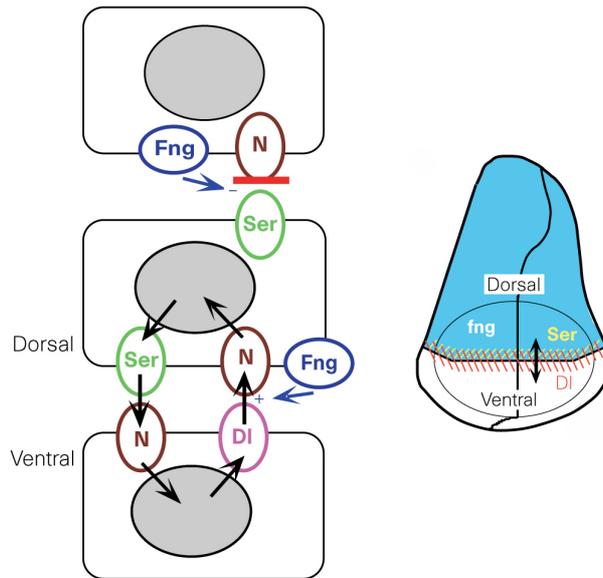
Figure 5.1.8 Notch component interactions at the dorsal–ventral compartment border

Delta (Dl in pink) and Notch (N in brown) are broadly expressed during early wing development, while Fringe (Fng in blue) and Serrate (Ser in green) are restricted to dorsal cells. For clarity, not all proteins and potential protein interactions have been included. Expression of Fng allows dorsal cells to respond to Dl, resulting in Notch activation, which leads to transcription of downstream genes, including Ser. Ser signals back from dorsal to ventral cells, activating Notch, which leads to the transcription of downstream genes, including Dl. Fng blocks the ability of Ser to signal to other dorsal cells.

In *Drosophila*, the final product of the *O*-linked fucosylation pathway is the GlcNAc-Fuc disaccharide, and this disaccharide is sufficient for binding to Delta and modulating the Notch signaling pathway (Xu et al, 2007). In mammals, *O*-fucose can be extended further to a tetrasaccharide containing Siaα2,3/6Galβ1,4GlcNAcβ1,3 Fucα1-*O*-Ser/Thr (Fig. 5.1.7) (Harris & Spellman, 1993). Elongation of the disaccharide is catalyzed by the Golgi-resident glycosyltransferases β1,4-galactosyltransferase (β4GalT-1) and α2,3/6-sialyltransferase (Fig. 5.1.7). It has been shown that in the CHO cell system, GlcNAcβ1,3Fuc disaccharide is necessary, but not sufficient, for the Notch receptor to inhibit Jagged-1-induced Notch activation. By restoring the tetrasaccharide on the EGF repeat by transfecting β4GalT-1 into the cell, the Notch receptor regained the ability to inhibit Jagged-1-induced signaling. Therefore, in mammalian cells, β4GalT-1 is required for Fringe to modulate Jagged1-induced Notch (Chen et al, 2001).

O-Glucose was first identified on the EGF repeat from bovine blood coagulation factors VII and IX (Hase et al, 1988; Hase et al, 1990), and was later found on the EGF repeat of the Notch receptor. *O*-Glucose on the EGF repeat can be extended to Xylα1,3Xylα1,3Glcβ1-*O*-Ser (Moloney et al, 2000b). In contrast to the well-

characterized O-fucosylation on the Notch signaling pathway, the importance of the O-glucose trisaccharide has emerged just recently. The glycosyltransferase responsible for transferring glucose from UDP-glucose to serine on the consensus C^1-X-$\underline{S}$-X-P-C^2 sequence on the EGF repeat was first identified in *Drosophila* and was named Rumi (Acar et al, 2008). Rumi is a soluble, ER-retained protein. Loss of Rumi activity in flies caused a temperature-sensitive Notch phenotype, since Notch could not be cleaved properly despite normal binding to Delta. Notch also accumulated inside the cell and at the cell surface in flies without Rumi activity. The authors suggested that the lack of Rumi activity did not affect Notch binding to its ligand, but prevented S2 cleavage and affected the surface expression of Notch (Acar et al, 2008; Jafar-Nejad et al, 2010). The ortholog of Rumi in humans has also been recently identified (Takeuchi et al, 2011). Interestingly, the authors observed that Rumi is not only a glucosyltransferase, but also a xylosyltransferase that can transfer xylose from UDP-xylose to proteins with almost the same efficiency as glucose. The specificity for the dual substrates of Rumi is partially affected by the X residue after the glycosylated serine. Rumi only adds xylose when the second X is serine in the consensus sequence (C^1-X-$\underline{S}$-S-P-C^2) (Takeuchi et al, 2011). This result suggests that the specificity of Rumi for donor substrates is influenced by the structure of the sugar acceptor. In a mouse model, the *Rumi$^{-/-}$* mouse embryos died before embryonic day 9.5 with posterior axis truncation and defects in neural development, somitogenesis, cardiogenesis, and vascular remodeling (Fernandez-Valdivia et al, 2011). These data indicate that O-glucosylation, like O-fucosylation on Notch, also plays a crucial role during mammalian development.

Other xylosyltransferases that extend Glc-O-Ser have been identified. Two enzymes named GXYLT1 and GXYLT2 (glucoside xylotransferase 1 and 2) are responsible for elongating Glc-O-Ser to Xyl-Glc-O-Ser (Sethi et al, 2010). The final glycosyltransferase that transfers the last xylose to the Xylα1,3-Glu disaccharide to generate the final trisaccharide product Xylα1,3Xylα1,3Glcβ1-O-Ser on the EGF repeat has recently been reported to be XXYLT1 (xyloside xylosyltransferase 1) (Sethi et al, 2012).

The last rare O-glycosylation on the EGF repeat is O-linked GlcNAc. Modification site mapping by mass spectrometry and amino acid substitution studies revealed that the O-GlcNAc modification occurs on a serine or threonine located between the fifth and sixth cysteine residues within the EGF domain (Fig. 5.1.7). Although O-GlcNAc has been found on cytoplasmic and nuclear proteins, O-GlcNAc on the EGF repeat was

the first example of the modification on an extracellular protein (Matsuura et al, 2008). The glycosyltransferase for *O*-GlcNAc on the EGF repeat is not OGT, which catalyzes the modification on intracellular proteins (Kreppel et al, 1997). The novel glycosyltransferase responsible for *O*-GlcNacylation was recently reported to be Eogt1 in *Drosophila* and mice (Sakaidani et al, 2012; Sakaidani et al, 2011). The function of *O*-GlcNAc on the EGF repeat is still unknown.

Glycosylation on the thrombospondin type 1 repeat

The TSR is a cysteine-rich motif of ~50 amino acids with three disulfide bonds (the detailed structure of the TSR is presented in Chapter 5.4). Two types of glycosylations have been demonstrated on TSRs (Fig. 5.1.9). The Glcβ1,3Fuc disaccharide was first identified from the urine of a normal individual with an *O*-linked fucose (Hallgren et al, 1975). More than 20 years later, the same disaccharide was identified on the protein from CHO cell line by size-exclusion chromatography and high pH anion-exchange chromatography (Moloney et al, 1997). Later Jan Hofsteenge and coworkers identified this disaccharide when they were analyzing another rare glycosylation, *C*-mannosylation on TSRs of thrombospondin-1 (Hofsteenge et al, 2001). Since then, the disaccharide has been reported to be present on TSRs of F-spondin, properdin, ADAMTS13, and ADAMTS-like punctin-1 (Gonzalez de Peredo et al, 2002; Ricketts et al, 2007; Wang et al, 2007). The consensus sequence $CX_{2-3}(\underline{S/T})CX_2G$ of *O*-fucosylation was determined by analyzing 14 TSR peptides from F-spondin, properdin, and TSP-1 by mass spectrometry (Gonzalez de Peredo et al, 2002). The database search with the consensus sequence identified 51 proteins with various numbers of the fucosylation sequence in mice (Table 5.1.2), which included 10 transmembrane proteins and 41 secreted proteins that regulate diverse cellular processes such as cell adhesion, differentiation, and anti-angiogenesis (Du et al, 2010; Tucker, 2004).
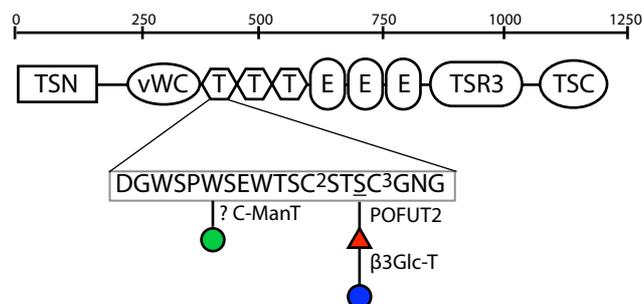
Figure 5.1.9 TSR *O*- and *C*-glycosylation on thrombospondin-1

A schematic representation of thrombospondin-1 is shown. The length of the protein corresponding to the amino acid is indicated on the top. The sequence with *C*-linked and *O*-linked glycosylation sites on TSR1 is provided in the box. The synthesis of the Glcβ1,3Fuc disaccharide is catalyzed by POFUT1 and β3Glc-T. The glycosyltransferase for *C*-mannosylation hasn't been identified. TSN, thrombospondin N-terminal domain; vWC, von Willebrand factor type A domain; T, thrombospondin type 1 repeat; E, EGF-like domain; TSR3, thrombospondin type 3 domain; TSC, thrombospondin C-terminal domain.

Both EGF repeats and TSRs are protein motifs that can be *O*-fucosylated in the ER, but these two proteins are fucosylated by different *O*-fucosyltransferases (Luo et al, 2006b). The fucosyltransferase for *O*-fucosylation of TSRs is protein *O*-fucosyltransferase 2 (POFUT2). POFUT2 is located to the ER, although no ER-retaining sequence has been identified. POFUT2 transfers fucose from GDP-fucose to properly folded TSRs (Luo et al, 2006a). While the function of *O*-fucosylation on EGF repeats has been explored extensively in the Notch signaling pathway, the molecular mechanism of *O*-fucose on TSRs remains unknown. Mutations in the fucosylation sites in ADAMTS13 and ADAMTS-like-1/punctin-1 decrease the secretion of these two proteins. Moreover, the secretion efficiency of these two proteins in Lec13 cells was restored when fucose was supplemented in the culture medium, suggesting that *O*-fucosylation catalyzed by POFUT2 is required for protein secretion. The importance of *O*-fucosylation on TSRs has recently been shown in *Pofut2$^{-/-}$* mice. The homozygous embryos died before E10.5, i.e., when gastrulation and organogenesis occur. These data demonstrated that *O*-fucosylation of TSRs is essential for restricting the epithelial-to-mesenchymal transition in the primitive streak, correct patterning of the mesoderm, and localization of the definitive endoderm (Du et al, 2010).

TSR class	Mouse protein	# Consensus sequences	Family
Group 1	ADAMTS1–4, ADAMTS5 (Wang et al., 2009), ADAMTS6–10, ADAMTS12, ADAMTS13 (Ricketts et al., 2007), ADAMTS14–20	1–12	A disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS)
	ADAMTSL1 (Wang et al., 2007, 2009), ADAMTSL2–6	1–10	ADAMTS-like
	BAI1–3	4	Brain-specific angiogenesis inhibitor
	Cfp (Properdin) (Gonzalez de Peredo et al., 2002)	4	Complement component
	CILP2	1	Cartilage intermediate layer protein/nucleotide pyrophosphohydrolase
	HMCN1	6	Hemicentin
	PAPLN (Papilin)	4	
	SEMA5A, SEMA5b	2	Semaphorin
	THBS1 (TSP1) (Hofsteenge et al., 2001; Gonzalez de Peredo et al., 2002), THBS2 (TSP2)	3	Thrombospondin
	ISM1, ISM2, XM_283765	1	Isthmin
	UNC5a	1	UNC5
Group1 and Group 2	C6	1	Complement component
	SSPO (Sco-spondin)	12	Spondin
	THSD7a, THSD7b	4–5	Thrombospondin, type I, domain-containing
Group 2	CCN1 (CYR61), CCN2 (CTGF), CCN3 (NOV), CCN4 (ELM1/WISP1), CCN5 (COP1/WISP2), CCN6 (WISP3)	1	CCN (Cyr61, Ctgf, Nov)
	SPON1 (f-spondin) (Gonzalez de Peredo et al., 2002)	4	Spondin

Adapted from (Du et al, 2010)

Table 5.1.2 Mouse proteins with a putative fucosylation site on TSRs

The mouse proteins containing putative fucosylation sites from a database search with the consensus sequence are listed in two groups according to different disulfide bond

patterns. Proteins with *O*-linked fucose that have been evaluated experimentally are underlined.

The elongation of *O*-linked fucose to the Glcβ1,3Fuc-O disaccharide is catalyzed by β1,3-glucosyltransferase (β3Glc-T) (Kozma et al, 2006; Sato et al, 2006). β3Glc-T uses UDP-glucose as the sugar donor and transfers it to the 3-position of *O*-linked fucose on the TSR. Like POFUT2, β3Glc-T is also localized to the ER compartment, but unlike POFUT2, it has a KDEL-like (REEL) retaining signal on the C-terminus. Deletion of REEL resulted in the accumulation of β3Glc-T in culture medium (unpublished data from Kozma). β3Glc-T can only glycosylate properly folded fucosylated TSR4 but not fucosylated EGF repeat although β3Glc-T share 28% sequence identity with Fringe (Kozma et al, 2006). In humans, mutations in the gene encoding β3Glc-T cause a rare developmental disease called Peters plus syndrome (details are in Chapter 5.3).

5.2 Structure and catalytic mechanism of glycosyltransferases

The diverse structures of glycans on molecules are primarily determined by the assembling and trimming of the sugars mediated by GTs and glycosidases, respectively. Glycosyltransferases are a group of enzymes that catalyze the transfer of a sugar from an activated sugar donor to a sugar acceptor. The most common activated sugar donors are the nucleotide diphosphate sugars such as GDP-fucose; other types of sugar donors, including nucleotide monophosphate sugars (e.g., CMP-Neu5Ac) or lipid phosphate sugars (e.g., Dolichol-P-mannose), are also used. Oligosaccharides, the most common sugar acceptor, as well as proteins, lipids, and nucleic acids, can act as substrates for glycosyltransferases (Fig. 5.1.1).

Glycosylation reactions

Acceptor   Glycosyltransferase   Glycosylated acceptor
\+   \+
Glycosyl donor   Nucleotide or isoprenoid-P

Glycosyl donors
CMP-sialic acid
GDP-fucose
GDP-mannose
UDP-glucose
UDP-galactose
UDP-xylose
UDP-*N*-acetylglucosamine
UDP-*N*-acetylgalactosamine
UDP-glucuronic acid
Dolichol-P-glucose
Dolichol-P-mannose
Dolichol-P-P-(glucose $_3$-mannose$_9$-GlcNAc$_2$)
Undecaprenyl-P-P-*N*-acetylmuramic acid-pentapeptide-GlcNAc

Acceptors
Oligosaccharides
Monosaccharides
Proteins
Lipids (ceramides)
DNA

Adapted from (Rini et al, 2009)

Figure 5.2.1 A variety of donors and substrates are used by glycosyltransferases

In vertebrates, glycosylation is a two-substrate reaction involving the transfer of the sugar moiety from glycosyl donors to the acceptors. The most common sugars and acceptors are listed.

Currently, more than 30,000 glycosyltransferase sequences are known. They constitute about 1% of the mammalian genome (Lowe & Marth, 2003). Based on protein sequence similarity, they have been further categorized into 94 families in the carbohydrate-active enzyme (CAZy) database (Cantarel et al, 2009).

Structural folds of glycosyltransferase

In contrast to the diverse substrate specificity of glycosyltransferases and the enormous diversity of glycan structures that can be produced, structural analysis has

shown that glycosyltransferases only adopt a limited number of folds. The structure of GT-A and GT-B folds of glycosyltransferases, which usually utilize nucleotide sugars as the donor, are shown in Fig. 5.2.2A and B, respectively. Recently, the structure of GT-C, which uses lipid sugars as the donor, was elucidated (Lizak et al, 2011), and is shown in Fig. 5.2.2C. Details of each structural fold and catalytic mechanism of glycosyltransferase are discussed below.

A. GT-A   B. GT-B   C. GT-C

—— 36 Å ——   ——— 55 Å ———   —— 50 Å ——
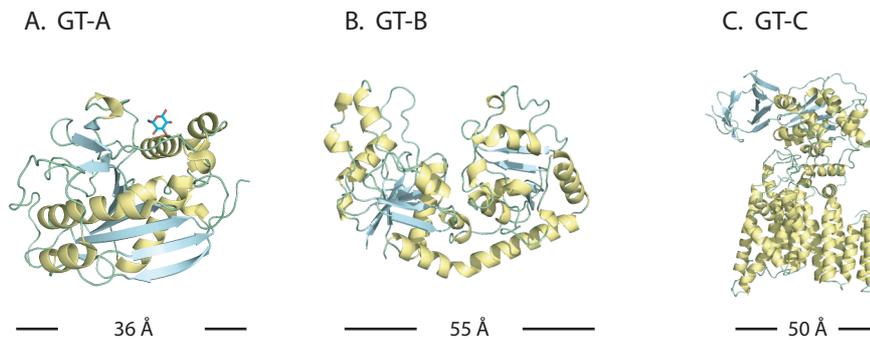
Figure 5.2.2 Structural comparison of glycosyltransferases with different folds

The GT-A fold in (A) is β4Gal-T1 from Bos taurus, PDB: 1fgx (Gastinel et al, 1999). The GT-B fold in (B) is from bacteriophage T4 β-glucosyltransferase, PDB: 1jg7 (Vrielink et al, 1994). GT-C is the bacterial oligosaccharyltransferase PglB from Campylobacter lari, PDB: 3rce (Lizak et al, 2011). All the structures are represented as ribbons. The α-helices are colored in yellow, and the β-sheets are colored in blue. The size of each GT is indicated below the structure.

Structure and catalytic mechanism of GT-A glycosyltransferase

The overall structure of glycosyltransferases with the GT-A fold consists of β-sheets surrounded by α-helices (Fig. 5.2.2A). It consists of two dissimilar Rossmann-like motifs tightly associated with each other. The Rossmann-like fold is found on proteins that bind nucleotides and is responsible for binding the nucleotide sugar-donor substrate in glycosyltransferases.

Most GT-A glycosyltransferases have the DXD (Asp-X-Asp) signature motif for coordinating divalent metal ions. The metal ion plays an important role in binding the sugar donor by interacting with one or both acidic residues of the DXD motif and with the oxygen atoms from the phosphate group of the nucleotide sugar; therefore, it plays a crucial role in sugar-donor binding.

One of the most well-characterized GT-A glycosyltransferases is β1,4-galactosyltransferase 1 (β4Gal-T1) (Gastinel et al, 1999; Qasba et al, 2005; Qasba et al, 2008; Ramakrishnan et al, 2002; Ramakrishnan & Qasba, 2002). β4Gal-T1 is a membrane-bound enzyme that resides in the Golgi; it transfers galactose from uridine diphosphate galactose (UDP-Gal) to the terminal GlcNAc residue to form N-

acetyllactosamine. The reaction catalyzed by β4Gal-T1 follows a sequential ordered mechanism (Fig. 5.2.3). The nucleotide sugar donor and metal ion bind to the enzyme, which induces a conformational change that converts the enzyme from an open conformation to a closed conformation. This conformational change creates an acceptor-binding site for the incoming sugar acceptor. After catalysis, the product release from the complex is also sequential. The glycosylated acceptor is released first, followed by the nucleotide, which converts the enzyme back to the open conformation for another cycle of catalysis.
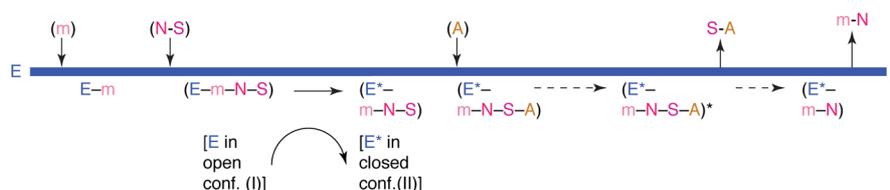

<div align="right">Adapted from (Ramasamy et al, 2005)</div>

Figure 5.2.3 The catalytic mechanism of β4Gal-T1

The catalytic cycle of glycosyltransferase starts with the enzyme (E) in the open conformation. The binding of nucleotide sugar (N-S) and metal ion (m) converts it to the closed conformation (E*). After product release, E* converts back to the open conformation.

The conformational change upon nucleotide sugar binding is not unusual in the proteins with a GT-A fold. Although the region that undergoes the conformational change may not be identical among glycosyltransferases, it is invariably located near the sugar donor-binding site. In β4Gal-T1, two flexible loops have been identified: a short loop made up of the residues 313–316 and a long loop made up of the residues 345–365 (Ramasamy et al, 2003). In the apo-structure (open conformation), the active site residues in the loops are located away from the catalytic site. UDP-Gal binds to the β4Gal-T1 with the metal ion in the open conformation, which induces a conformational change in the flexible loop region (Fig. 5.2.4). The large loop (I345–H365) flips from the open conformation to close the UDP-Gal binding site. The closed conformation also provides interactions between the long-loop residues and UDP-Gal and reduces accessibility to water in the active site. In addition, W314 in the short loop is oriented to bind to UDP in the closed conformation. The conformational change also creates an acceptor-binding site. Thus, an ordered binding of the donor and acceptor substrates is ensured.
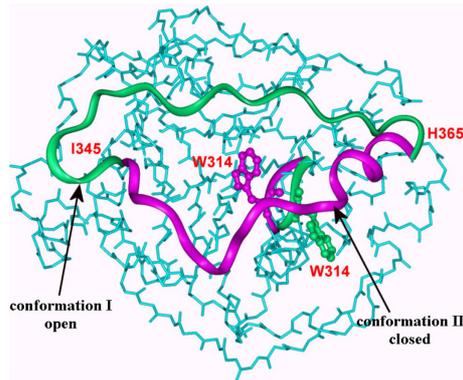
Figure 5.2.4 Conformational change upon sugar nucleotide binding

The loops shown by green ribbons are in the open conformation. In the closed conformation (magenta), W314 faces the catalytic pocket.

Structure and catalytic mechanism of GT-B glycosyltransferase

Similar to the GT-A fold, the architecture of the GT-B fold consists of two Rossmann-like domains. While the two domains have dissimilar sizes and are in close proximity in the GT-A fold, the sizes of the two Rossmann-like domains in the GT-B fold are similar and they are further apart, which creates a cleft between the two domains (Fig. 5.2.2B).

The prototypical GT-B fold of glycosyltransferases is β-glucosyltransferase (BGT) from the bacterial phage T4 (Vrielink et al, 1994), which uses UDP-glucose as the sugar donor and 5-hydroxymethylcytosine (5-HMC) in the viral DNA as the acceptor. Glycosylation of DNA can protect the DNA from host restriction enzymes.

The structure of BGT is comprised of two Rossmann-like folds (Fig. 5.2.5), separated by a central cleft. The structure of BGT in complex with DNA and UDP revealed that UDP and the 13-mer DNA bind to the inter-domain cleft (Lariviere & Moréra, 2002).
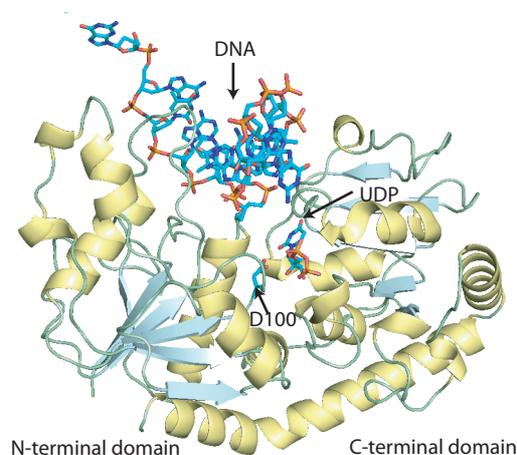
Figure 5.2.5 Structure of BGT in complex with DNA and UDP

The structure of BGT in complex with a 13-mer DNA and UDP (Lariviere & Moréra, 2002). BGT is shown by cartoon representation with α-helices colored in yellow and β-sheets colored in blue. The DNA and UDP are shown as sticks. D100 indicated in the structure acts as the catalytic base in the direct-displacement catalytic mechanism.

The conformational change observed in GT-A glycosyltransferases upon sugar-donor binding has also been observed in GT-B glycosyltransferases. Comparison of the two apo-BGT structures in different crystal forms with the UDP-BGT binary complex indicated that the UDP-BGT complex structure is more compact (Fig. 5.2.6). In the presence of UDP, BGT adopts a closed conformation. The conformational change originates from rigid body movement of the C-terminal to the N-terminal domain; the C-terminal domain of the UDP-BGT rotates 14º to the N-terminal domain compared to the apo-BGT structure. In contrast to the conformational change induced by binding to UDP, binding of a metal ion had no obvious effect on the structure of BGT (Moréra et al, 2001).
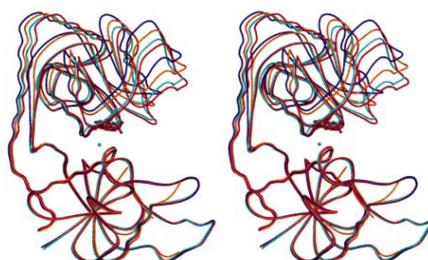
Adapted from (Moréra et al, 2001)

Figure 5.2.6 Stereo view of rigid body movement in BGT

Superposition of BGT on the N-terminal domains from apo-BGT (dark blue: tetragonal crystal, and orange: orthorhombic crystal), BGT-UDP-Mg^{2+} complex (light blue), and metal-free UDP-BGT complex (red). The Mg^{2+} is shown as a light blue sphere. UDP is shown by red sticks.

The combined structural and biochemical data of BGT indicate the catalytic mechanism of BGT. Asp100 (D100) acts as the catalytic base, which is in close proximity to the C1 carbon of UDP-glucose and 5-hydroxymethyl group of 5-HMC (Larivière et al, 2003). D100 deprotonates 5-HMC, which then attacks the anomeric C1 carbon of UDP-glucose, and forms the oxocarbenium ion transition state (see below). The crystal structure of the D100A-BGT mutant in complex with UDP-glucose further confirms that D100 is the catalytic base. UDP-glucose was hydrolyzed to UDP and glucose in the crystallization process in the presence of an active enzyme; however remained intact in the D100A complex structure (Larivière et al, 2003).

**Structure and catalytic mechanism of GT-C glycosyltransferase**

Glycosyltransferase with the GT-C fold were first predicted by amino acid sequence analysis; 5 families were determined to contain the GT-C fold (Liu & Mushegian, 2003). Currently, family 22, 39, 50, 57, 58, 59, 66, 83, 85 and 87 in the CAZY database are predicted to adopt the GT-C fold. The architecture of the GT-C fold was predicted to have an integral transmembrane domain composed of 8–13 α-helices located in the ER or plasma membrane and a soluble region with the active site in the lumen. Very recently, the structure of PglB (Fig. 5.2.2C), which is an oligosaccharide transferase (OST) for *N*-glycan synthesis, was determined and provided experimental data for the structure of the GT-C fold (Lizak et al, 2011).

The structure of PglB was solved with the sugar acceptor, which is a six-amino acid-long peptide (DQ<u>N</u>ATF). The peptide is located in the region between the transmembrane and cytoplasmic domains (Fig. 5.2.7A). The peptide forms a loop that turns 180º in the peptide-binding cavity, and presents the glycosylation site to the active site. This turn restricts the glycosylation site of PglB, so it must be presented on a flexible and surface-exposed loop that allows the peptide to be accommodated into the active site. The threonine at the +2 position of the glycosylation site stabilizes the interaction with PglB by forming three hydrogen bonds with the conserved WWD motif.

Like other ion-dependent GTs, the activity of OST is dependent on metal ion binding (Sharma et al, 1981). A metal ion was identified in the structure of PglB, which is located on the right side of the active site cavity (Fig. 5.2.7B). Consistent with the DXD motif in the GT-A fold, the GT-C fold has a DXD motif on the connecting loop of the transmembrane domain to coordinate the metal ion (D154 and D156). The metal ion in the active site also involves binding of the sugar donor. Unlike metal ion binding in the GT-A fold, E319 and D56 directly interact with the metal ion through the formation of a salt bridge with the metal ion, in addition to the binding interactions provided by the DXD motif. These data are consistent with mutational studies where single amino acid mutations of D154, D56, or E319 did not completely inactivate the enzyme, but double mutations such as D56A/E319A did abolish activity.

The authors proposed a three-step catalytic mechanism for PglB (Fig. 5.2.7C). The binding of the peptide to PglB induces a conformational change of the EL5 loop on the transmembrane domain. The EL5 in the apo-PglB is predicted to be flexible and disordered; binding orders the loop and restricts the movement of the acceptor in the

active site. This conformational change also moves E319 on EL5 to the active site. The binding of the sugar donor lipid-linked oligosaccharide (LLO) in the active site leads to nucleophilic attack of the amide nitrogen and creates an *N*-glycan on the peptide. The *N*-glycosylated peptide is then released from PglB, and EL5 moves back to the apo-PglB conformation.
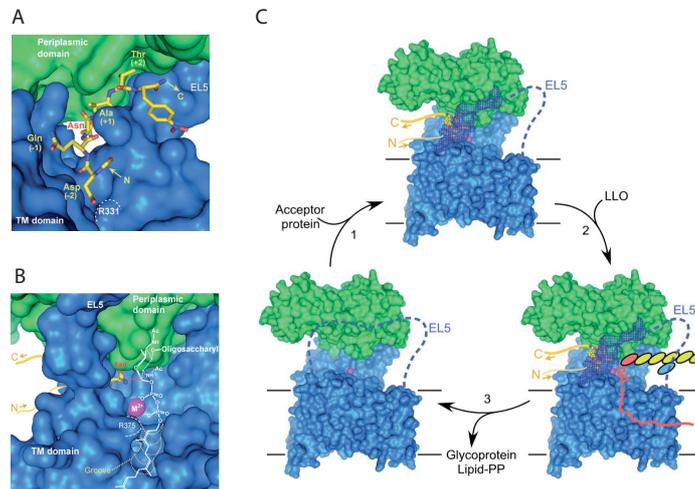
Figure 5.2.7 Structure and catalytic mechanism of PglB

The periplasmic domain of PglB is colored in green, and the transmembrane domain is colored in blue. In (A), the acceptor peptide, shown as sticks, turns 180º in the active site. The metal ion is shown in purple, and the sugar donor is modeled in the active pocket in white (B). The proposed catalytic mechanism of PglB is shown in (C). The first step is the binding of the peptide acceptor (yellow ribbon) in the active site, which also induces the conformational change of EL5. The second step is the binding of LLO, which results in the glycosylation. The products in the third step return EL5 to its original state.

Catalytic mechanism of glycosyltransferases

Glycosyltransferases catalyze the formation of a glycosidic bond between the sugar donor and the acceptor. During the reaction the configuration of the glycosidic bond can be either inverted, which has a different anomeric stereochemistry from the sugar donor, or resulted in retention, which retains the same stereochemistry (Fig. 5.2.8). Based on the outcome of the stereochemistry, glycosyltransferases can be classified as either retaining or inverting enzymes, which is correlated to different catalytic mechanisms of the glycosyltransferases (Coutinho et al, 2003).
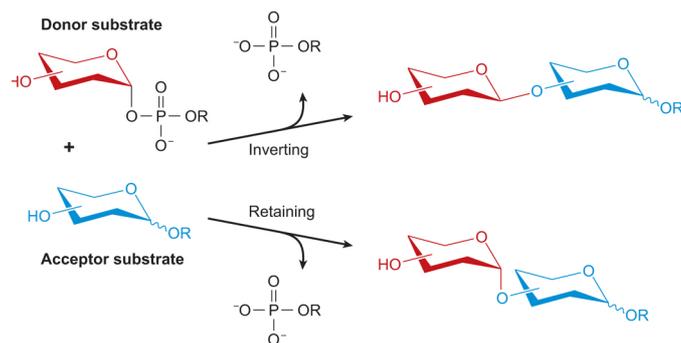
Figure 5.2.8 Inverting and retaining reactions

Glycosyltransferase catalyzes glycan transfer with either inversion or retention of the anomeric stereochemistry.

Mechanism of inverting glycosyltransferases

Similar to the well characterized catalytic mechanism of the inverting glycoside hydrolases, the inverting glycosyltransferases use an S_N2 reaction mechanism (Murray et al, 1996), which usually uses a general base residue (such as aspartic acid, glutamic acid, or histidine) to deprotonate the hydroxyl group on the acceptor. The deprotonated acceptor works as a nucleophile and attacks the electrophilic center on the sugar donor. A new bond with an inverted anomeric configuration is created through the oxocarbenium ion-like transition state (Fig. 5.2.9)
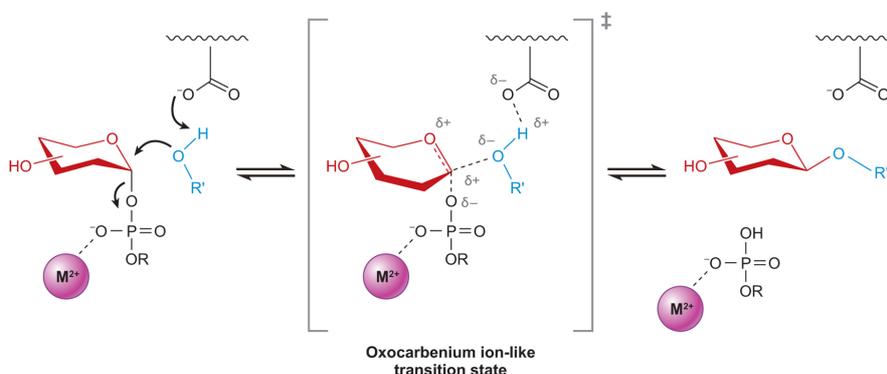
Figure 5.2.9 Catalytic mechanism of inverting glycosyltransferases

The inverted anomeric configuration is created *via* SN2-like reaction mechanism through the oxocarbenium ion-like transition state. R', in light blue, is the acceptor, and the sugar is colored in red.

Mechanism of retaining glycosyltransferases

The mechanism of retaining glycosyltransferases has been proposed to use a double-displacement mechanism with a glycosyl-enzyme intermediate (Fig. 5.2.10). In the double-displacement mechanism, an aspartic acid or glutamic acid side chain

in the active site attacks the sugar donor resulting in the inversion of the anomeric configuration. At this stage, the sugar is covalently linked to the active residue of the enzyme, and forms the glycosyl-enzyme intermediate. The phosphate group from the sugar donor acts as a base catalyst that deprotonates the sugar acceptor, which then attacks the glycosyl-enzyme intermediate in the second displacement step. In the second step, the conformation of the glycosyl-enzyme intermediate is again inverted, which results in overall retention of the stereochemistry.
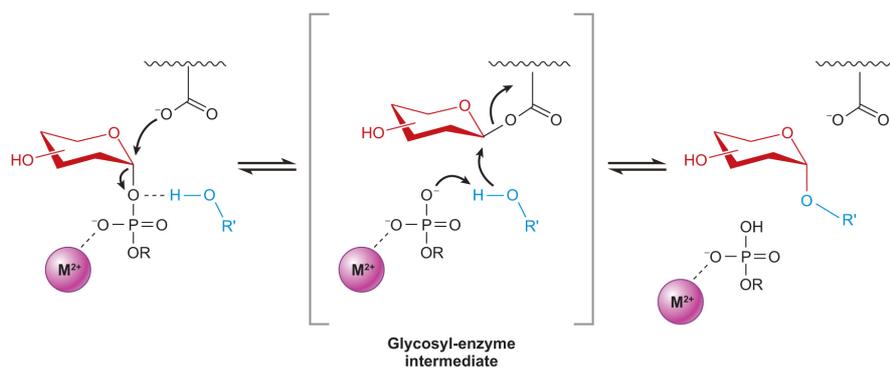
Figure 5.2.10 Double-displacement mechanism for retaining glycosyltransferases

In the first step, an acidic residue on the enzyme acts as a catalytic residue to attack the sugar donor at the C1 position and forms a glycosyl-enzyme intermediate with inverted stereochemistry. In the second step, the sugar acceptor attacks the intermediate and displaces the sugar to the acceptor with another inversion of the configuration. Overall, the stereochemistry is retained.

5.3 Dysregulation of Glycosylation and Human Diseases

The importance of glycosylation has increased in the last decade because of its emerging role in human diseases. Based on the differences in the onset of dysregulation of glycosylation, diseases can be divided into two types: acquired diseases and inherited diseases. This chapter will summarize the human diseases related to glycosylation defects.

Acquired human diseases

In tumor environment, changes in the glycosylation levels are associated with increase in the invasive and metastatic properties of tumor cells. Aberrant glycosylation has been shown to be involved in different stages of tumor development, such as tumor proliferation, invasion, angiogenesis, and metastasis. The function of glycoconjugates in these steps will be discussed in this section.
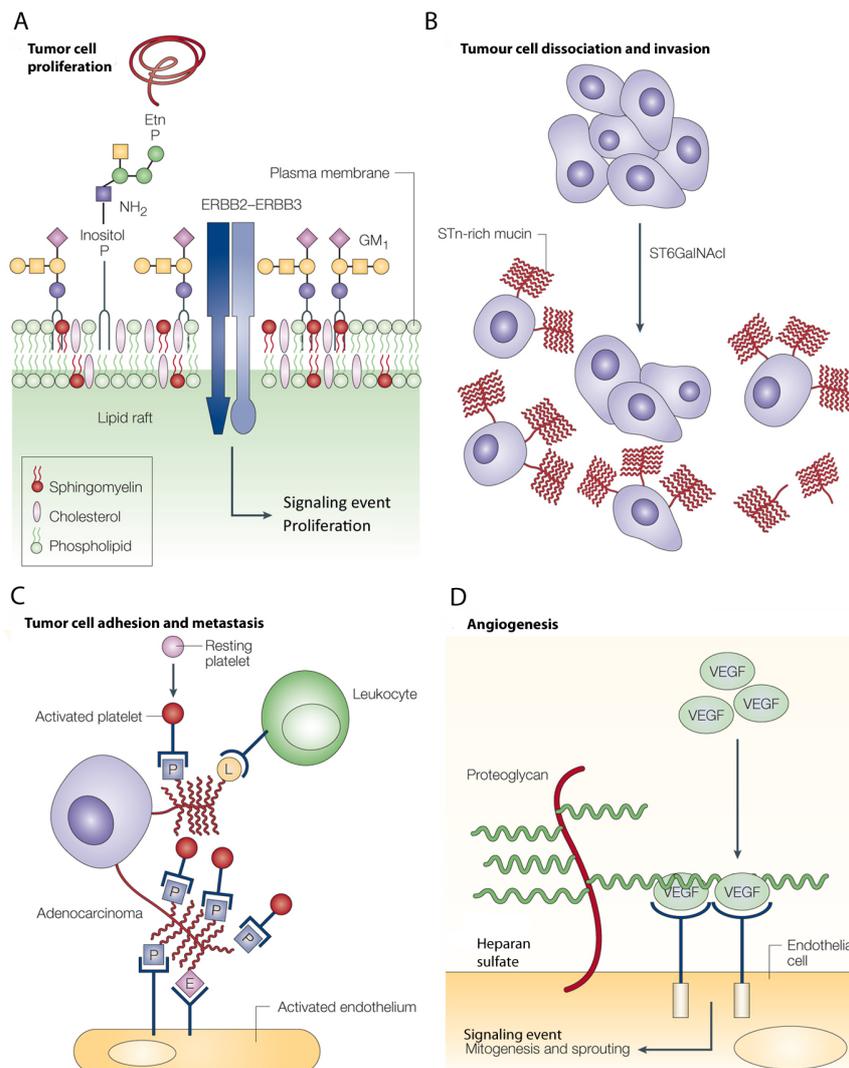
Adapted from (Fuster & Esko, 2005)

Figure 5.3.1 Glycans are involved in tumor development

Various growth-factor receptors are modulated by glycans through oligomerization of their respective receptor tyrosine kinase receptors in lipid rafts; this process is mediated by gangliosides. (A) GM_1 can modulate heterodimerization of the ERBB2-ERBB3 receptor, and therefore, activate downstream signaling pathways. (B) Up-regulation of ST6GalNAc I increases the presentation of STn antigen on mucin-type *O*-glycans, which facilitates the dissociation of tumor cells. (C) Dissemination of tumor cells through the bloodstream is facilitated by the association of tumor cells with leukocytes and platelets. The association is modulated by SLe^a and SLe^x epitopes on the tumor cells and L- and P-selectin on leukocytes and platelets, respectively. E-selectin on the endothelium arrests tumor cells and facilitates metastasis. (D) Angiogenesis is the process of sprouting new blood vessel endothelial cells in response to pro-angiogenic factors such as vascular endothelial growth factor (VEGF). Heparan sulfate on the endothelial cell surface facilitates VEGF binding and activation of downstream signaling.

Glycoconjugates in tumor proliferation

Mucins are glycoproteins that contain clustered *O*-glycans on the core protein (Chapter 5.1). Mucins are expressed in most carcinomas in a secretory form or as a membrane-bound protein. Altered expression of mucins has been reported in tumors, including the up-regulation of MUC4 on the cell surface of adenocarcinoma cells (Sherblom et al, 1980). MUC4 contains two EGF-like repeats that can act as ligands for the receptor tyrosine kinase ERBB2 (Carraway et al, 1992). The up-regulated EGF-like motif can activate ERBB2 in the absence of typical ligands in breast carcinoma. In a nude mice model, MUC4 overexpression was shown to autophosphorylate ERBB2 and contribute to the suppression of apoptosis (Komatsu et al, 2001).

Gangliosides are a group of sialic acid-containing glycosphingolipids. Gangliosides can interact with tyrosine kinase receptors and regulate their phosphorylation. Overexpression of ganglosides has been reported in many cancer types (Hakomori, 1990). In the membrane of breast cancer cells, gangliosides regulate the formation of heterodimer receptors as lipid rafts on the membrane (Fig. 5.2.1A). The formation of the heterodimer receptor is crucial for responding to growth factor stimulation, thereby inducing cell proliferation (Nagy et al, 2002).

Glycoconjugates in tumor invasion

Invasion is a process whereby the tumor cell detaches from the original location and migrates to other tissues, which requires the cell to modify the receptors and ligands on its cell surface in order to reduce the affinity for its original location. Proteases and glycosidases may also be released for remodeling the ECM and increasing the dissociation of tumor cells. Modification of the structures of *N*- and *O*-glycans has

been identified to be involved in the invasion of tumor cells (Fuster & Esko, 2005). One of the most prominent changes regarding glycosylation on tumor cells is the enhanced expression of sialic acid. Because of the negative charge of sialic acid, the enhanced expression of sialic acid could promote cell detachment through charge repulsion. Sialic acid has been found to affect both *N*- and *O*-glycans. On *N*-glycans, the capping of the terminal galactose with sialic acid, catalyzed by ST6Gal I (Lin et al, 2002), can reduce cell-cell interactions, and therefore, promote invasion. One of the most well studied tumor-associated antigen is sialyl Tn (STn) antigen, which is an *O*-linked GalNAc (Tn antigen) with a sialic acid on the 6-OH group of GalNAc. The emergence of STn antigen in many cancers has been related to a defect in COSMC, which is a chaperone involved in the maturation of the galactosyltransferase acting on the Tn antigen (Ju & Cummings, 2002). Lack of galactosyltransferase activity results in the accumulation of Tn antigen, thereby increasing STn formation. Transfection of breast cancer cell lines with cDNA encoding ST6GalNAc I that is responsible for transferring sialic acid to Tn antigen can increase STn expression and also increase the migration potential of the transfected cells (Julien et al, 2001). The STn antigen on mucins generates a negative charge on the cell surface and induces cell-cell dissociation by repulsion (Fig. 5.2.1B).

Glycoconjugates in adhesion and metastasis

After entry of the tumor cell into the blood vessel, it can adhere to the platelets and leukocytes, and eventually to the endothelial cells in distant tissues (Fig. 5.2.1C) (Läubli & Borsig, 2010). The cell-cell interaction between tumor cells and other cells is modulated by ligand-receptor interactions. The receptor that regulates the interaction is called selectin, which is a lectin that binds to sialyl Lewis x (SLe[x]) and sialyl Lewis a (SLe[a]) structures. Different cell types have different subtypes of selectin: platelets, leukocytes, and endothelium have P-, L-, and E-selectin, respectively. Solid tumors such as colon and pancreatic carcinomas frequently overexpress SLe[a] and SLe[x] on glycoproteins or glycolipids on the cell surface. These overexpressed SLe[a] and SLe[x] act as ligands that bind to P-, L-, and E-selectin on different cells, and enhance migration and adhesion of tumor cells. The level of expression of these ligands is inversely correlated with patient survival (Baldus et al, 1998).

Glycoconjugates in angiogenesis

Angiogenesis is a crucial process for solid tumors to acquire nutrition and oxygen through newly synthesized blood vessels. Up-regulation of heparan sulfate proteoglycans (HSPGs) had been found on the microvascular endothelia, but not on macrovascular endothelia, which implicates a role of HSPGs in angiogenesis (Marcum & Rosenberg, 1985). HSPGs can bind to pro-angiogenic factors such as fibroblast growth factor (FGF), vascular endothelial growth factor (VEGF), and platelet-derived growth factor (PDGF). In case of VEGF (Fig. 5.3.1D), binding of HSPGs to VEGF regulates the interaction with the VEGF receptor (Gitay-Goren et al, 1992). The binding can be abolished by treating the cell with heparinase, which digests the cell surface HSPGs and prevents the binding of VEGF to its receptor (Gitay-Goren et al, 1992).

Inherited human diseases

Glycosylation of proteins is the most common form of post-translational modification. More than 50% of proteins are glycosylated, and 1% of the genetic material is involved in glycosylation processes (Lowe & Marth, 2003). These glycoconjugates are involved in a variety of biological roles such as cell-cell interactions, protein folding, and signal transduction. Therefore, dysregulation of glycoconjugate biosynthesis can often cause diseases. Congenital disorders of glycosylation (CDG) are a group of disorders marked by defects in the glycosylation of proteins or lipids. Since the first patient identified in 1980 (Jaeken et al, 1980), the number of patients with these disorders has grown tremendously (Fig. 5.3.2). Defects in the glycosylation pathway are not limited to hypoglycosylation of *N*- and *O*-glycans on proteins; recent cases have identified hyperglycosylation of glycoprotein as another defect (Jaeken, 2010).
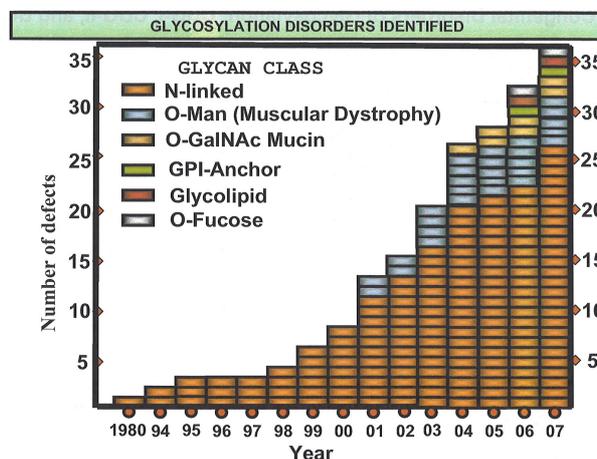
Figure 5.3.2 The growing number of CDG since 1980

Since the first case discovered in 1980 on *N*-linked proteins, the number of CDG cases has rapidly increased owing to the availability of gene sequences and cooperation between clinicians and scientists.

Defects in *N*-glycosylation

N-glycosylation is a complex process comprising three parts: assembly, attachment, and processing. Moreover, these steps occur in three different compartments: assembly of LLO in the cytosol, attachment of LLO to nascent proteins in the ER, and processing of *N*-glycan structures in the ER and Golgi complex (Fig. 5.3.3). Conventionally, CDG with defects in the *N*-glycan biosynthesis pathway can be categorized into two sub-types: type I CDG and type II CDG (Fig. 5.3.3). In type I CDG, the defect abolishes or reduces the efficiency of one of the sequential steps in the biosynthesis of LLO in the cytosol. Isoelectric focusing (IEF) analysis of serum transferrin from patients with type I CDG shows an increase in the levels of di- and asialotransferrin (Fig. 5.3.4). Type II CDG defects impair the steps involved in trimming, remodeling, and decorating *N*-glycan structures with a variety of monosaccharides (Fig. 5.3.3). The transferrin pattern of type II CDG shows the isoforms of transferrin with one to six sialic acids (Fig. 5.3.5). Differences in the results of IEF analyses of transferrin between type I and type II CDG are caused by the defects in different parts of *N*-glycosylation. For type I CDG, mutations were found in the assembly of LLO, and aberrant LLOs may result in proteins with unoccupied *N*-glycosylation sites. In case of type II CDG, the occupancy of the *N*-glycosylation site is not altered, but the structure of *N*-glycans is impaired, which leads to different degrees of sialylation on the *N*-glycans of transferrin (Fig. 5.3.3 and 5.3.4). In this section, the most prevalent CDGs will be introduced.
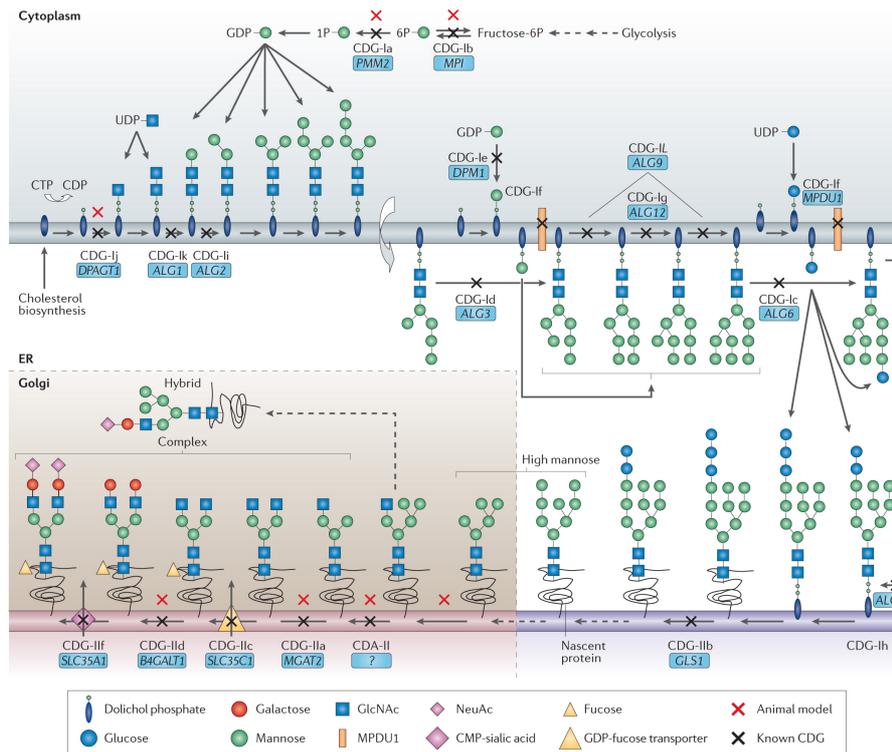
Adapted from (Freeze, 2006)

Figure 5.3.3 Defects in the *N*-linked glycosylation pathway

The *N*-glycan biosynthesis pathway is illustrated in three different cellular compartments. The names of the CDG (CDG-X) and corresponding gene (boxed with blue background) are also indicated. The nomenclature of CDG in the text is in accordance with a new system, which is not consistent with this figure. The new nomenclature uses the official gene symbol follow by "-CDG" (Jaeken et al, 2009).

PMM2-CDG is the most prevalent CDG with more than 500 patients worldwide and an estimated incidence of 1/20,000. The mutation occurs in phosphomannomutase 2 (PMM2) that converts mannose-6-phosphate to mannose-1-phosphate, which is the precursor of GDP-mannose (Fig. 5.3.3). The mutation in PMM2 causes a reduction in the levels of GDP-mannose and LLO. Shortage of substrate for oligosaccharidetransferase leads to the production of glycoproteins with unoccupied *N*-glycosylation sites. Nearly 90 mutations have been identified in PMM2, which have been shown to decrease PMM2 activity, reduce affinity for mannose-6-phosphate, or result in complete loss of activity in a few cases (Kjaergaard et al, 1999). The most commonly identified mutation in patients is R141H on PMM2, which is an inactive enzyme in *in vitro* assays (Kjaergaard et al, 1999; Pirard et al, 1999). The R141H mutation on PMM2 has never been found in the homozygous state, suggesting that PMM2 activity is crucial for survival and total loss of PMM2 activity is likely lethal (Kjaergaard et al, 1999; Pirard et al, 1999).
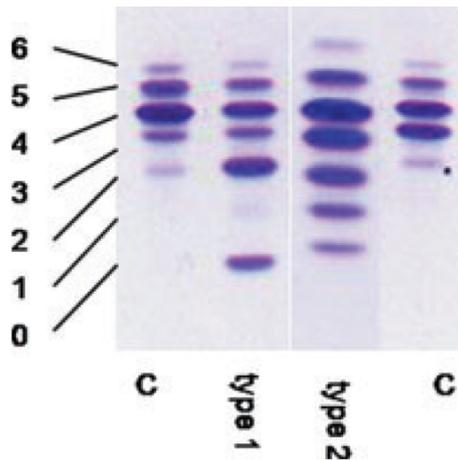
Figure 5.3.4 Isoelectric focusing of transferrin

The number of sialic acids on *N*-glycans of transferrin is indicated on the left from 0 to 6. The samples from healthy individuals are termed as controls (C), and those from patients with type 1 CDG and type 2 CDG with *N*-linked glycosylation defects are termed as type 1 and type 2.

Another cytosolic enzyme in the GDP-mannose synthesis pathway is phosphomannose-isomerase (MPI) (Fig. 5.3.3). MPI-CDG patients have a milder phenotype compared to those with PMM2-CDG, and MPI-CDG is the only known CDG with an efficient treatment protocol. Orally supplemented mannose (1 g/kg body weight) can be converted to mannose-6-phosphate by hexokinases, thus bypassing the defect and effectively treating the disease (de Lonlay & Seta, 2009).

Type II CDG seems to be less frequent than type I CDG, but this may reflect the lack of awareness or lack of efficient analytical methods rather than actual prevalence (Freeze, 2007). Two patients with MAG2-CDG (CDG-IIa) have been independently reported with similar phenotypes including severe mental retardation, chronic feeding problems with severe diarrhea, and growth retardation (Jaeken et al, 1993; Ramaekers et al, 1991). The gene responsible for the disease was identified as *N*-acetylglucosaminyltransferase II (Fig. 5.3.3) that transfers GlcNAc to Man, which is one of the key steps for synthesizing hybrid and complex type *N*-glycans.

Recently, mutations in the OST complex have been identified in patients with autosomal recessive mental retardation (Garshasbi et al, 2008; Molinari et al, 2008). The mutations are not in central components such as Stt3, but in two mutually exclusive paralogs, i.e., Tusc3 and IAP. These two proteins share 70% sequence identity and possess identical membrane topology. The function of these two proteins in humans has not been clarified. The homologous proteins of Tusc3 and IAP in yeast are OST3p and OST6p, respectively. A study on the OST3p and OST6p in yeast showed oxidoreductase activity (Schulz et al, 2009). Based on sequence

similarity, Tusc3 and IAP are proposed to possess oxidoreductase activity. Interestingly, neither the IEF analysis of the transferrin nor the *N*-glycan profile obtained by the mass spectrometry analysis of serum proteins showed difference between samples from patients and normal individuals. Furthermore, OST complexes from the patients also had similar activity as the wild-type OST complex, in *in vitro* assay (Molinari et al, 2008). The authors proposed that Tusc3 and IAP in humans, with their oxidoreductase activity, might aid in the folding of brain specific proteins. Misfolding of these proteins causes mental retardation syndromes in patients.

Defects in *O*-glycosylation

O-Glycosylation defects that cause human diseases have been noted in *O*-mannose, *O*-xylose, and *O*-fucosylation pathways (Jaeken, 2010). Defects in *O*-mannosyltransferase 1 and *O*-mannosyltransferase 2 can cause Walker-Warburg syndrome (POMT1/POMT2-CDG), a rare disease characterized by congenital muscular dystrophy. Glycans that are *O*-linked through xylose to serine are called glycosaminoglycans. Xylose can be extended to tetrasaccharide by the addition of two galactose residues and one glucuronic acid residue, which is the precursor of heparan sulfate and dermatan sulfate (Freeze, 2007). Mutations in EXT1 and EXT2 (EXT1/EXT2-CDG), which are the enzymes required to extend xylose with galactose and glucuronic acid, cause multiple cartilaginous exostoses in patients (Ahn et al, 1995).

Two defects have been identified in the glycosyltransferase in the *O*-fucosylation pathway (Fig. 5.3.5). Under normal circumstances, the *O*-fucose on the EGF-like repeat can be extended by Fringe, which is a β1,3-*N*-acetylglucosaminyltransferase that adds GlcNAc to position 3 of fucose. Patients with spondylocostal dysostosis type 3 disease (SCD3) have a homozygous mutation in Fringe (Sparrow et al, 2006), wherein a conserved phenylalanine at position 188 is mutated to leucine. Although F188 is not directly involved in substrate binding and is not an active site residue, the F188L mutation still abolishes enzyme activity (Sparrow et al, 2006). The activity of Fringe modulates the substrate specificity of the Notch receptor, and therefore, plays an important role in development and other important biological events (Luther & Haltiwanger, 2009; Stanley, 2007).

Another disease related to the *O*-fucosylation pathway is Peters Plus syndrome, which is caused by inactivating biallelic truncating mutations in β1,3-

glucosyltransferase (β3Glc-T) (Heinonen et al, 2003; Kozma et al, 2006; Lesnik Oberstein et al, 2006). β3Glc-T is involved in the synthesis of an unusual Glc-β1,3-Fuc-*O*- dissacharide unit on the thrombospondin type 1 repeat (Fig. 5.3.5). TSR is found in extracellular and membrane proteins that function in angiogenesis, neuronal guidance, and tissue remodeling during development (Tucker, 2004). Although TSR-containing proteins are presumed to be involved in these biological processes, the detailed mechanism underlying the regulation of the function of TSR-containing proteins by glycosylation is still unclear.
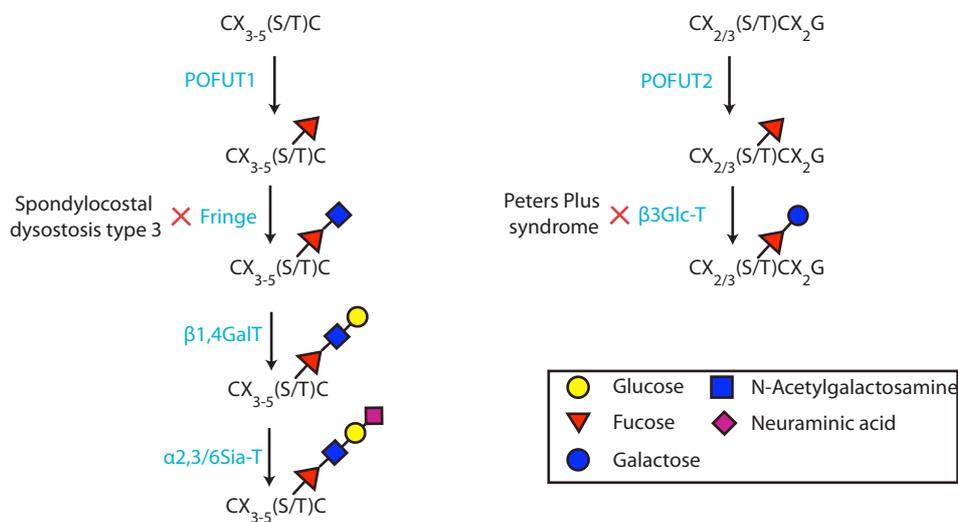

Figure 5.3.5 The CDG related to *O*-fucosylation pathways

O-Fucosylation on EGF-like repeats is initiated by POFUT1 and can be extended to GlcNAc-β1,3-Fuc-*O*-disaccharide by Fringe. Fringe mutations cause spondylocostal dysostosis type 3 syndrome. In the *O*-fucosylation pathway of TSR, *O*-fucose is extended to Glc-β1,3-Fuc disaccharide by β3Glc-T. Defects in β3Glc-T cause Peters Plus syndrome.

Other type of CDGs

Most CDGs result from hypoglycosylation. Diseases that result from the gain of glycosylation sites are relatively unexplored. Three children with susceptibility to mycobacterial diseases were identified as homozygous for a missense mutation in the gene of interferon γ receptor ligand-binding chain (IFN$_\gamma$R2) (Vogt et al, 2005). The mutation creates a new *N*-glycosylation site at the position 168 (T168N) in IFN$_\gamma$R2. This additional *N*-glycan in IFN$_\gamma$R2 does not affect the expression of the protein; it, however, is necessary and sufficient to abolish the function of IFN$_\gamma$R2. By searching the Human Gene Mutation Database for mutations resulting in the gain of *N*-glycosylation sites on secretary proteins, 142 mutations in 77 genes that

correspond to 1.4% of mutations and 13.3% of the secretary protein genes were identified. The high prevalence of human proteins in secretory pathways with gain-of-*N*-glycosylation-site mutations might be responsible for uncharacterized human genetic diseases. Their pathogenic effects may be a direct consequence of the additional *N*-glycan in the protein.

Clinical approaches to identify patients with CDG

The most popular analytical approach to identify a CDG related to *N*-glycosylation pathways is IEF analysis of transferrin. Transferrin is a glycoprotein with two *N*-linked glycans. In healthy individuals, the majority of *N*-linked glycans on each glycosylation site are biantennary, containing one or two sialic acids; a small fraction of the sugar chain has a triantennary structure with three sialic acids. The transferrins in IEF gels are separated according to the differences in the number of negatively charged sialic acids, which may range from 2 to 6 in healthy individuals (Fig. 5.3.4), which represents 1–3 sialic acid residues on each *N*-glycan.

IEF analysis of transferrin is an inexpensive, efficient, and reproducible approach feasible in an experienced laboratory. However, false-negative results have been obtained using this method. Moreover, defects in CDG are not limited to the *N*-glycosylation pathway. Apolipoprotein, a serum protein with mucin-type *O*-glycosylation, has been proposed to be used as a indicative marker to detect defects in the mucin-type *O*-glycosylation pathway (Wopereis et al, 2007), which can complement IEF analysis of transferrin. IEF analysis of transferrin combined with that of apolipoprotein provides an efficient method to identify defects in the biosynthesis pathways of *N*-glycans and mucin-type *O*-glycans.

Mass spectrometry has also been used to characterize CDG. An automated online immunoaffinity liquid chromatography-mass spectrometry method has been developed to analyze transferrin (Lacey et al, 2001). This method was first used to purify transferrin from 25 μL of serum by using an affinity column. The transferrin elution fraction was loaded onto a C4 column with a 9-min gradient, and then subjected to mass spectrometry. This approach was used to analyze the *N*-glycosylation site occupancy of intact transferrin. Other approaches that have analyzed tryptic peptides with *N*-glycan and mucin-type *O*-glycan have also been reported (Tajiri et al, 2005). An approach using immunopurification-mass spectrometry has been established to analyze samples from patients with Peters Plus syndrome (Hess et al, 2008). The properdin from serum samples was

immunopurified and digested with trypsin. The digested samples were analyzed by mass spectrometry using multiple-reaction monitoring that offers high sensitivity and specificity. The results showed that only Fuc-*O*- was found on all four *O*-fucosylation sites from the properdin purified from patients' serum, but in normal individuals, properdin carried Glc-β1,3-Fuc-*O*- disaccharide on the fucosylation sites. This approach highlights the importance of dedicated analytical approaches to identify rare glycosylation defects that cannot be elucidated by using general methods.

5.4 Thrombospondin type 1 repeat

Thrombospondin (TSP) was first identified as a glycoprotein released from the α-granules of platelets in response to thrombin stimulation (Baenziger et al, 1971). TSP is a multi-modular protein with two major groups. In humans, two genes encode Group A TSPs (TSP-1 and TSP-2). The TSPs in Group A can form homotrimers with their N-terminal oligomerization domain. Following the oligomerization domain, there is a von Willebrand Factor Type C (vWC) domain, three thrombospondin type 1 repeats (TSR), three epidermal growth factor (EGF)-like repeats, a calcium-binding wire, and a lectin-like C-terminal domain (Fig. 5.4.1A). TSP-3, TSP-4, and TSP-5 in Group B TSPs form pentamers with their N-terminal oligomerization domain, with one additional EGF-like repeat, but they lack TSR and vMC modules.
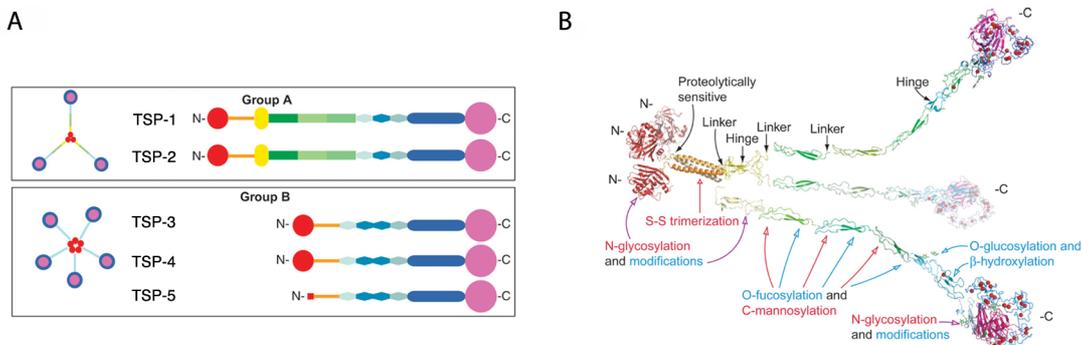
(Carlson et al, 2008)

Figure 5.4.1 Modular structures of thrombospondin families and a model of Group A TSP

(A) The modules of TSPs. The members of group A TSPs (TSP-1 and TSP-2) form a homotrimer. TSP-3, TSP-4, and TSP-5 comprise Group B and form pentamers. The modules are color-coded as follows: TSP-N (red), coiled-coil (orange), vWC (yellow), TSR1-3 (green), EGF-like repeat (blue), calcium-binding wire (blue), and lectin-like module (purple). (B) The structure of a Group A TSP was modeled *in silico*; the module structures were solved individually. The colors of the modules in the structure are the same as in (A). The flexible sites are labeled with black text, and the sites with co- and post-translational modifications are colored with red and cyan, respectively.

Because of the large size, flexibility between modules, and post-translational modifications on the protein, NMR-based elucidation of the structure of entire TSP is impossible, while it is quite challenging by X-ray crystallography. Electron microscopy and low-resolution biophysical studies have been used to elucidate the structure of the individual modules and describe the molecular organization of TSP-1 (Fig. 5.4.1B) (Carlson et al, 2008).

The TSR was first described in human TSP-1 in 1986 (Lawler & Hynes, 1986). However, TSRs are not limited to TSP-1 and TSP-2, but are also present in other multi-modular proteins. In humans, there are 65 proteins that contain TSRs

(Pfam:00090), such as the proteins involved in the formation of extracellular matrix (Spondin-1), metalloproteases (ADAMTS family), and the proteins in the complement pathway (C6, C7, C8). The number of TSR modules in different proteins varies from 1 to 24. These TSR-containing proteins constitute a TSR superfamily with functions related to matrix organization, cell-cell interactions, and cell guidance (Tucker, 2004). This chapter addresses the structure, post-translational modifications, and function of TSRs, as well as the pharmaceutical applications of the TSR-derived peptides.

Structure of TSR

TSRs are about 55-amino acid long, of which 12 residues are highly conserved. The sequence alignment of different TSRs and the 3D structure of TSR revealed that these conserved residues provide crucial interactions for the unique three-strands anti-parallel structure (Tan et al, 2002). The structure of TSR[2,3] was first solved in 2002, from human TSP-1, to 1.9 Å resolution (Tan et al, 2002) (Fig. 5.4.2). The overall structure of each TSR is compact with dimensions of 15 × 20 × 55 Å and contains three pairs of conserved disulfide bonds. Each TSR has three anti-parallel strands: the irregular A-strand with a unique ripple conformation and the B- and C-strands form a β-sheet structure. Pro472 between TSR2 and TSR3 creates a 180º turn that allows the two TSRs to face in opposite directions. This linkage between two TSRs is relatively rigid and inflexible because of the Cys422-Cys471 disulfide bond and a hydrophobic interface on the loop between two TSR modules. Interestingly, the linkage between TSR1 and TSR2 also has a Pro at a similar position, which may also create the 180º turn between TSR1 and TSR2.
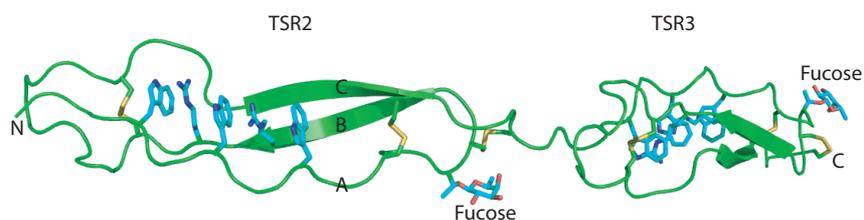
Figure 5.4.2 Structure of hTSR[2,3] from human thrombospondin-1

The disulfide bonds (in yellow) and *O*-linked fucose moieties (in cyan) at Thr432 and Thr489 are drawn in stick representation. The A-B-C-strands and fucose in the TSR2 domain are indicated. Side chains that are involved in the WR layers in TSR2 and TSR3 are shown as sticks.

The anti-parallel structure of TSR is maintained by unique CWR layers among the three strands. On the A-strand, three conserved tryptophan residues in a W**W**W motif constitute the W layers. The guanidinium group of arginines form the R*R*R motif on the B-strand, sandwiched by two adjacent W layers comprise the R layers

(Fig. 5.4.2). The alternative stacking of aromatic tryptophans and the planar cationic groups of arginine form multiple cation-π interactions that provide the main stabilizing energy for TSR folding. The distance between the tryptophan and arginine layers is about 4.6 Å throughout TSR2 and TSR3. Furthermore, the WR layers are capped by disulfide bonds on both sides of the layers. The first disulfide bond between the B-C-strand (Cys444–Cys456 in TSR2) caps the left side of the WR layer, and the second disulfide bond between the A-C-strand (Cys429–Cys466 in TSR2) caps the other side of the layer (Fig. 5.4.2). The stacking side chains from the tryptophan and arginine residues and the disulfide bonds create the CWR layers between three strands; these layers comprise the characteristic core of the TSR structure. The CWR layers provide stabilization energy for the TSR structure, and it creates a front face for proposed interactions. Because of the inherent strand twist, these layers have a right-handed spiral groove with a positive charge on the surface, which is proposed to interact with negatively charged ligands such as heparin sulfate. Based on the disulfide bond patterns, TSRs can be divided into two main groups (Fig. 5.4.3) (Tan et al, 2002). In group 1 TSRs, the disulfide pattern is C_1-C_5, C_2-C_6, and C_3-C_4. This group of TSRs includes TSP, ADAMTS, and properdin. Another structural feature of group 1 TSRs is two hydrogen bonds between a structural motif called jar handle and the N-terminus, which pulls the A-strand closer to the TSR core structure and stabilizes the conformation. In group 2 TSRs, which includes F-spondin and TRAP, the disulfide bond pattern is C_1-C_4, C_2-C_5, and C_3-C_6. The group 2 TSRs do not have a jar handle created by hydrogen bonds; instead, the N-terminus of the A-strand is stabilized by a disulfide bond between C_1 and C_4 (Fig. 5.4.3).
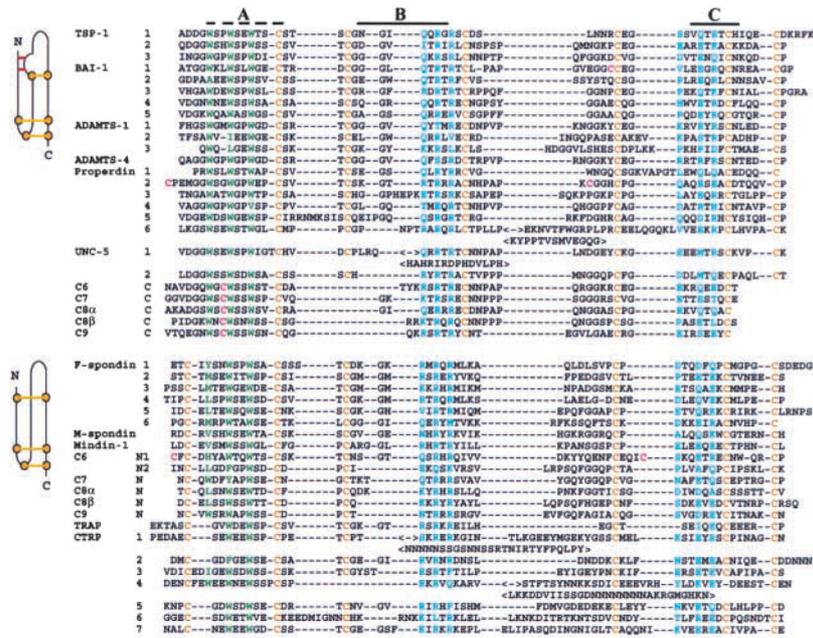
Adapted from (Tan et al, 2002)

Figure 5.4.3 Multiple sequence alignments of TSRs

The alignment of TSRs from the same or different proteins is categorized into different groups according to their disulfide bond pattern. Group 1 and 2 TSRs are shown in the upper and lower panels, respectively. Two hydrogen bonds between the N-terminus and jar handle are drawn as red lines. The cysteine residues that are involved in forming the disulfide bonds are in yellow. The regions of A-, B-, and C-strands are indicated above the sequence. The tryptophan and arginine residues that create the core region are colored in green and cyan, respectively, in the alignment.

Although there is a different disulfide bond pattern between group 1 and 2 TSRs, the overall structures are similar. Three structures from group 2 TSRs, i.e., TSR1 and TSR4 from F-spondin and TRAP-TSR from *P. falciparum*, were determined by NMR (Pääkkönen et al, 2006; Tossavainen et al, 2006). Comparison of the structures of TSRs from F-spondin (group 2), TSP-1 (group 1), and TRAP-TSR (group 2) shows that the loop between the A- and B-strands superimpose with RMSDs from 1.0 Å to 2.1 Å between the structures. The overall structure shows high structural similarity with an RMSD from 2.3 Å to 2.9 Å (Fig. 5.4.4).


A   B   C

Modified from (Tossavainen et al, 2006)
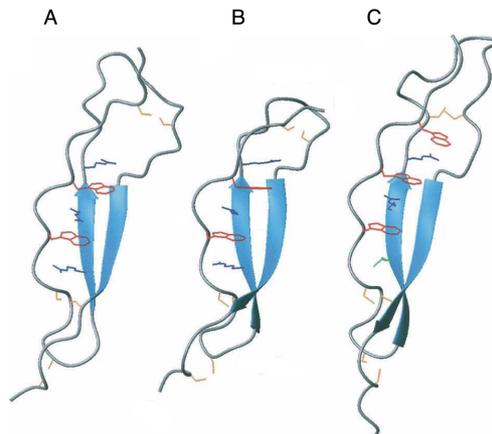
Figure 5.4.4 Comparison of TSR structures

Cartoon structures of (A) TSR4 from F-spondin (1VEX), (B) TRAP-TSR (2BBX), and (C) TSR2 from TSP-1 (1LSL) are shown. The cysteine, arginine, and tryptophan residues in the CWR layers are shown in sticks and colored in yellow, green, and blue, respectively. The overall fold of the TSRs is similar, although there are differences in the disulfide-bonding pattern.

Glycosylation on the thrombospondin type 1 repeat

TSRs undergo two types of unique glycosylation: one is *C*-mannosylation on a tryptophan of the A-strand and the other is *O*-fucosylation on a serine/threonine on the loop between the A- and B-strands (Gonzalez de Peredo et al, 2002; Hofsteenge et al, 2001). *C*-Mannosylation is a co-translational modification. The α-mannosyl residue is linked to the C-2 atom of the tryptophan through a C-C bond (scheme on the right). The modification occurs in the ER by a glycosyltransferase, which has not been identified. The unknown *C*-mannosyltransferase can recognize a synthetic W-X-X-W peptide and transfer the mannose to the first W, which indicates that *C*-mannosylation can transfer mannose to an unfolded protein with the consensus motif (Doucey et al, 1998). A study on TSP-1 from human platelets showed that *C*-mannosylation was very abundant; the occupancy of *C*-mannose on the consensus sequence was more than 90% in the determined sites. However, these sites on the recombinant TSRs from insect cells revealed lower or no *C*-mannosylation (Hofsteenge et al, 2001). In the structure of TSR2-TSR3 from F-spondin (Tan et al, 2002), no electron density was observed for *C*-mannosylation, probably because the TSR was expressed in insect cells, which lacks the *C*-mannosylation machinery. However, the C2 positions on the side chains of all tryptophans were pointing away from the Cα backbone, which makes the modification site more accessible for the unidentified *C*-mannosyltransferase. Recently, the crystal structure of TSR with *C*-mannosylated-tryptophan was determined from the human complement component C8 (Lovelace et al, 2011). Compared to the structure of TSR[2,3] expressed in insect cells, C8 was purified from human plasma; hence, the tryptophan on the W-X-X-W motif could be modified by mannose (Fig. 5.4.5). The structure confirmed that the mannose was on the C2 position of tryptophan, but it lacked the *O*-fucose that was found on the structure of TSR[2,3] because of a proline instead of a serine or threonine at the consensus site. Interestingly, they also observed a clear electron density that fit better to a β- rather than α-mannosyl substituent (Lovelace et al, 2011). However, the

44

authors did not provide an explanation for this unique finding. It should be noted that *C*-mannosylation is not limited to the TSR module, it has also been found on proteins such as ribonuclease 2 (de Beer et al, 1995), and mucins (Perez-Vilar et al, 2004).
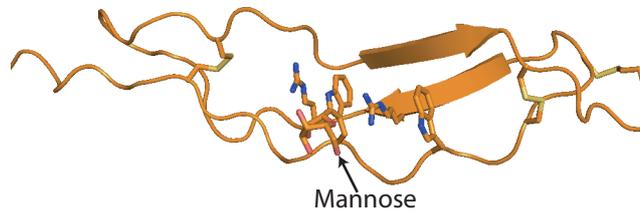

Mannose

Figure 5.4.5 *C*-Mannosylation on complement C8

The structure of TSR from the complement C8α is shown as a cartoon representation (PDB: 3OJY). The side chains of the tryptophan and arginine residues involved in the core layer formation are presented as sticks. The mannose is found on the first tryptophan of the W**W motif.

The function of *C*-mannosylation is still not known mainly because the *C*-mannosyltransferase has not been identified. The results from a recent study suggest that *C*-mannosylation affects the secretion of the TSR-containing protein punctin-1. A point mutation on the *C*-mannosylation site or the expression of punctin-1 in *C*-mannosylation-deficient cells (Lec35.1) decreased the secretion of punctin-1 (Wang et al, 2009). They proposed that *C*-mannosylation provides a quality control mechanism for punctin-1 secretion. Interestingly, a similar function of *C*-mannosylation was also proposed on MUC5AC and MUC5B (Perez-Vilar et al, 2004). The MUC5AC mutant (AXXW instead of WXXW) showed decreased secretion and accumulation in the cell. These data suggest that the function of *C*-mannosylation in secretion/folding is not limited to TSRs, but it has a global effect on *C*-mannosylated proteins. An alternative approach to investigate the function of *C*-mannosylation has been to identify its protein-binding partner by using a synthetic peptide with *C*-mannose (C-Man-WSPW) in cell culture (Ihara et al, 2010); heat shock cognate protein 70 (Hsc70) was found to have a higher affinity for the *C*-mannosylated peptide than for the non-mannosylated peptide. Furthermore, a downstream signaling pathway induced by Hsc70 was enhanced more in the presence of the *C*-mannosylated peptide. These results demonstrate the TSR-derived peptide with *C*-mannosylation may regulate cellular signaling through binding to Hsc70.

The other type of glycosylation on TSR is *O*-fucosylation. *O*-Fucosylation takes place in the ER by *O*-fucosyltransferase 2 (POFUT2), which adds a fucose by an α-linkage to a serine or threonine on the loop between the A- and B-strands (Luo et al, 2006a; Luo et al, 2006b; Tan et al, 2002). The fucose can be extended to Glc-β1,3-Fuc

disaccharide by β1,3-glucosyltransferase (β3GlcT) that is also present in the ER (Kozma et al, 2006; Sato et al, 2006). The biological function and importance of the *O*-fucosylation pathway are in Chapter 5.1.

Function of thrombospondin type 1 repeat

The functions of the TSR are diverse. TSR is involved in many important biological processes such as inflammation, neuronal development, angiogenesis, and parasite infection (Adams & Tucker, 2000; Lopez-Dee et al, 2011; Matuschewski et al, 2002; Tucker, 2004). The function of the TSR depends on the protein that contains the TSR, its cellular location, and the timing of the protein expression. In this section, I will only address the role of the TSR module plays in *Plasmodium* infection and anti-angiogenesis.

Plasmodium is the causative agent of malaria. During its life cycle, the sporozoite is the transmissible form of the parasite, spreading the disease from mosquitoes to humans. Sporozoites are released into human blood vessels from the salivary glands of mosquitoes, and later they infect liver cells. The thrombospondin-related anonymous protein (TRAP) is a transmembrane protein with a TSR domain in the cytosol. The TSR module is necessary for the entry of sporozoites into the hepatocytes; replacement of the residues in the WR layers (Fig. 5.4.2) with alanine in the TSR of TRAP resulted in non-invasive sporozoites (Matuschewski et al, 2002). Glycosaminoglycan chains (GAGs) on the host cell surface have been reported to be the receptors for invasion of *Plasmodium* sporozoites (Frevert et al, 1996). GAG is a sulfated molecule with a negative charge on the surface, which binds to the positively charged residues on the N-terminal region of TSR (Tossavainen et al, 2006). The TSR presumably facilitates invasion by recruiting GAGs *via* a charged interaction, and this interaction allows sporozoites to invade the host cell (Matuschewski et al, 2002).

TSP-1 was the first natural protein with anti-angiogenic properties that can restrain tumor growth (Good et al, 1990). The domains responsible for anti-angiogenic activity are predominantly located in the TSRs (Fig. 5.4.6). Using synthetic peptides, WSHWSPW, VTGC, and GVITRIR motifs were identified to have the ability to bind vascular endothelial growth factor (VEGF) (Lawler & Detmar, 2004). The binding of TSP to VEGF can mediate the uptake and clearance of VEGF, and therefore, inhibit angiogenesis (Greenaway et al, 2007). Another pathway of TSRs to inhibit angiogenesis is through binding to CD36 (Dawson et al, 1997), which is an integral

membrane receptor expressed on microvascular endothelial cells. The binding of TSRs to CD36 activates the caspase-3-like protease pathway, which then induces receptor-mediated apoptosis in activated microvascular endothelial cells (Jimenez et al, 2000). The CSVTCG motif on TSR (Fig. 5.4.6) was identified as the region for CD36 binding (Asch et al, 1992). Interestingly, this peptide harbors the *O*-fucosylation site on TSRs, but the effect of fucose or fucose-glucose on TSRs on the binding to CD36 remains to be clarified. The ability of TSP-1 to activate transforming growth factor β (TGFβ) may also contribute to arresting tumor growth and inhibiting tumor progression (Crawford et al, 1998). The motifs responsible for the binding and activation of TGFβ were identified on the TSR2 of TSP-1. As a first step in the activation of TGFβ, the sequence WSHWSPW on TSR binds to the VLAL sequence on TGFβ, which allows the RFK sequence on the N-terminal of TSR to activate TGFβ (Young & Murphy-Ullrich, 2004a; Young & Murphy-Ullrich, 2004b) (Fig. 5.4.6).
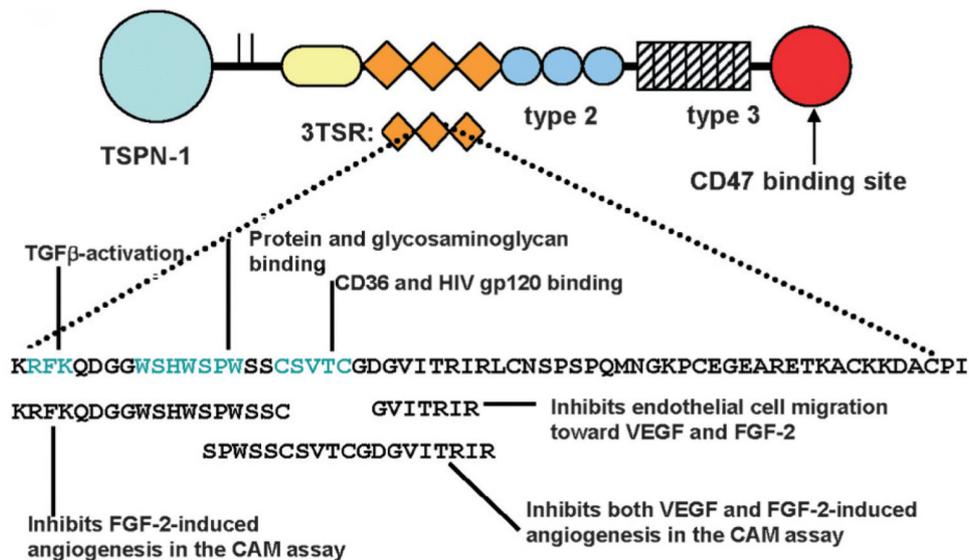
Adapted from (Kazerounian et al, 2008)

Figure 5.4.6 Domain structure of TSP-1

TSP-1 has multiple functional domains. The highlighted amino acid sequences of TSR2 are involved in the activation of TGFβ and inhibition of angiogenesis.

The anti-angiogenic and apoptosis-inducing properties of TSR make it an attractive drug candidate for cancer therapy. Because of the size and multiple biological functions of intact TSP-1, it is not suitable to use intact TSP-1 for therapy. Several peptides derived from the different motifs of TSRs have been developed as angiogenic inhibitors (Henkin & Volpert, 2011). ABT-510 is a peptide derived from the sequence GVITRIR in the TSR2 of TSP-1 (Reiher et al, 2002). It was the first compound shown to mimic the anti-angiogenic function of TSR in a clinical trial. ABT-

510 was shown to effectively block angiogenesis in a mouse model and to regress tumor lesions in dogs (Haviv et al, 2005). However, ABT-510 did not have a significant effect on phase II clinical trials as a monotherapy and is no longer in clinical development (Baker et al, 2008; Markovic et al, 2007).

Other proteins with TSR domains have been found to have anti-angiogenic activity. A disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) are a group of multi-domains proteins with the ability to cleave aggregating chondroitin sulfate proteoglycans (Nakamura et al, 2000). ADAMTS-1 inhibits angiogenesis through two mechanisms. It binds to the VEGF receptor, which prevents VEGF binding to the VEGF receptor, and it releases anti-angiogenic peptides from TSP-1 and -2 through its protease activity (Luque et al, 2003; Vazquez et al, 1999).

Failure of anti-angiogenic peptides derived from TSP-1

Tumor cells are able to evolve drug resistance to peptides that interrupt tumor angiogenesis. Recent investigations of peptides targeted in the VEGF pathway in preclinical studies suggested that although they inhibited primary tumors, the peptides could accelerate metastasis and decrease the overall survival rate in mice (Ebos et al, 2009; Paez-Ribes et al, 2009). ABT-510 derived from TSR of TSP-1 failed to show antitumor effects in phase II clinical trials (Baker et al, 2008; Markovic et al, 2007). To identify why these peptides have anti-angiogenic properties, but do not show antitumor activity, the role of TSP-1 in tumor progression has been re-evaluated using a prostate cancer model (Firlej et al, 2011). In this interesting work, they first found that TSP-1 strongly stimulated migration of cultured cancer cells, and the CD36 receptor was necessary for migration to occur. In addition, the anti-angiogenic property of TSP-1 was found to increase hypoxia of the tumor cells, which, in turn, enhanced TSP-1 production and secretion. The cycle of hypoxia and increased TSP-1 expression in prostate tumors promoted cell migration. They proposed that the ability to increase hypoxia and to trigger cell migration were inherently linked, and that this link was shared by the endogenous and pharmacologic inhibitors of angiogenesis (Firlej et al, 2011).

In summary, the diverse functions of TSR-containing proteins in biological activities are related to the TSR modules. An in-depth understanding of TSR, including co- and post-translational modifications, will aid in identifying the functions and roles of TSR-containing proteins.

6. Results

6.1 Expression and purification of rat F-spondin TSR4

Introduction

O-Fucosylation is a relatively rare protein glycosylation modification that is only found on EGF-like repeats, TSRs, and PMP-C. The EGF repeat is a protein motif about 40 amino acids in length with six conserved disulfide bonds. These cysteines form a specific disulfide bond pattern: Cys_1-Cys_3, Cys_2-Cys_4, and Cys_5-Cys_6 (Campbell & Bork, 1993). TSR, which is ~60-amino acid-long in length, also contains six conserved cysteines with three disulfide bonds. The disulfide pattern in F-spondin TSR4 is Cys_1-Cys_4, Cys_2-Cys_5, and Cys_3-Cys_6. The fucosyltransferases that transfer fucose from GDP-fucose to EGF repeats and TSRs have been identified. POFUT1 is responsible for the fucosylation of EGF repeats, and POFUT2 specifically recognizes TSRs. Although there are similarities between the two O-fucosylation pathways, POFUT1 and POFUT2 have high specificity for their own substrate and do not cross-react (Luo et al, 2006b). One of the aims of my thesis was to characterize the mechanism of human POFUT2 and elucidate the substrate specificity of POFUT2.

To achieve this goal, a correctly folded TSR is required in the activity assay to determine the activities of wild-type and mutant POFUT2. Previously in our laboratory, TSR4 from rat F-spondin has been recombinantly expressed in HEK293T cells with a C-terminal Myc-His$_6$-tag (Gonzalez de Peredo et al, 2002). The TSR4 was purified and used to characterize C-mannosylation on tryptophan. The same protein module has also been expressed and purified in an $E.\ coli$ expression system to characterize the activity of β1,3-glucosyltransferase (Kozma et al, 2006). The rat and human F-spondin TSR4 modules are highly conserved. Of 53 amino acids, only a single amino acid is changed in the C strand: human TSR4 has an alanine and rat TSR4 has a valine (Fig. 6.1.1). The position of this amino acid is not on the consensus sequence of O-fucosylation, and we do not expect the difference in a single amino acid will affect the specificity of POFUT2.

Human F-spondin TSR4   PCLLSPWSEWSDCSVTCGKGMRTRQRMLKSLAELGDCNEDLEQAEKCMLPECP
    Rat F-spondin TSR4   PCLLSPWSEWSDCSVTCGKGMRTRQRMLKSLAELGDCNEDLEQVEKCMLPECP

Figure 6.1.1 Sequence alignment of human and rat F-spondin TSR4

The amino acid difference between human and rat TSR4 is highlighted in blue.

Cysteine scrambling and multimeric TSR4 concatemer formation occurred when rat TSR4 was expressed in *E. coli* (Kozma et al, 2006). To circumvent these problems, we expressed the protein at a lower temperature (20 ºC instead of 37 ºC) by using an autoinduction expression system and BL21 Rossetta 2 cells, which have been engineered to enhance the expression of eukaryotic proteins containing rare codons in *E. coli*. The His$_6$-tag on the C-terminus was moved to N-terminus following a human rhinovirus 3C cleavage site. We established a three-step protocol to purify His$_6$-3C-TSR4 from *E. coli*. Moreover, we were also interested in determining the fucosylation efficiency of mini-TSR (a minimal POFUT2 substrate, Chapter 6.2) and TSR23*. TSR23* consists of the TSR2 and TSR3* modules from human thrombospondin-1 with a T to A mutation on the *O*-fucosylation site of TSR3. The results of preliminary purification of GST-His$_6$-3C-mini-TSR and GST-His$_6$-3C-TSR23* and initial fucosylation efficiency will be described in this chapter.

Materials and methods

Expression of rat TSR4 in *E. coli*

Rat F-spondin TSR4 containing an N-terminal His$_6$-3C-tag was cloned into the pOPINF vector (Berrow et al., 2007) and was used to transfect *E. coli* (Rosetta 2) cells. The freshly transfected *E. coli* cells were grown overnight in 20 ml of GS96 medium containing 1% glucose. The overnight culture was diluted 1:100 into 0.5 l of auto-induction medium (Studier, 2005) containing 34 μg/ml chloramphenicol and 50 μg/ml carbenicillin. Cells were grown at 37 ºC at 225 rpm for 4 h, then the temperature was decreased to 20 ºC, and the cells were grown for another 20 h. Cells were harvested by centrifugation at 6500 g for 30 min. The cell pellet from 1 l of culture was resuspended in 36 ml of lysis buffer [50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.2% Tween20, and complete protease inhibitor (Roche)], and stored at -80 ºC for later purification. In total, 48 L of *E. coli* culture was prepared for large-scale TSR4 production.

Purification of His$_6$-3C TSR4

For purification, 8 L of the cell pellet was resuspended in 400 ml of lysis buffer and lysed by a high-pressure cell disruptor. The lysate was cleared by centrifugation at 30,000 *g* for 30 min at 4 ºC. The cleared lysate was decanted into a fresh tube and incubated with Ni-NTA agarose (Qiagen) for 2 h at 4 ºC with gentle rolling. The resin and lysate mixture was poured into a column (Econo-Column, Bio-Rad), and the

column was connected to an ÄKTA purifier. The unbound fraction was collected at a flow rate of 1 ml/min. The resin was washed five times with the gel volume of lysis buffer, and TSR4 was eluted in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, and 500 mM of imidazole. The eluted fraction was loaded onto a Superdex-75 16/60 PG column (GE Healthcare) in 20 mM Tris-HCl pH 7.4, 150 mM NaCl. Fractions containing TSR4 were pooled and buffer exchanged into 20 mM Tris-HCl pH 7.4, 10 mM NaCl by using a HiPrep 26/10 desalting column. The sample was then loaded onto a MonoQ 5/50 column (GE Healthcare), and TSR4 was eluted using a linear gradient of 0–500 mM NaCl over 50 min. TSR4 was concentrated using an Amicon-4 centrifugal filter (3000 NMWL) to 3.6 mg/ml or 7 mg/ml and stored at -80 ºC for later applications.

In vitro fucosylation assay

The description of the method is included in Chapter 6.2.

Mass spectrometry (MS) analysis of mini-TSR and TSR23*

The description of the method is included in Chapter 6.2.

Results

Expression and purification of His$_6$-TSR4 by Ni-NTA affinity chromatography and size exclusion chromatography

For the _in vitro_ POFUT2 assay, one of the key steps is to prepare properly folded, pure TSR. The methods for studying the expression and purification of C-terminally tagged rat F-spondin TSR4 from _E. coli_ has been established previously (Kozma et al, 2006). Using the structures of POFUT2 and TSR2-TSR3 from TSP-1 (Tan et al, 2002), we modeled the POFUT2-TSR complex (Fig. 6.2.5). Based on the model, we predicted that the C-terminal Myc-His$_6$ tag of TSR4 might be located in the inter-domain cleft, which would interfere with the co-crystallization of POFUT2 with GDP-fucose. Therefore, we designed a new construct with the His$_6$-3C tag on the N-terminus and used an auto-induction system for protein expression.

The pellet was thawed from -80 ºC in lysis buffer and lysed by mechanical force. The clarified lysates were incubated with Ni-NTA agarose, and the elution fraction from the Ni-NTA agarose resin was loaded onto a SDS-PAGE gel to check the purity of the protein. Staining of the gel indicated that the sample contained a moiety of the predicted size (lane 1 in Fig. 6.1.2B), but the sample contained contaminants. The

size of most contaminants was larger than TSR4, so we could further purify the protein by size-exclusion chromatography (SEC).

From the UV280 absorbance shown on the chromatogram (Fig. 6.1.2A), we found two major absorbent peaks, the first at 40 ml, which is the void volume of the column, and the second at 80 ml that corresponds to the size of monomeric TSR4. The select fractions from the SEC column were analyzed by SDS-PAGE; we observed that the majority of the contaminants were separated from TSR4 (Fig. 6.1.2B). Proteins larger than 50 kDa were fractionated into #8–12. The smearing signal in the fraction #8 at a high-molecular weight could indicate protein aggregation. In the fractions #13–#18, the molecular weight of contaminating proteins decreased with the elution time. The signal with a size close to His_6-3C-TSR4 was present in the fraction #19 at ~10 kDa under non-reducing conditions. However, we observed another protein that ran higher than TSR4 in the non-reducing gel. In the reducing gel, however, these two proteins could not be separated, and they ran as a single band.
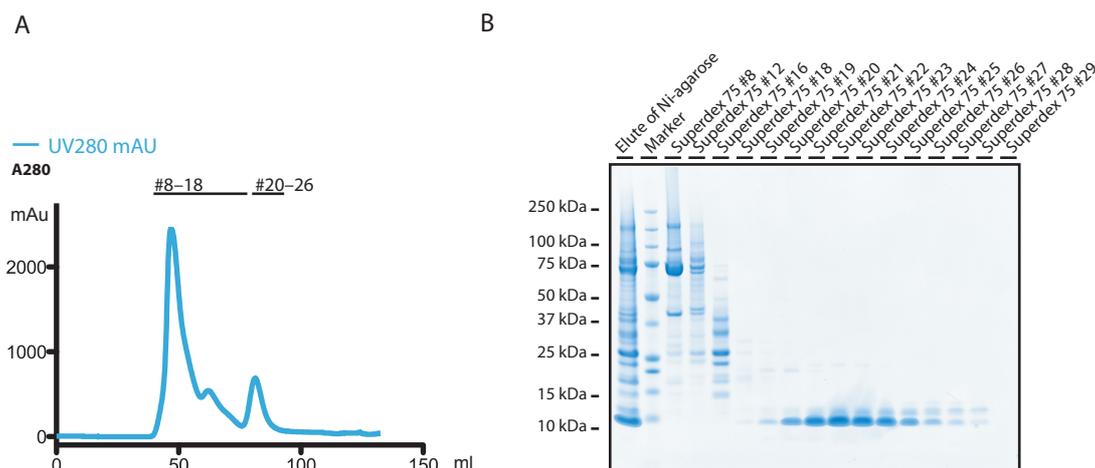

Figure 6.1.2 Purification of His_6-3C-TSR4 by SEC

After Ni-NTA affinity purification, the sample was further purified by size-exclusion chromatography to isolate TSR4. The chromatogram in (A) shows that the proteins were separated into two main peaks, which are shown in the non-reducing SDS-PAGE gel in (B). Most contaminant proteins were separated from TSR4 by the SEC column; only a small fraction of other proteins was eluted with TSR4.

To confirm that the major band on the gel consists of TSR4 and to identify the contaminant protein, we cut these two bands from the gel and performed in-gel trypsin digestion for protein identification by MS. The MASCOT search indicated that the major protein was rat TSR4 from F-spondin; the contaminant protein was 30S ribosomal protein S15. The predicted molecular weight of the S15 is 10,268 Da, which is very close to the molecular weight of His_6-3C-TSR4 (8194 Da).

S15 was removed by a Mono Q anion exchange column

The similarity in the sizes of S15 and His$_6$-TSR4 makes these two proteins inseparable by SEC. However, the isoelectric point (pI) of S15 (10.4) and His$_6$-TSR4 (5.8) were significantly different; therefore, the two proteins could be separated by ion-exchange chromatography. The fractions #20–#26 from the SEC column were pooled and buffer exchanged into 20 mM Tris-HCl pH 7.4, 10 mM NaCl, by using a HiPrep™ 26/10 desalting column before loading onto the MonoQ column.

MonoQ anion exchange chromatography was used to remove *E. coli* S15 (upper band in Fig. 6.1.2B lane Superdex 75 #20–#26) from TSR4. Since S15 is a basic protein, it should not be retained on the anion exchange column under neutral conditions. We injected the mixed sample onto the MonoQ anion exchange column under neutral, low salt conditions. On the chromatogram (Fig. 6.1.3A), we observed a peak by UV280 absorbance before the gradient was started, indicating the presence of a protein that did not bind to the MonoQ column. When we started the gradient, a peak was observed at 40 ml by using the gradient of 0% Buffer B to 50% Buffer B within 60 min. The purity of the sample separated by the MonoQ anion exchange column was checked on a SDS-PAGE gel. After staining the gel, we determined that the flow-through fraction contained S15 (Fig. 6.1.3B). On the other hand, His$_6$-TSR4 was retained on the column in Buffer A, and eluted later with increasing conductivity. The major peak from fraction #13 to #18 contained TSR4. On the gel, the protein from these fractions was a single band with >99% purity at the predicted size for TSR4. Therefore, fractions #13–#18 were pooled and stored at -80 ºC.
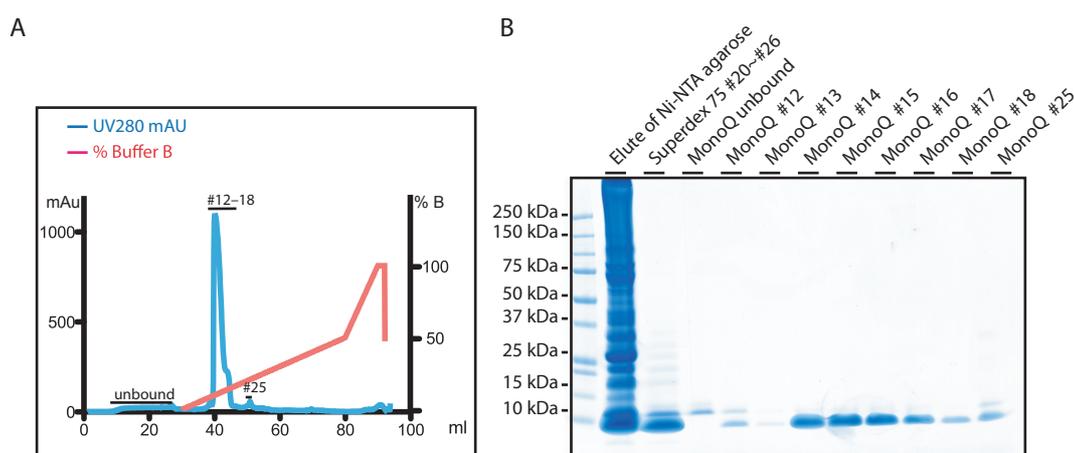

Figure 6.1.3 Purification of His$_6$-TSR4 by MonoQ anion exchange chromatography

The pooled samples after size-exclusion gel filtration were further purified on a MonoQ anion exchange column. From the absorbance at UV280 (A), we observed two major fractions: the unbound fraction before starting the gradient and major fraction from #12

to #18. The SDS-PAGE gel in (B) confirms that the unbound fraction is S15 and that the main peak is TSR4.

Mass spectrometry analysis of purified TSR4

To confirm that the protein we purified was TSR4 and to investigate the fucosylation of N-terminally tagged TSR4, we used reverse-phase HPLC coupled with MS to analyze the purified TSR4 sample. In addition, we analyzed the sample after *in vitro* fucosylation. MS results revealed that the molecular weight of our sample was 8194 Da (Fig. 6.1.4A), which is 152 Da lower than the expected molecular weight of TSR4 (8349 Da). We hypothesized that 146-Da difference was due to N-terminal methionine excision, which has been observed in the cytosolic proteins of *E. coli* (Waller, 1963). N-Terminal methionine excision is catalyzed by a methionyl-aminopeptidase that removes the methionine if the following amino acid is glycine or alanine (Hirel et al, 1989). In our TSR4 construct, alanine is the amino acid following methionine, which may lead to excision of the methionine in the purified TSR4. The other 6-Da difference likely comes from the formation of three disulfide bonds, which may indicate that TSR4 folds properly in *E. coli*.

A

B
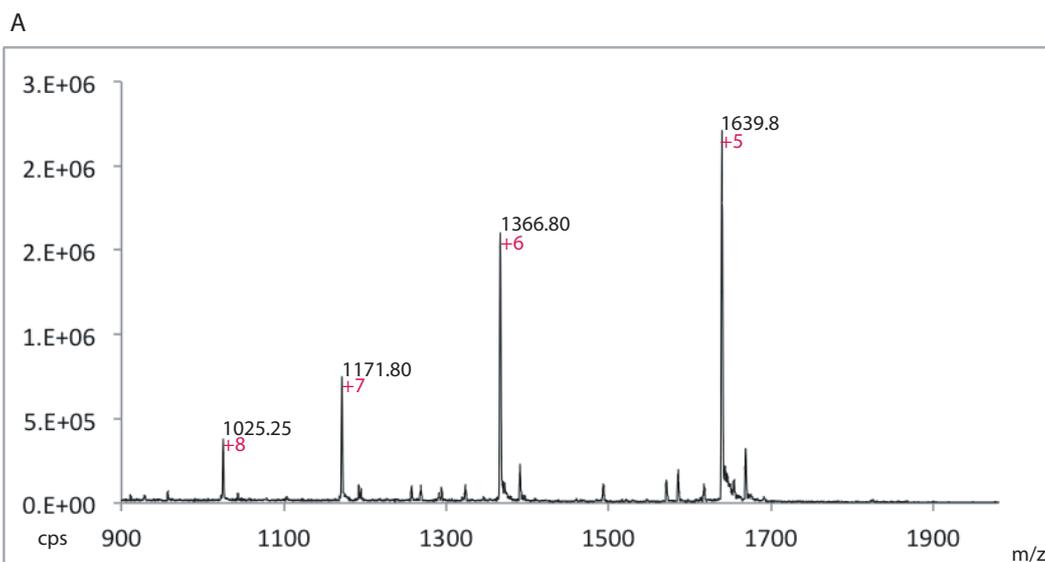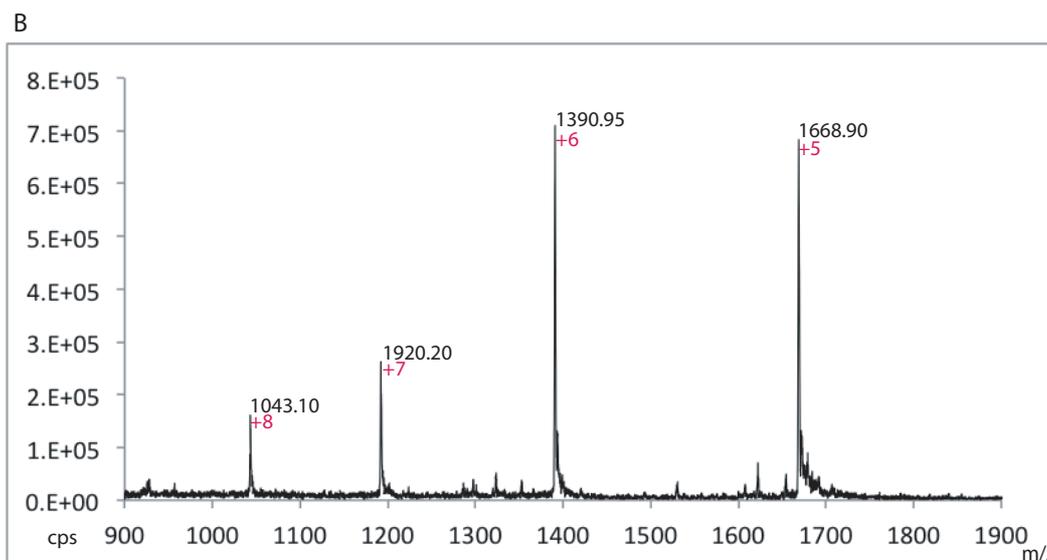
Figure 6.1.4 ESI-MS analysis of purified rat F-spondin TSR4

The average mass of purified TSR4 (A) was 8194.1 Da, which corresponds to the molecular weight of TSR4 after N-terminal methionine excision and loss of six protons following the formation of three disulfide bonds. After *in vitro* fucosylation assay, the molecular weight of TSR4 increased by 146 Da (B), indicating that TSR4 can be fucosylated by POFUT2.

POFUT2 can only transfer fucose to correctly folded TSR4 (Luo et al, 2006b). We incubated purified His_6-TSR4 with POFUT2 and GDP-fucose and analyzed the product by ESI-MS. The molecular weight increased by 146 Da, compared to the molecular weight of TSR4 before fucosylation, which corresponds to the attachment of one fucose on His_6-TSR4. The successful *in vitro* fucosylation of purified His_6-TSR4 with an excess of GDP-fucose indicates that the His_6-TSR4 expressed in *E. coli* is folded correctly.

The total amount of pure TSR4 isolated from 8 l of *E. coli* culture with two steps of purification was ~2 mg. In total, ~12 mg of His_6-TSR4 was prepared from six batches of 8-L *E. coli* culture. His_6-TSR4s from different batches were pooled and concentrated to 3 mg/ml or 7 mg/ml. The concentrated samples were stored at -80 ºC for use in the activity assay of wild-type and mutant POFUT2 (Chapter 6.2).

Expression and purification of GST-His_6-WT-TSR, GST-His_6-mini-TSR, and GST-His_6-TSR23*

In Chapter 6.2, we discuss the *in vivo* and *in vitro* fucosylation of mini-TSR. However, our results suggested that mini-TSR can be fucosylated by POFUT2, but they do not explain why the fucosylation rate on mini-TSR is lower than that on wild-type TSR4 (Fig. 6.2.6). These data were determined from the small-scale expression of mini-TSR; we had not optimized the large-scale purification of mini-TSR in *E. coli*. In this

section, we describe the ongoing purification and characterization of mini-TSR. We also expressed and optimized the purification of TSR23*, which has thrombospondin repeats 2 and 3 from TSP-1 with a T to A mutation at the fucosylation site of TSR3. One of our goals was to investigate the *in vitro* fucosylation of these two modules by POFUT2.

TSR23*, wild-type TSR4 (WT-TSR4), and mini-TSR were digested by HRV 3C protease, and incubated with Ni-NTA Agarose to remove the HRV 3C protease, the GST-His$_6$ tag, and any non-specific contaminant proteins. As shown in Figure 6.1.5A, the 3C digestion efficiently removed the GST-His$_6$ tag from TSR23*, as demonstrated by the evident band shift from ~37 kDa to 26 kDa. The band at ~26 kDa corresponded to the GST-His$_6$ tag, and the band at 16 kDa corresponded to TSR23*. Incubation of the digested TSR23* with Ni-NTA resin efficiently removed the N-terminal tag and 3C protease. The same procedure was applied to mini-TSR and WT-TSR4. Similar to TSR23*, we observed that WT-TSR4 appeared at ~10 kDa after cleavage and Ni-NTA incubation. However, there were several problems associated with the purification of mini-TSR. First, it was not completely digested by 3C protease; the tag was not cleaved from ~50% of the protein, despite an excess of protease being supplied for the digestion. Second, the digested mini-TSR still bound the N-terminal tag. The digested mini-TSR co-eluted with the 26 kDa N-terminal tag when we eluted the protein from the Ni-NTA resin (Fig. 6.1.5B).
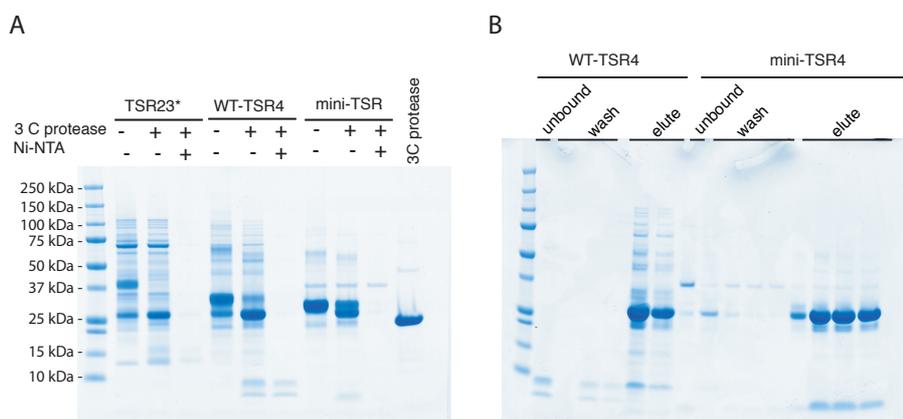

Figure 6.1.5 Tag removal of WT-, TSR23*, and mini-TSR

The proteins were digested by 3C protease and digested samples were incubated with Ni-NTA Agarose to remove the tag and protease, and then analyzed by SDS-PAGE (A). The proteins were eluted from Ni-NTA Agarose, and then analyzed by SDS-PAGE (B).

Although there was significant sample loss in the purification of mini-TSR, we could still observe a blurry signal corresponding to mini-TSR after concentrating the protein (Fig. 6.1.6).
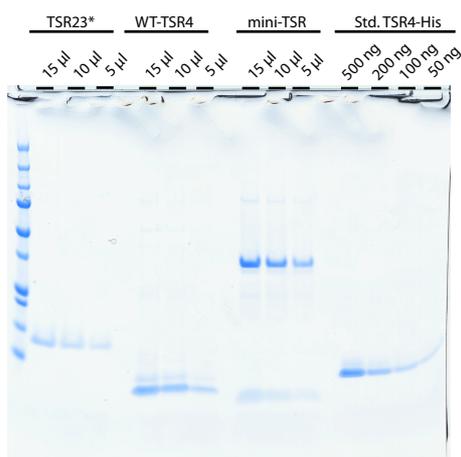
Figure 6.1.6 Protein quantification

Different volumes of TSR23*, WT-TSR4, and mini-TSR were loaded onto the gel for protein quantification by comparing them to the standard TSR4 of known concentration.

The concentrations of TSR23*, WT-TSR4, and mini-TSR were determined by comparing them to standard TSR4. The concentration of TSR23*, WT-TSR4, and mini-TSR were 10 ng/μl, 40 ng/μl, and 10 ng/μl, respectively. Overall, the final yields were 15 μg of TSR23*, 60 μg of WT-TSR4, and 0.5 μg of mini-TSR from 250 ml of *E. coli* culture.

In vitro activity assay of POFUT2 on mini-TSR and TSR23*

The same concentration (4 ng/μl) of WT-TSR4, TSR23*, and mini-TSR was incubated with an excess of GDP-fucose (200 μM) and 2 ng of POFUT2 to test the efficiency of fucosylation. The reactions were quenched by acidification at different time points and analyzed by LC-MS with the multiple-ion monitoring mode.

We observed that the fucosylated WT-TSR4 (red line) eluted earlier than the unfucosylated WT-TSR4 (blue line) did (Fig. 6.1.7). About 40% of the WT-TSR was fucosylated at 7.5 min. The ratio of fucosylated WT-TSR increased with the reaction time. However, after 90 min, we still observed ~3% of unfucosylated WT-TSR. The peak shape of unfucosylated WT-TSR4 at 90 min showed two peaks; we hypothesized that one of the peaks corresponded to misfolded TSR4, which could not be fucosylated.
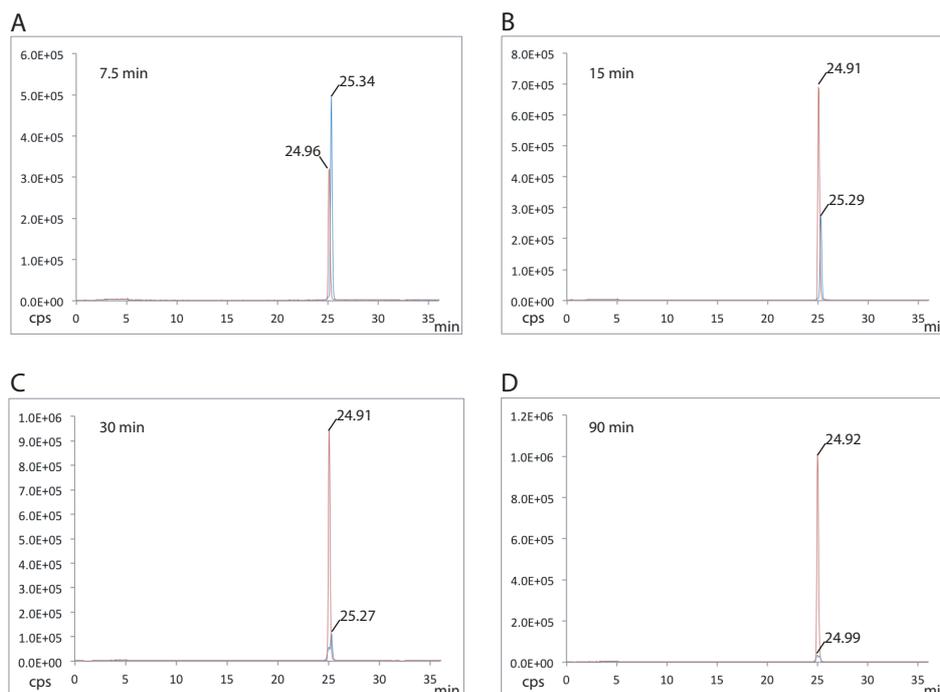
Figure 6.1.7 *In vitro* fucosylation assay of wild-type TSR4

The reaction mixture was quenched at different time points and analyzed by LC-MS. We monitored the non-fucosylated TSR4 at m/z = 1622 (blue line) and fucosylated TSR4 at m/z = 1585 (red).

When we analyzed the sample from mini-TSR, we observed that mini-TSR was eluted in three peaks. The molecular weight of each of these three peaks was 3268 Da, which corresponds to mini-TSR with two disulfide bonds (Fig. 6.1.8A). The ratio of the three peaks was 25.3%, 55%, and 18.7%.

The ratio of fucosylated mini-TSR (red line) increased with the reaction time. At 7.5 min, the overall fucosylated mini-TSR was ~1.3% of the total mini-TSR (sum of three peaks), and increased to 2.6%, 3.8%, and 15% at 15, 30, and 90 min, respectively. Interestingly, by comparing the three peaks of the non-fucosylated form (blue line), the fucosylated mini-TSR gives only one peak at 19.8 min, which implies that only one of the three non-fucosylated peaks can be fucosylated. To identify a non-fucosylated peak that can be recognized and fucosylated by POFUT2, we repeated the mini-TSR assay with 5-fold more POFUT2 and extended the reaction time to 120 min (Fig. 6.1.8E). Under these conditions, we found that the non-fucosylated peak at 20.17 min disappeared, and that the ratio of the fucosylated mini-TSR was 55%, which is the ratio of the second non-fucosylated peak. These data suggest that under these reaction conditions, all the mini-TSR at the second peak (20.17 min) was fucosylated by POFUT2, while the mini-TSR corresponding to the other two peaks could not be fucosylated.

58
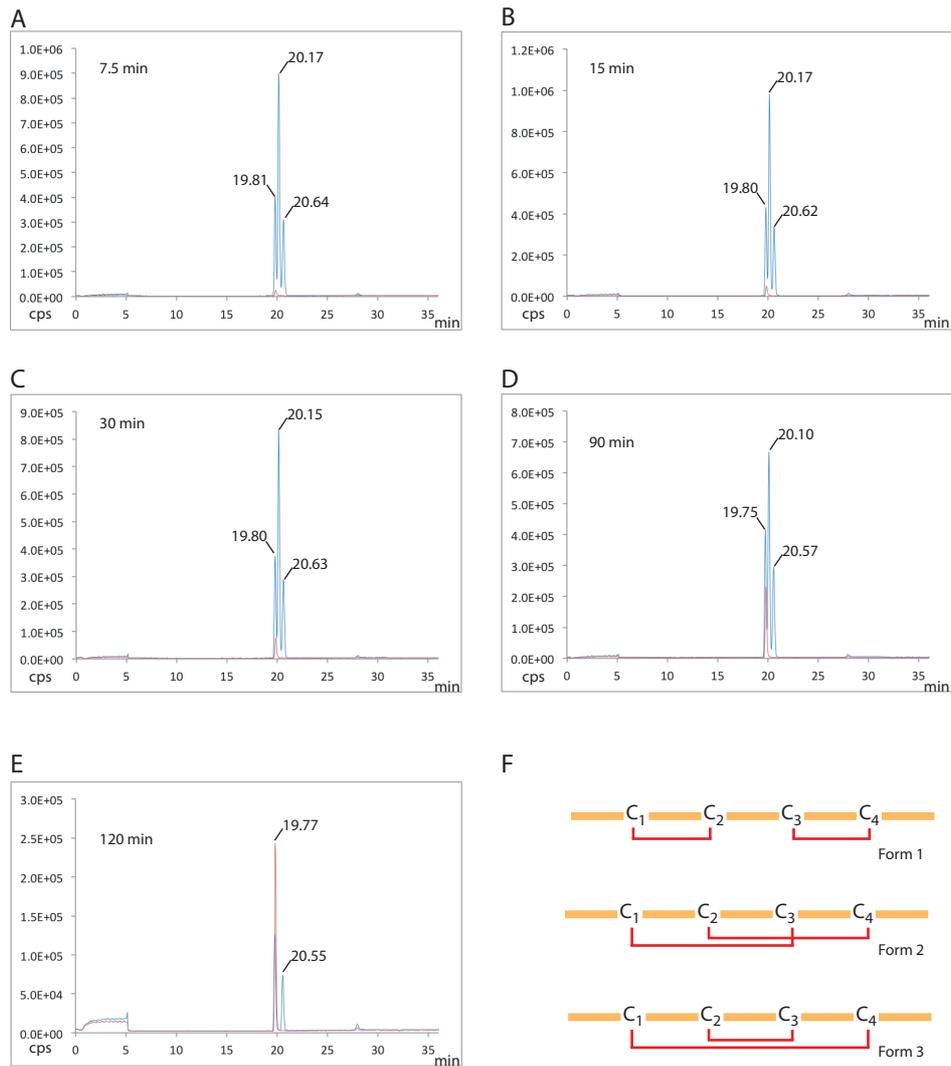
Figure 6.1.8 *In vitro* fucosylation of mini-TSR

The reaction mixture was quenched at different time points and analyzed by LC-MS. We monitored the non-fucosylated mini-TSR at m/z = 1090 (blue line) and fucosylated mini-TSR4 at m/z = 1138 (red). Mini-TSR with three different disulfide bond patterns are illustrated in F. Form 2 is the correctly folded form that is present in WT-TSR.

Based on the results presented in Figure 6.1.8A-E, we propose that the mini-TSR expressed in *E. coli* may have three different disulfide bond patterns. Mini-TSR has four cysteines that can form two pairs of disulfide bonds (Fig. 6.1.8F). In form 1, C1 bonds to C2, and C3 bonds to C4, thereby creating a long loop between C2 and C3. Form 2 is the correct disulfide pattern that can maintain the three-strand structure in TSR (Tan et al, 2002). The disulfide pattern in form 3 may create a big loop with two strands. According to our model (Fig. 6.2.5), only form 2 can be recognized and fucosylated by POFUT2. Therefore, we propose that form 2 is the peak that was eluted at 20.17 min, but forms 1 and 3 have not been assigned yet.

The same reaction conditions that were used for WT-TSR4 and mini-TSR were used for TSR23*. In this elution gradient, fucosylated TSR23* could not be separated from

the non-fucosylated form. Although the fucosylated TSR23* and non-fucosylated TSR23* were inseparable, they could still be quantified individually by their molecular weight determined by mass spectrometry. At 7.5 min, 60% of the TSR23* was fucosylated, which increased to 80%, 97%, and 100% after 15, 30, and 90 min of reaction, respectively (Fig. 6.1.9).
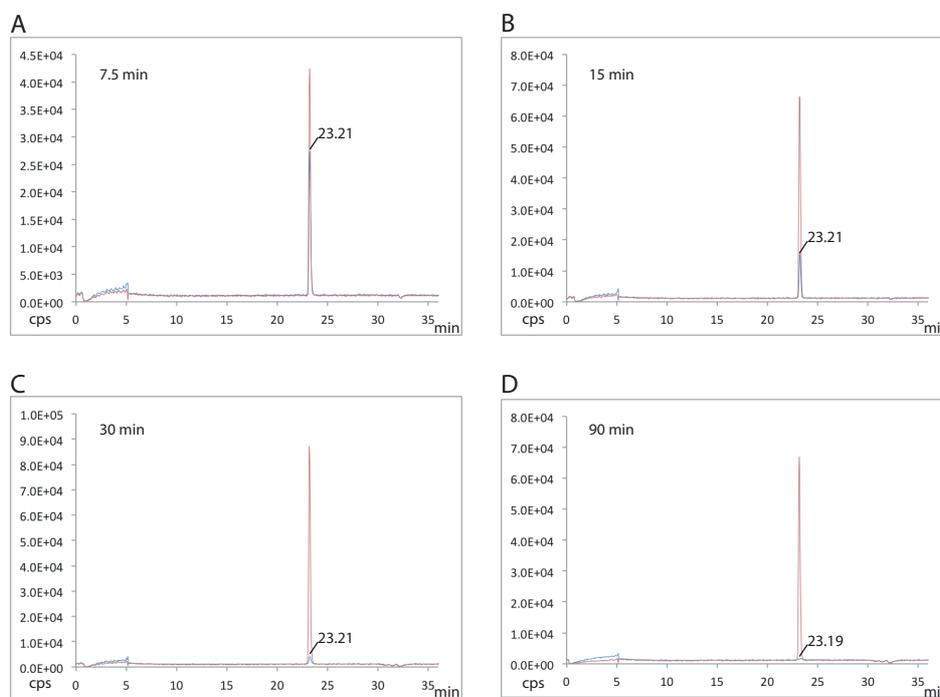

Figure 6.1.9 *In vitro* fucosylation of TSR23*

The reaction mixture was quenched at different time points and analyzed by LC-MS. We monitored non-fucosylated TSR23* at m/z = 1784 (blue) and fucosylated TSR4 at m/z = 1805 (red).

Interestingly, if we compare the ratio of fucosylated TSR at 7.5 min between the three substrates, TSR23* has the highest fucosylation rate, where 80% of the substrate was turned into product. About 40% of WT-TSR4 was fucosylated at 7.5 min, while only 1.3% of mini-TSR (or 2.1% of the second peak) was fucosylated. The difference in the fucosylation rate of POFUT2 on different substrates under the same reaction conditions indicates that POFUT2 may catalyze the reaction with different efficiencies.

Discussion

In this chapter, we described the expression and purification of different TSR constructs in *E. coli* for different purposes. We first described the large-scale expression and purification of His$_6$-3C-TSR4 from rat F-spondin. We purified ~12 mg of His$_6$-3C-TSR4 from 48 l of *E. coli* culture for kinetic assay studies (Chapter 6.2), activity assays of POFUT2 (Chapter 6.2), and co-crystallization experiments with

POFUT2 and GDP-fucose (that did not crystallize). We also described the purification of GST-His$_6$-WT-TSR, GST-His$_6$-mini-TSR, and GST-His$_6$-TSR23*, which were the preliminary experiments to establish purification procedures and test the fucosylation efficiency of POFUT2. Although the expression and purification of His$_6$-3C-TSR4 was tedious, the procedure was very reproducible, and TSR4 was properly folded. On the other hand, the expression and purification of GST-His$_6$-TSR23* and GST-His$_6$-mini-TSR still need optimization. The discussion below will focus on mini-TSR and TSR23*.

Disulfide patterns of the mini-TSR expressed in *E. coli*

The characteristic feature on the TSR module is the CWR-layered core structure (Tan et al, 2002), where the side chains of cysteine, tryptophan, and arginine stack in layers (Fig. 6.1.10).
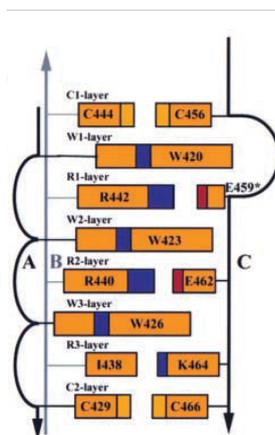
Adapted from (Tan et al, 2002)

Figure 6.1.10 CWR-layered core structure of the TSR domain

A schematic drawing of the CWR-layered structure; each layer and layer-forming residues are labeled. Three tryptophans, namely, W420, W423, and W426, (W layers) sandwich with R422, R330, and I438 (R layers). Two disulfide bonds cap the layers.

The alternate stacking of the planar cationic guanidinium group of arginine and the aromatic side chains of tryptophan forms multiple cation-π interactions. This unique layered structure provides critical stabilization energy for TSR to maintain its compact antiparallel structure. However, in mini-TSR, there is only one tryptophan on the A strand and two arginines on the B strand, which could significantly reduce the stabilization energy of the structure, and lead to a less stable structure instead of a compact TSR domain. The pairing of the disulfide bonds relies on the layers bringing the cysteine residues into close proximity. Consequently, the scrambled structure of the mini-TSR could result in the incorrect pairing of disulfide bonds. Therefore, we propose the three peaks shown in Figure 6.1.8 correspond to mini-TSR with different

disulfide bond patterns. However, we have not determined the isoform of each peak. Since POFUT2 can recognize only properly folded TSR, we hypothesize that the second peak is form 2 in Figure 6.1.8F. Since we can only purify 0.5 μg of mini-TSR from 1 l of *E. coli* culture, we will likely use a highly sensitive method such as MS to identify the isoform corresponding to each peak.

Analysis of peptides digested by a protease is a sensitive approach to identify proteins. Ideally, a protease that can hydrolyze mini-TSR under native conditions and with characteristic fragments to distinguish the disulfide bond patterns will be useful. Thermolysin is a metalloprotease from the gram-positive bacteria *Bacillus thermoproteolyticus* that can cleave proteins at the N-terminus of hydrophobic residues such as phenylalanine, leucine, isoleucine, methionine, and alanine.

Mini-TSR cleavage by thermolysin can generate five fragments (6.1.11). Different forms of mini-TSR would have characteristic fragments for identification. Therefore, the disulfide bond pattern in each elution peak can be possibly identified by peptide mapping with thermolysin digestion.
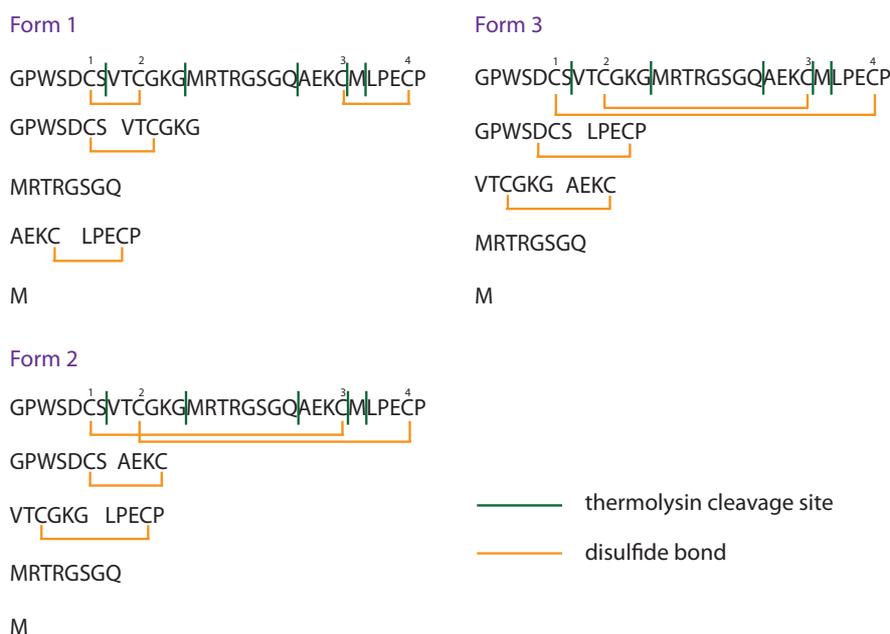
**Figure 6.1.11 Thermolysin cleavage of mini-TSR**

The predicted peptides of mini-TSR with three disulfide bond patterns after cleavage by thermolysin

Fucosylation on TSR23*

TSRs are usually present in proteins as consecutive modules. For example, human thrombospondin-1 has three consecutive TSR modules, followed by three EGF-like repeats, and a thrombospondin type-3 repeat (Fig. 6.1.12A). Most TSRs with a

fucosylation motif in the consecutive repeats can be *O*-fucosylated (Gonzalez de Peredo et al, 2002). In our POFUT2-TSR model (Fig. 6.1.12B), POFUT2 only binds to TSR2 around the fucosylation site on the AB loop; the other part of TSR2 is solvent-exposed. TSR3, on the other hand, contacts POFUT2 through part of the A strand and BC loop, and leaves the fucosylation site on the AB loop exposed. According to the model, we predict that a TSR with two modules may be a better substrate because of the larger binding surface with POFUT2, which may increase the affinity between the two molecules.
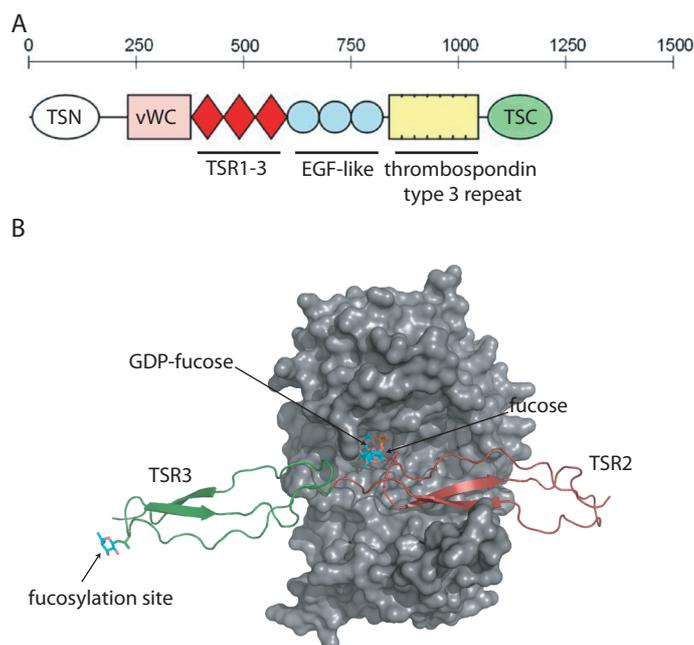
(A) is modified from (Tucker, 2004)

Figure 6.1.12 The model of POFUT2 in complex with GDP-fucose and TSR23 of TSP-1

Domain structure of TSP-1 is shown in (A). TSN, thrombospondin N-terminal domain; vWC, von Willebrand factor type A domain; TSC, thrombospondin C-terminal domain. Amino acid numbers are indicated above. (B)The structure of POFUT2 is shown in gray. TSR23 from TSP-1 (PDB:1LSL) is shown as a cartoon model. TSR2 is colored in red, and TSR3 is colored in green. GDP-fucose and fucose are shown as sticks.

The preliminary results of fucosylation on TSR23* shown in Figure 6.1.9 suggest that it could be a better substrate than WT-TSR4 (Fig. 6.1.7). At the same time point, the ratio of fucosylated TSR23* is always higher than WT-TSR4. However, the experiment still needs to be optimized to obtain conclusive results.

WT-TSR4 was from rat F-spondin, while TSR23* was cloned from human TSP-1. Although the structures of the AB loop between TSR2 from TSP-1 and TSR4 from F-spondin are very similar, the disulfide bond pattern of the two proteins differ. Furthermore, the additional jar handle on TSRs from TSR-1 (Tan et al, 2002) that lacks in TSR4 from F-spondin may cause the difference in binding to POFUT2. An

appropriate way to compare the efficiency of fucosylation will be to express one-module and two-module TSRs from the same protein such as TSR2 and TSR23* from TSP-1. Therefore, differences in the fucosylation efficiency between the proteins would be solely due to the number of TSR modules.

We compared the fucosylation efficiency of POFUT2 on different TSR modules. Nevertheless, the catalytic reaction consists of three separate steps; substrate association, catalysis, and product dissociation. We did not specifically determine how the additional TSR module affects fucosylation efficiency. Other approaches such as isothermal titration calorimetry (ITC) that can measure the association rate (K_a), enthalpy (ΔH) and binding stoichiometry or surface plasma resonance that is used to determine k_a and dissociation rate (k_d) may help in clarifying how the additional TSR module affecting POFUT2-catalyzed fucosylation efficiency.

THE
EMBO
JOURNAL

Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and *O*-fucosylation

Chun-I Chen, Jeremy J Keusch, Dominique Klein, Daniel Hess, Jan Hofsteenge and Heinz Gut*

Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

Protein *O*-fucosylation is a post-translational modification found on serine/threonine residues of thrombospondin type 1 repeats (TSR). The fucose transfer is catalysed by the enzyme protein *O*-fucosyltransferase 2 (POFUT2) and >40 human proteins contain the TSR consensus sequence for POFUT2-dependent fucosylation. To better understand *O*-fucosylation on TSR, we carried out a structural and functional analysis of human POFUT2 and its TSR substrate. Crystal structures of POFUT2 reveal a variation of the classical GT-B fold and identify sugar donor and TSR acceptor binding sites. Structural findings are correlated with steady-state kinetic measurements of wild-type and mutant POFUT2 and TSR and give insight into the catalytic mechanism and substrate specificity. By using an artificial mini-TSR substrate, we show that specificity is not primarily encoded in the TSR protein sequence but rather in the unusual 3D structure of a small part of the TSR. Our findings uncover that recognition of distinct conserved 3D fold motifs can be used as a mechanism to achieve substrate specificity by enzymes modifying completely folded proteins of very wide sequence diversity and biological function.

Introduction

Protein glycosylation is the most abundant and diverse co- and post-translational modification in life. In eukaryotes, >50% of proteins are modified with carbohydrates (Apweiler *et al*, 1999) which together regulate myriad biological processes. Altered or defective protein glycosylation pathways cause various developmental defects as reflected in the rapidly growing number of congenital disorders of glycosylation (Freeze, 2007; Jaeken and Matthijs, 2007).

*Corresponding author. Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland.
Tel.: +41 61 696 70 38; Fax: +41 61 697 39 76;
E-mail: heinz.gut@fmi.ch

The unusual protein *O*-linked fucosylation has been described on thrombospondin type 1 repeats (TSR) (Hofsteenge *et al*, 2001; Gonzalez de Peredo *et al*, 2002) and epidermal growth factor-like (EGF) repeats (Bjoern *et al*, 1991; Buko *et al*, 1991; Harris *et al*, 1992; Nishimura *et al*, 1992; Harris and Spellman, 1993) and is catalysed by the protein *O*-fucosyltransferase 2 (POFUT2) and protein *O*-fucosyltransferase 1 (POFUT1), respectively (Harris and Spellman, 1993; Luo *et al*, 2006a). Both enzymes transfer the fucose moiety from GDP-fucose to a serine or threonine residue of the properly folded acceptor molecule, recognizing the consensus sequences $CX_{2-3}(S/T)CX_2G$ (Hofsteenge *et al*, 2001) in TSR or $CX_{4-5}(S/T)C$ (Harris and Spellman, 1993) in EGF repeats, respectively. The fucose residue on TSR can be elongated to a glucose-β1,3-fucose disaccharide by the β1,3-glucosyltransferase (β3GlcT) (Kozma *et al*, 2006; Sato *et al*, 2006). In EGF repeats, the fucose may be extended to an NeuAc-α2,3/α2,6-Gal-β1,4-GlcNAc-β1,3-Fuc tetrasaccharide catalysed by the sequential enzymatic activity of Fringe, β1,4-galactosyltransferase 1 and α2,3/α2,6-sialyltransferase (Nishimura *et al*, 1992; Harris and Spellman, 1993; Stanley, 2007; Luther and Haltiwanger, 2009; Rana and Haltiwanger, 2011). Both TSR and EGF repeats are small cysteine-rich, layered structural motifs with three conserved disulphide bonds and little secondary structural elements. TSR and EGF repeat proteins are sequence-wise very diverse with only a few structural key residues being conserved. The glycosyltransferases involved in the *O*-fucosylation pathways of TSR and EGF repeats are specific and do not crossreact (Luo *et al*, 2006b).

The importance of protein glycosylation on EGF repeats has been extensively studied in the Notch signalling pathway (Luther and Haltiwanger, 2009) where the EGF modification was shown to regulate embryonic development and tissue renewal by controlling the ligand specificity of Notch (Stanley, 2007; Stahl *et al*, 2008). Crystal structures of *C. elegans* POFUT1 alone and in complex with GDP-fucose or GDP have been solved recently and give insight into overall protein structure and the enzymatic mechanism (Lira-Navarrete *et al*, 2011). The role of POFUT2-dependent fucosylation of TSR on the other hand is less clear. Progress was made recently by Du *et al* (2010) using *Pofut2* knockout mice where they found that *O*-fucosylation of TSR is critical for restricting epithelial-to-mesenchymal transition, correct patterning of the mesoderm, and localization of the endoderm in embryonic development. In *C. elegans*, POFUT2-dependent TSR fucosylation was found to be involved in the regulation of distal tip cell migration (Canevascini *et al*, 2006). TSR proteins are expressed in the secretory pathway with *O*-fucosylation occurring within the endoplasmic reticulum. In cell culture experiments, mutation of fucosylation sites on TSR of ADAMTS13 (Ricketts *et al*, 2007) and Punctin-1 (Wang *et al*, 2007) reduced or completely abolished secretion of the proteins, indicating

that POFUT2-dependent _O_-fucosylation on TSR might be required for optimal secretion of these proteins. No disorder has yet been directly linked to a genetic defect of the _Pofut2_ locus in humans. However, mutations in the _B3GALTL_ gene that encodes the β3GlcT enzyme responsible for glucose transfer onto _O_-fucosylated TSR cause the autosomal recessive disorder Peters Plus syndrome (Lesnik Oberstein _et al_, 2006; Hess _et al_, 2008). This disorder is characterized by anterior-eye-chamber abnormalities, disproportionate short stature and developmental delay.

Protein _O_-fucosylation raises three fundamental questions about the interaction between glycosyltransferases and their protein substrate: How does a glycosyltransferase accommodate a fully folded protein substrate in its active site? Which structural features are used to discriminate between the different families of protein substrates and how can specificity be achieved in the case of sequence-wise degenerated protein substrates? We have addressed these questions by determining the structure of human POFUT2 (alone and in complex with the sugar donor GDP-fucose) and steady-state kinetic analysis of wild-type and mutant transferase. To investigate further how POFUT2 interacts with its TSR sugar acceptor, we have analysed _O_-fucosylation of wild-type and mutant TSR in an _in-vitro_ assay and in mammalian HEK293T cells. The crystal structure shows that POFUT2 belongs to the classical GT-B fold family of glycosyltransferases with two closely interacting Rossmann-like domains. The C-terminal domain binds the GDP-fucose moiety while the TSR substrate is recognized by a large cavity in the centre of the bilobal structure. Based on our structural data and steady-state kinetic measurements, we suggest that the conserved E54 residue acts as the catalytic base, and describe key catalytic residues located in the active site. Structural and biochemical knowledge was used to clarify why only TSR modules can bind to the sugar acceptor site and to design an artificial minimal TSR module which we show to be sufficient as sugar acceptor for common TSR glycan modifications (_O_-fucose-glucosylation and _C_-mannosylation). Furthermore, we investigated how POFUT2 substrate specificity is achieved despite the large sequence diversity present in TSR containing the $CX_{2-3}(S/T)CX_2G$ fucosylation motif. We present the structure of a protein glycosyltransferase modifying a completely folded protein substrate and propose a novel mechanism of enzyme-protein substrate specificity, based on recognition of a small conserved 3D structural motif. It explains how site-specific modifications can take place in the absence of a conserved protein sequence.

Results

Crystal structure of human POFUT2

We have expressed and purified human Δ21-POFUT2 from mammalian cell culture and have determined its crystal structure at 3.0 Å resolution. The protein crystallized in space group P3$_2$21 with two molecules in the asymmetric unit (a.u.) and the structure was solved by the single isomorphous replacement with anomalous scattering (SIRAS) method using a platinum derivative. Data collection, phasing and refinement statistics are presented in Table I. The refined POFUT2 crystal structure displays clear electron density for residues 41–429 (out of 22–429) and the two molecules in the a.u. are almost identical with an r.m.s.d. of only 0.49 Å. The

structure of POFUT2 is composed of two Rossmann-like domains with β/α/β topology typical of the GT-B fold of glycosyltransferases (Figure 1A). N- and C-terminal domains encompass residues 22–242 and 243–429, respectively. The two domains interact closely with each other (buried surface area of 1416 Å^2) forming an extended protein unit. Fully structured loops originating from both the N-terminal (Q141–V156, E158–N189) and C-terminal domain (T407–Y429, L293–L309) form a large central cavity in the molecule with two disulphide bonds stabilizing loop conformations in each domain (C161–C192 and C412–C419). A second narrower cleft is present in the C-terminal domain, formed by helices α13 and α14, loop Q93–Q99 and the N-terminal tip of helix α1 (E54–N57). Electron density for three N-acetylglucosamine (GlcNAc) moieties is present at residues N189, 209, and 259 revealing all predicted N-glycosylation sites occupied. The quality of the electron density allowed model building of GlcNAc moieties at N189 and N259.

In order to identify functional POFUT2 regions involved in catalysis and substrate binding, we mapped conserved residues onto the protein surface and also analysed the electrostatic surface potential (Figure 1B–D). Martinez-Duncker _et al_ (2003) identified three conserved peptide motifs, which are shared among all four families of fucosyltransferases. These peptide motifs (I, II, and III) map onto the bottom and one wall of the narrow cleft in the C-terminal domain that branches away from the central large cavity (Figure 1B). The fact that this cavity also shows a highly positive electrostatic potential at its entrance up to the middle (Figure 1C) and that superposition of the _C. elegans_ POFUT1 GDP-fucose complex placed the nucleotide sugar in the same region, made it very likely that it harbours the GDP-fucose binding site. Additional conserved residues (Figure 1B and D) mapped onto a second extended surface patch located at the bottom of the large cavity formed by N- and C-terminal loops in the centre of the two domains. Considering the shape and dimensions of this cavity, we hypothesized the TSR substrate to bind in this central area.

We searched the Protein Data Bank (PDB) to identify structurally closely related proteins using DALI (Holm and Rosenström, 2010; Supplementary Table SI; Figure 2; Supplementary Figure S1). A search with the entire POFUT2 structure revealed the structure of _C. elegans_ POFUT1 to be most similar (PDB 3ZY2; Lira-Navarrete _et al_, 2011) followed by the nodulation fucosyltransferase NODZ (PDB 2HHC) (Brzezinski _et al_, 2007), the lipopolysaccharide heptosyltransferase I WaaC (PDB 2H1H) (Grizot _et al_, 2006), and the α1,6-fucosyltransferase FUT8 (PDB 2DE0) (Ihara _et al_, 2007). If the N-terminal domain alone was used in the search, then structures of POFUT1 and NODZ gave the highest Z-scores followed by very distantly related Rossmann-like fold proteins with low scores. A search with the C-terminal domain alone on the other hand yielded POFUT1, NODZ, WaaC, and FUT8 as close structural neighbours. _C. elegans_ POFUT1 and human POFUT2 (21% sequence identity) have a very similar core structure in the two Rossmann fold domains and also share the same arrangement of N- and C-terminal domains but differ significantly in many surface exposed structural elements (Figure 2; Supplementary Figure S1). N-terminally, the POFUT2 loop 85–103 that is in a coiled conformation is replaced by an additional short β-hairpin in POFUT1. The

Δ **Table I** Data collection and refinement statistics

	POFUT2 native	POFUT2 Pt derivative	POFUT2 GDP-fucose complex
Data collection			
Space group	P3$_2$21	P3$_2$21	P3$_2$21
Cell constants *a, b, c* (Å)	118.6, 118.6, 196.2	118.5, 118.5, 195.0	153.0, 153.0, 185.7
Wavelength λ (Å)	1.000	0.890	1.000
Resolution range (Å)[a]	30.0–3.0 (3.11–3.00)	20.0–5.5 (5.69–5.50)	40.0–3.4 (3.63–3.40)
Unique reflections	31 375	9737	32 751
Completeness (%)[a]	96.4 (71.6)	100 (99.9)	93.3 (93.6)
Multiplicity	11.3	10.4	4.7
R_{sym} (%)[a,b]	12.4 (39.7)	22.4 (43.4)	16.6 (76.6)
$I/\sigma(I)$[a]	21.0 (2.4)	11.3 (3.7)	11.9 (2.1)
Phasing power iso/ano		1.04/0.53	
Refinement			
Resolution range (Å)	30.0–3.0		40.0–3.4
Reflections (all)	31 318		32 745
Reflections (test set)	1593 (5.1%)		1622 (5.0)
R_{crys} (%)	17.4		19.3
R_{free} (%)	23.6		23.8
r.m.s.d.			
Bond lengths (Å)	0.008		0.011
Bond angles (deg)	1.28		1.31
Wilson *B*-factor (Å^2)	50.9		—
Mean B-factor (Å^2)			
Protein	54.6		102.2
Ligand	—		113.4
Ramachandran plot (%)			
Favoured	97.0		97.0
Allowed	3.0		3.0
Outliers	0		0

[a]Values in parentheses refer to the highest resolution shell.
[b]$R_{sym} = \Sigma_{hkl}\Sigma_j |I_{j,hkl} - \langle I_{hkl} \rangle| / \Sigma_{hkl}\Sigma_j I_{j,hkl}$ where $\langle I_{hkl} \rangle$ is the average of the intensity $I_{j,hkl}$ over $j = 1, \dots, N$ observations of symmetry equivalent reflections *hkl*.

two structures differ dramatically in the POFUT2 region 140–200 where the long structured POFUT2 loop comprising residues 140–156 is missing in POFUT1. In the C-terminal domain, three striking structural differences can be identified. First, the long POFUT2 loop (260–287) that reaches over to the N-terminal domain opposite of the substrate binding cleft is not present in the *C. elegans* POFUT1 structure. Second, the POFUT2 loop 293–307 that builds the C-terminal wall of the central protein cleft is replaced by an additional small domain in POFUT1 (239–283) formed by three short helices that together restrict access to the central POFUT1 protein cavity. Third, the POFUT1 sequence is much shorter and the structure ends after the last C-terminal β strand where POFUT2 continues with a large disulphide linked turn followed by a long stretch of residues in a rippled conformation defining the entry to the central protein cavity on the C-terminal side. Superposition of C-terminal domains of POFUT2, NODZ, and FUT8 reveals these domains to be similar with the central β-sheet and surrounding helices superimposing very well (r.m.s.d. 3.1 and 2.7 Å, respectively). Nevertheless, more detailed analyses identify structural differences: Again, the POFUT2 loop 260–287 opposite of the substrate binding cleft is not present and also the two last strands of the central β-sheet of NODZ and FUT8 differ by forming a regular β-sheet while they are in a rippled conformation with an SS-bridge connecting the two strands in POFUT2. Superposition of the entire POFUT2 structure with DALI hits that rank after POFUT1 show that the N-terminal domains of FUT8 and

WaaC are structurally very different. Only the first and last helix of the N-terminal domain and some strands of the central β-sheet overlap well with NODZ and WaaC.

POFUT2 enzymatic activity

To validate our model of POFUT2 interaction with the GDP-fucose sugar donor and the TSR sugar acceptor, we established an LC-MS-based enzyme activity assay and tested the capability of wild-type and mutant POFUT2 to fucosylate TSR4 from rat F-spondin. In an initial set of experiments, we analysed the effect of two different N-terminal boundaries (Δ21- and Δ36-POFUT2) and of varying N-glycan structures, as well as the influence of different divalent cations on the enzymatic activity of wild-type POFUT2 (Figure 3). While neither changing the N-terminal boundary nor having a different glycoconjugate composition had an effect on the enzymatic activity, we found that different metal ions influence catalytic activity in different ways. Mg^{2+}, followed by Mn^{2+} and Ca^{2+}, activated the enzyme in decreasing order (100, 90, and 80% relative activity) but Zn^{2+} completely abolished its activity. The enzyme was still active in the presence of EDTA, albeit at a very low level ($\sim$5% relative activity). Having a sensitive enzymatic activity assay available that monitors directly TSR fucosylation, we determined the steady-state kinetic parameters for GDP-fucose and TSR4 using wild-type high mannose type Δ21-POFUT2 (Supplementary Figure S2). POFUT2 is an efficient enzyme with K_M values of 9.8 and 29.5 μM for GDP-fucose and TSR4,

Structure of human protein *O*-fucosyltransferase 2
C-I Chen *et al*

respectively, and a k_{cat} of 144 per minute. Based on our structural results, we designed 14 mutations targeting specific putative functional residues of the enzyme and tested enzymatic activity (Figure 3D and E). Out of 14 mutations, 2 did not yield any soluble protein pointing to a critical function of these residues in the folding pathway of the protein (D333A and Δ265–285). All other mutants expressed and purified well and equal amounts were used for the activity assay.
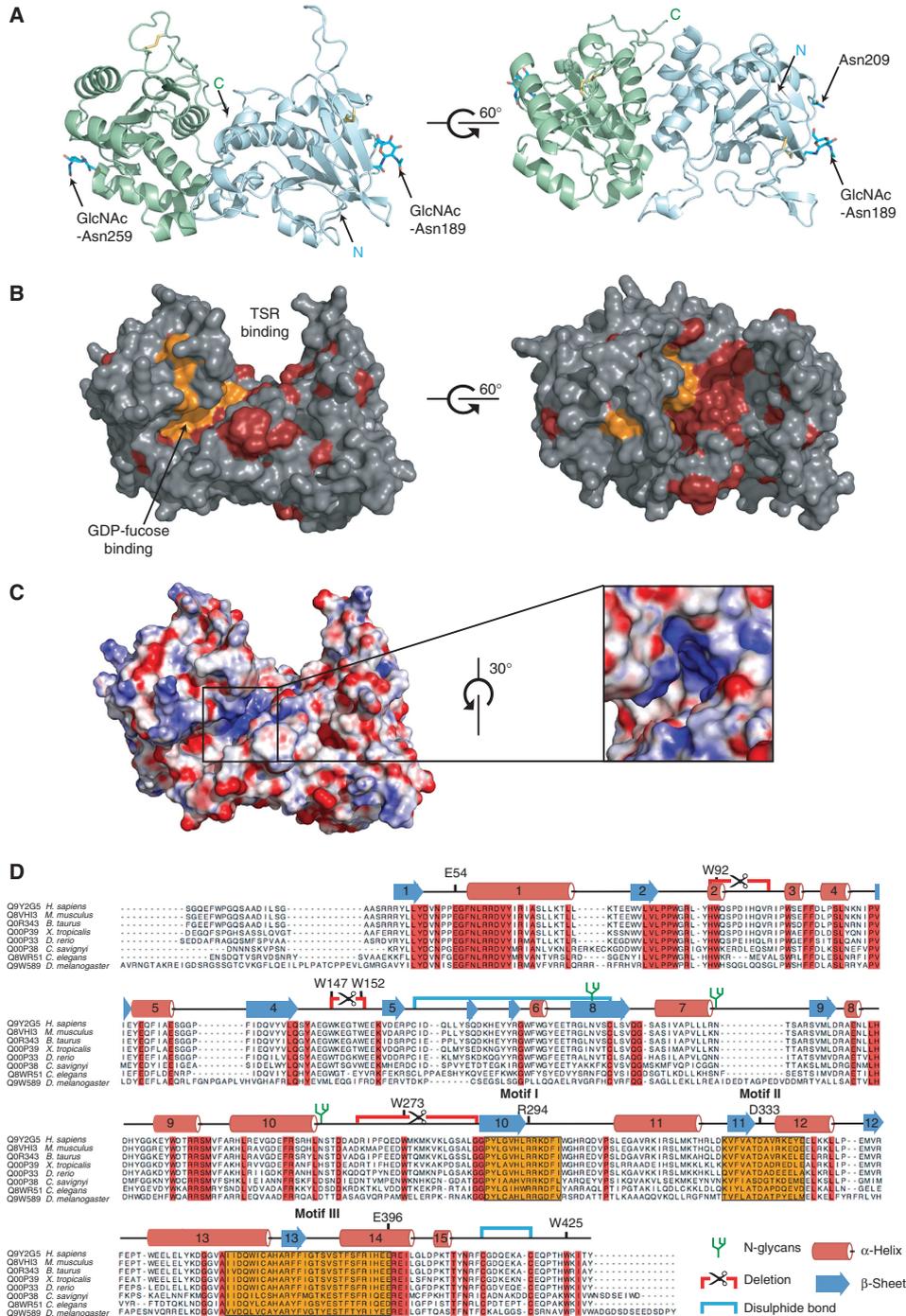
From the POFUT2 mutants targeting the catalytic mechanism, E54A and R294A resulted in complete loss of activity while the D297A and E396A mutants remained active (15 and 8%, respectively). A change of the highly conserved W92 to alanine as well as deletion of the entire loop (Δ90–100) abolished enzymatic activity. POFUT2 features a unique loop (265–285) located on the opposite side of the large cleft (Figure 3D), which protrudes from the C-terminal domain and which is attached to the N-terminal domain *via* a completely conserved tryptophan residue. We hypothesized that this residue, W273, is involved in controlling movements of the N- and C-terminal domain relative to each other during the catalytic cycle and indeed lost 90% activity when we mutated W273 to alanine. A series of mutations targeted the large loop forming one wall of the central cavity (residues 147–152) with the aim of disturbing TSR binding. While the point mutations reduced the catalytic activity to ∼10–73%,

removal of the entire loop (Δ147–152) yielded a completely inactive enzyme. Alanine mutation of another conserved residue (W425) located at the entry of the central cavity and potentially involved in TSR binding also reduced POFUT2 activity to 38%.

Crystal structure of human POFUT2 in complex with GDP-fucose

In crystals of native POFUT2, we found the putative GDP-fucose binding pocket to be partially covered by a loop from the second molecule present in the a.u. This made it impossible to obtain a structure of the binary complex by soaking experiments and despite extensive efforts we did not obtain crystals in co-crystallization experiments. We only obtained crystals when we used the catalytically incompetent POFUT2 E54A mutant that allowed us to solve the structure of POFUT2 in complex with the sugar donor. Analysis of the
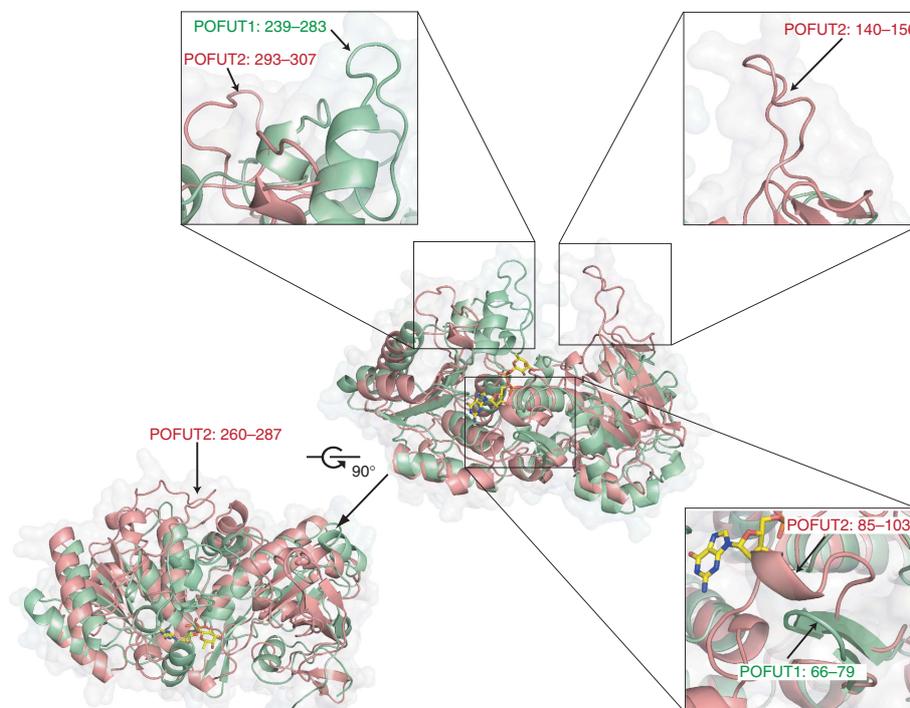

Figure 2 Structural superposition of human POFUT2 and *C. elegans* POFUT1 in two orientations rotated by 90°. POFUT2 (red) and POFUT1 (green) are displayed as cartoon models with transparent surfaces. GDP-fucose bound to POFUT1 is shown as sticks (yellow and atom colours). Main structural differences are highlighted in close-up images and are labelled.

Figure 1 Crystal structure of human POFUT2 and multiple sequence alignment of orthologue sequences. (**A**) The human POFUT2 structure is displayed as cartoon model in two orientations 60° apart. N- and C-terminal domains are in grey and green, respectively, N- and C-termini are labelled. Disulphides and covalently bound GlcNAc molecules are displayed as sticks in atom colours. (**B**) Conserved residues from the multiple POFUT2 sequence alignment (**D**) are mapped onto the surface of the human POFUT2 structure (100% conservation: red, conserved motifs I–III among fucosyltransferases: orange). Substrate binding sites are labelled. (**C**) Mapping of the electrostatic surface potential onto the surface of the POFUT2 structure (scale: − 20 to + 20 kT/e from red to blue). The highly positive surface patch involved in GDP-fucose binding is boxed, zoomed-in and rotated by 30° in the right image. Computed with the APBS plugin of PyMol (Baker *et al*, 2001). (**D**) Multiple sequence alignment of selected POFUT2 sequences (ClustalW; Larkin *et al*, 2007). Structural features present in the POFUT2 structure and mutated residues are indicated.
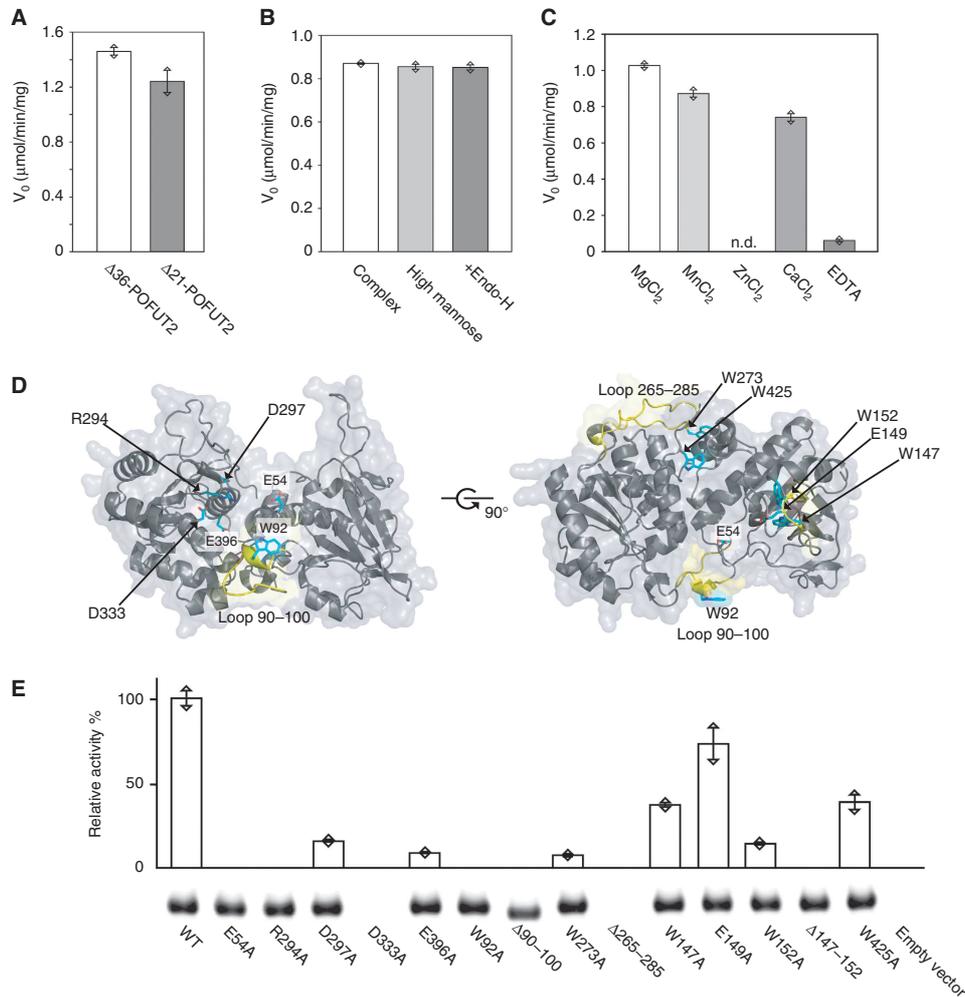
Figure 3 Initial rate of fucosylation of purified wild-type and mutant POFUT2. (**A**) Influence of different N-terminal POFUT2 truncations. (**B**) Influence of the glycan composition (complex, high mannose, + Endo-H). (**C**) Presence or absence of varying divalent metal ions (n.d.: not detected). Columns in (**A–C**) represent the average of two independent measurements while individual measurements are shown as triangles. (**D**) POFUT2 mutations used in the enzymatic *in-vitro* assay. POFUT2 is shown as a cartoon model with transparent surface in two orientations. Point mutations (sticks, cyan) and deletions (yellow) are indicated. (**E**) Relative enzymatic activity of different point and deletion mutations compared with the wild-type (WT) protein. Experiments were repeated twice and the mean of the two measurements and the individual measurements are shown. The western blot signal used for quantification of the enzyme input is indicated.

crystal packing revealed the presence of a crystallographic dimer similarly to the non-crystallographic-symmetry dimer present in the a.u. of the apo structure but with a reduced interface enabling access to the GDP-fucose binding site (Supplementary Figure S3). Clear electron density for GDP-fucose was present in all four molecules in the a.u. (space group P3$_2$21, 3.4 Å resolution) and located at the predicted position in the narrow cleft leading from the N- and C-terminal domain interface into the C-terminal domain (Figure 4A). The guanine purine base is mainly held in place by stacking interactions with F389 and hydrogen bonds between the N1 nitrogen and the D371 side chain while other residues of the pocket additionally contribute hydrophobic interactions (Figure 4B). The ribose moiety

bulges up from the bottom of the cleft and does not show any tight interaction with the protein. Instead, the main affinity for the sugar donor comes from the interaction of the diphosphate group with the protein. The guanidinium moiety of R294 forms a salt bridge with the β-phosphate while the positive dipole located at the N-terminal end of the last helix (387–400) tightly attaches the diphosphate moiety to the helix tip where it hydrogen bonds side chain (T388) and backbone atoms of residues T388 and F389. Strikingly, the fucose is arranged almost perpendicular to the nucleotide diphosphate *via* hydrogen bonds between the O3 hydroxyl and the P53 carbonyl group and the O2 hydroxyl and the G55 amide nitrogen of the N-terminal domain. This arrangement presents the activated phosphoester bond at the anomeric
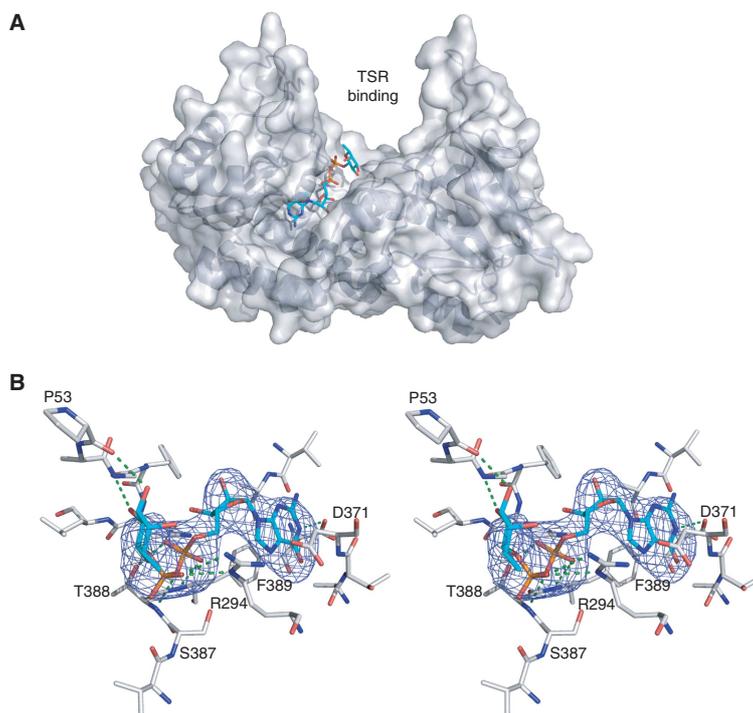
Figure 4 GDP-fucose binding in human POFUT2. (**A**) Structure of the POFUT2 GDP-fucose complex displayed as a cartoon model in grey with transparent surface. GDP-fucose is shown as sticks in atom colours. (**B**) Stereo figure of the GDP-fucose (cyan sticks) binding mode in POFUT2 (white sticks). NCS averaged mF_o-DF_c electron density calculated after molecular replacement and one round of refinement (in absence of GDP-fucose) is shown as mesh in blue (3σ). Hydrogen bonds are displayed as green dotted lines. The orientation is different than in (**A**) for better clarity.

carbon to the large open channel where the TSR substrate is postulated to bind. Both the 1C4 and 4C1 fucose ring conformations (C1-O1 bond in equatorial and in axial position, respectively) have been refined against the low resolution data (Supplementary Figure S4) and since the 1C4 conformation resulted in a slightly better fit to the experimental data we included it in the final model of the complex. The overall GDP-fucose binding mode in POFUT1 and POFUT2 is similar but a detailed analysis uncovers important differences which likely have an impact on the catalytic mechanism (Lira-Navarrete *et al*, 2011). In POFUT2, the fucose moiety is freely accessible from the large central protein cavity where we expect TSR to bind. This is not the case in POFUT1 where the additional small helical domain (239–283), that provides the F261 residue holding the fucose in place, blocks access together with F199 from the N-terminal domain. In addition, POFUT1 residue R40 completely covers with its side chain GDP-fucose from the top thereby limiting access from the other side. This residue is replaced by G55 in POFUT2 and GDP-fucose is solvent exposed. In general, GDP-fucose is much more buried in POFUT1 compared with POFUT2 also leading to different dihedral angles of the diphosphate group. Although the sugar donor could only be modelled with limited accuracy in POFUT2 due to limited resolution, the binding mode clearly differs in several details.

Structural restraints in TSR for productive fucose attachment

Our structure of the POFUT2 GDP-fucose complex together with the structures of fucosylated TSR2-TSR3 of human TSP-1 (Tan *et al*, 2002) and fucosylated TSR1 of ADAMTS13 (Akiyama *et al*, 2009) enabled us to build a model of the full enzyme substrate ternary complex. Superposition of anomeric fucose carbons in these structures with the anomeric carbon in the POFUT2 GDP-fucose complex, followed by manual adjustment, yielded the overall TSR position. This initial model was used to overlay the TSR4 structure from rat F-spondin (PDB 1VEX) (Pääkkönen *et al*, 2006) and energy minimization of the full complex in CNS (Brunger *et al*, 1998). We obtained a plausible model of the ternary complex in which the elongated TSR unit lies in the deep interdomain cavity of POFUT2 spanning across the glycosyltransferase (Figure 5A). The TSR module contacts the highly conserved POFUT2 surface *via* its flat hydrophobic side, opposite of the SS-bond pattern and the tryptophan-arginine stacking (CWR layer), where the B and C strand show a regular antiparallel β-sheet. In addition, the rippled A strand contacts the bottom and the side wall of the cleft. The entire AB loop harbouring the $CX_{2-3}(S/T)CX_2G$ motif is in close contact with POFUT2. Only half of the TSR module was predicted to interact with POFUT2 whereas the N-terminal
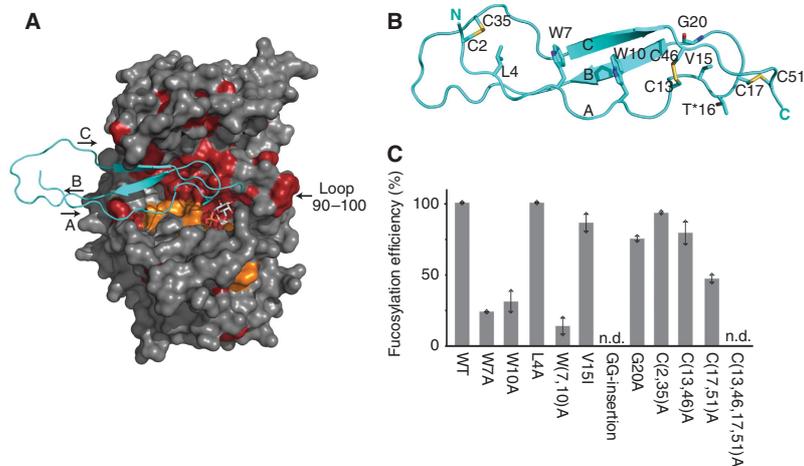
Figure 5 Model of the POFUT2 GDP-fucose TSR complex and POFUT2 enzymatic activity using mutant TSR as substrate. (**A**) Surface representation of the POFUT2 GDP-fucose structure with POFUT2 residues coloured as in Figure 1B and GDP-fucose as sticks. TSR4 from rat F-spondin (PDB 1VEX) is displayed as a cartoon model in cyan. (**B**) Structure of TSR4 from rat F-spondin. Residues that were mutated for fucosylation measurements are numbered. (**C**) Relative fucosylation efficiency of TSR mutants. Wild-type and mutant TSR was expressed and purified from HEK293T cells and analysed by mass spectrometry for fucosylation efficiency. n.d.: not detected. Experiments were repeated twice and the mean of two measurements and the individual measurements are displayed.

part of the A strand and the BC loop including the jar handle are solvent exposed. Interestingly, we find a second conserved POFUT2 surface patch located on the 90–100 loop that could interact with an additional TSR domain (e.g., TSR2–TSR3 in TSP-1) or with other protein domains on the C-terminal side of the TSR unit (e.g., EGF repeat in TSP-1 or C_A domain in ADAMTS13).

In order to validate our model of TSR binding, we expressed a series of F-spondin TSR4 mutants in HEK293T cells, purified the secreted protein from the medium and analysed its fucosylation state by mass spectrometry (Figure 5B and C). Changes to the conserved W*XX*W*XX*W motif in the A strand (in the case of TSR4 of F-spondin L$_4$*XX*W$_7$*XX*W$_{10}$), the key element of the TSR fold that forms the multi-layered delocalized π-system with the conserved arginine residues from the B strand (WR of CWR), reduced the efficiency of fucosylation significantly. Single mutations of W7 and W10 to alanine produced only 24 and 31% of fucosylated TSR4, double mutations on both sites further reduced the fucosylation to 14% compared with wild type whereas the L4A mutation did not have any influence. Next, we introduced mutations in the AB loop close to the threonine (T*16) that undergoes fucosylation: replacing the valine (V15) next to it with an isoleucine or changing the glycine (G20) at the end of the turn to alanine had only minor effects (86, 75% fucosylation). On the other hand, when we introduced two additional glycine residues right before the threonine (*GG*T*, mimicking the fucosylation sequon in EGF repeats) to create a larger loop between the disulphide forming cysteine and the threonine, fucosylation was completely abolished. The disulphide bond pattern is a hallmark in the fold of TSR therefore we investigated how fucosylation is affected by removing SS-bonds. Mutation of the cysteines forming the SS-bond between the A and C strand, where we postulated no interaction with POFUT2, was well tolerated with no reduction in fucosylation

levels (C2,35A). To our surprise, the removal of the second SS-bond between the A and C strand had no dramatic effect with 80% fucosylated product (C13,46A). On the other hand, removal of the SS-bond that connects the AB loop (C17,51A) with the C-terminal end of the C strand resulted in 55% reduced fucosylation levels. When we removed both SS-bonds, which together stabilize the 3D conformation of the AB loop, we could not detect any fucosylated TSR product.

We superimposed all available structures of TSR modules (PDB 1LSL, Tan *et al*, 2002; 1VEX and 1SZL, Pääkkönen *et al*, 2006; 3GHM, Akiyama *et al*, 2009; 2BBX, Tossavainen *et al*, 2006) at the predicted substrate binding site of POFUT2 and analysed structural and sequence restraints potentially involved in substrate recognition. The structural information was compared with the profile hidden Markov model (HMM) of the Pfam entry of the thrombospondin type 1 domain (PF00090, http://pfam.sanger.ac.uk; Supplementary Figure S5) and multiple sequence alignments of all human TSR type 1 domains present in the Uniprot database (http://www.uniprot.org) that contain the putative fucosylation motif (C*X$_{2-3}$*(S/T)C*X$_2$*GG). The sequence data revealed the enormous sequence diversity besides only a few conserved residues that seem to be absolutely necessary for the TSR fold. Taking into account our experimental data about the contribution of selected TSR structural elements to POFUT2 fucosylation efficiency, we hypothesized that substrate specificity is not primarily encoded in the protein sequence but rather in the unusual 3D structure of one half of the TSR module.

Consequently, we investigated whether a minimal TSR module containing only seven conserved structural residues of the TSR hallmark elements (disulphide pattern and side-chain stacking) would be sufficient as a POFUT2 sugar acceptor. We designed and expressed a truncated TSR module based on F-spondin TSR4 consisting of approximately half
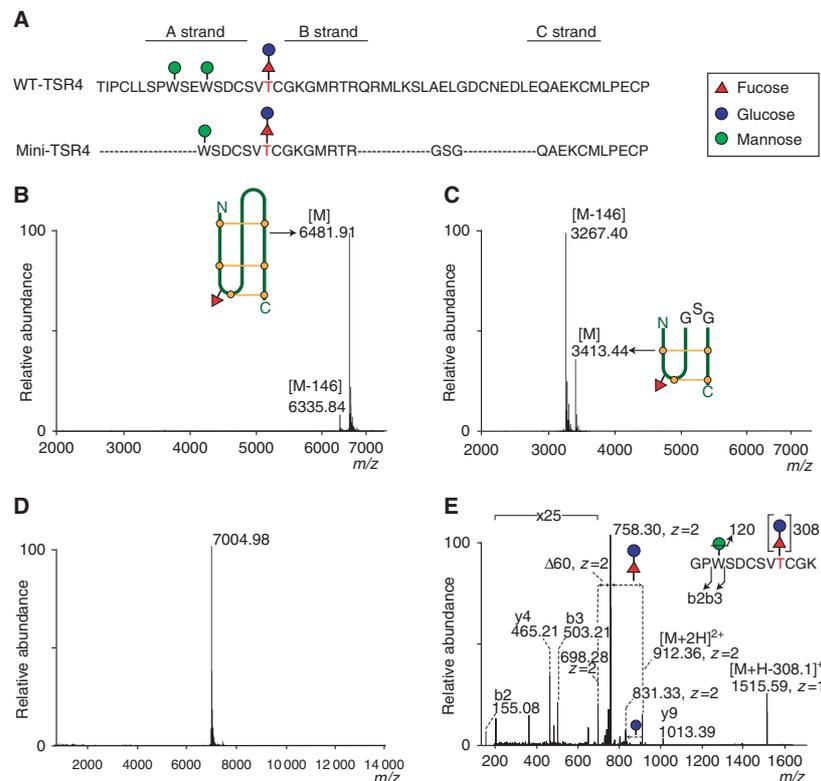
Figure 6 *O*-fucosylation of mini-TSR. (**A**) Amino-acid sequences of wild-type TSR4 from rat F-spondin and engineered mini-TSR are compared. Amino acids that form the three strands are indicated. Wild-type TSR and mini-TSR were expressed and isolated from *E. coli*. These modules were used as acceptor substrates in the POFUT2 *in-vitro* assay and the reaction products analysed by mass spectrometry. Both the wild-type TSR (**B**) and mini-TSR (**C**) show two peaks that differ by 146 Da, indicating the presence or absence of one fucose molecule. Wild-type and mini-TSR are drawn as cartoons with disulphide bonds in yellow and fucose as a red triangle. Mini-TSR is approximately half the size of wild-type TSR, with a truncated A strand and the deleted BC loop replaced by a three residue linker, GSG. (**D**) An N-terminal His$_6$–FLAG-3C-tagged mini-TSR was expressed in mammalian HEK293T cells, isolated from the medium and analysed by mass spectrometry. The mono-isotopic mass (7004.98) represents the intact, tagged mini-TSR with two hexoses and one deoxyhexose. (**E**) Secreted and purified mini-TSR from HEK293T cell expression was digested with human rhinovirus 3C and Lys-C protease and analysed by mass spectrometry. MS/MS analysis confirmed the sequence (GPWSDCSVTCGK) and revealed the glycan modifications. Loss of one hexose and one hexose-deoxyhexose disaccharide from the parent ion (*m/z* 912.36) was observed in the MS spectrum (*m/z* 831.33 and 758.30, respectively) and the mass difference between the b2 and b3 ion corresponded to one hexose molecule attached to the tryptophan. The characteristic − 120 Da fragment (*m/z* 698.28) represents the typical signature motif for *C*-hexosylation.

the length and tested fucosylation by mass spectrometry. The first TSR residue of this artificial minimal TSR corresponds to W10 in the A strand (Figure 5B) and includes the entire AB loop where the consensus sequence is located. Residues predicted not to be involved in POFUT2 interaction connecting the B and C strand were deleted and replaced with a short GSG linker (Figure 6A). This artificial minimal TSR comprised only 29 residues compared with 55 for wild-type TSR and is referred to as mini-TSR in this article.

Wild-type TSR4 and mini-TSR were expressed and isolated from *E. coli* and used as acceptor substrates for purified POFUT2 in a qualitative *in-vitro* fucosyltransferase assay. The reaction products were monitored for fucose incorporation by mass spectrometry. An increase in mass of 146 Da corresponding to the addition of one fucose moiety was observed for both the wild-type (Figure 6B) and mini-TSR substrate (Figure 6C). Fucosylation was only observed in reactions that included POFUT2. Thus, POFUT2 recognizes

and modifies both wild-type and truncated artificial mini-TSR. We further validated our findings by expressing mini-TSR with an N-terminal His–FLAG tag in HEK293T cells. The secreted mini-TSR protein was purified from the culture medium and analysed by mass spectrometry. We found a homogenous species of 7004.98 Da corresponding to the tagged mini-TSR with two hexoses and one deoxyhexose attached (Figure 6D). To further analyse the glycosylation state of the mini-TSR, secreted and purified protein was digested with human rhinovirus 3C to remove the tag, reduced and alkylated, and digested with Lys-C protease yielding a peptide covering the predicted glycosylation sites. MS/MS analysis confirmed the sequence (GPWSDCSVTCGK) and revealed the glycan modifications (Figure 6E). Loss of one hexose and one hexose-deoxyhexose disaccharide from the parent ion (*m/z* 912.36) was observed in the MS spectrum (*m/z* 831.33 and 758.30, respectively). In addition, the mass difference between the b2 and b3 ion corresponded to one

hexose molecule attached to the tryptophan and the characteristic $-120\,$Da fragment (m/z 698.28) revealed the tryptophan hexosylation to be *C*-linked (Hofsteenge *et al*, 1994). In summary, the mass spectrometry analysis demonstrates that mini-TSR is modified with an *O*-linked fucose-glucose disaccharide and a *C*-linked mannose on the tryptophan as it is the case for wild-type TSR4 (Hofsteenge *et al*, 2001). These experiments confirmed that mini-TSR is a substrate for POFUT2 both in HEK293T cells and in our *in-vitro* fucosyltransferase assay.

Discussion

POFUT2 protein structure and TSR substrate recognition

Here, we present the crystal structure of human POFUT2 that together with orthologues forms the GT68 family of inverting protein *O*-fucosyltransferases of the GT-B fold (Cantarel *et al*, 2009). Although many structures of glycosyltransferases are solved there is only very limited structural information available for glycosyltransferases that transfer the sugar moiety to a peptide or protein acceptor. For the three glycosyltransferase folds (GT-A, B, C) only structures of GALNT2 and GALNT10 (GT-A), AglB and PglB (GT-C), and MurG, OGT, and POFUT1 (GT-B) have been solved (Hu *et al*, 2003; Fritz *et al*, 2006; Kubota *et al*, 2006; Maita *et al*, 2010; Lazarus *et al*, 2011; Lira-Navarrete *et al*, 2011; Lizak *et al*, 2011). Interestingly, all of these enzymes (except POFUT1) use flexible solvent exposed protein regions as sugar acceptor whereas POFUT2 was shown to fucosylate only properly folded TSR (Luo *et al*, 2006b). The POFUT2 structure now gives insight into how substrate recognition, specificity, and catalysis are achieved with the special requirements of a properly folded 3D protein sugar acceptor that transiently forms a protein–protein interface with a glycosyltransferase.

Our data suggest that POFUT2 recognizes key 3D structural TSR elements formed by the disulphide pattern and side-chain stacking common to sequence-wise degenerated TSR modules. A search of the PDB using DALI with coordinates of a minimal TSR poly-alanine module (regions predicted to interact with POFUT2) identified only known structures of TSR domains without discovering this structural motif in any other protein. Therefore, substrate specificity seems to be achieved by the structural complementarity of a part of the TSR fold with the POFUT2 binding site and the wide TSR sequence diversity does not play a role as long as the critical TSR fold motif is intact. From our experimental data and the structural models, we conclude that disrupting the conformation of the rippled A strand (formed by the W*XX*W*XX*W motif, L*XX*W*XX*W in TSR4 from F-spondin) and of the AB loop (defined by the second and third disulphides) impairs TSR substrate recognition and fucosylation efficiency (Figure 5C). Our data of the mini-TSR also show that starting the A strand directly at the third tryptophan of the W*XX*W*XX*W motif has no negative effect on fucosylation as the conformation of the shortened A strand is still intact. Proposed key interactions are located at the entry of the large TSR binding cavity where conserved POFUT2 residues W152 from the N-terminal domain and W425 from the very C-terminal part define the most narrow part of the cleft ($\sim15\,$Å) allowing only space for a two stranded β-sheet to enter the cavity (Figure 7A). C-terminal residues seem to lock

the position of the bound TSR module by interacting with the backbone bulge formed by the second and third tryptophan of the L*XX*W*XX*W motif of the rippled A strand. Thereby, C-terminal POFUT2 residues act as a ruler to position the S/T residue undergoing modification exactly at the right position for E54-dependent deprotonation and nucleophilic attack at the anomeric GDP-fucose carbon. This model is supported by our observation that fucose attachment is reduced by 90 and 61% for the W152A and W425A mutation, respectively (Figure 3E), and that exchange of the second and third F-spondin TSR4 tryptophan (responsible for bulge formation) to alanine results in 75 and 69% reduction of fucosylated product, respectively (W_7, W_{10} in Figure 5C). The second key interaction is predicted to take place between the AB loop (where the consensus motif $CX_{2-3}(S/T)CX_2G$ is located) and the conserved POFUT2 residues Asn51, Pro52, Pro53, Glu54, Leu58, Asp61, and Glu221 (Figure 7A). These residues, together with L224 that inserts its side chain exactly where the C strand starts to crossover the B strand, ensure *via* TSR backbone interactions that only the unique 3D motif at the very tip of the TSR module can undergo fucosylation. This structural motif is mainly defined by the length and conformation of the AB loop, which is held in place by the two disulphide bonds and which is encoded in the $CX_{2-3}(S/T)CX_2G$ sequon. Our observation explains experimental data where insertion of two additional glycines before the threonine (CX_{2-3}**GG**$(S/T)CX_2G$, changing the length of the loop) as well as removing the two SS-bonds, which together are responsible for pulling the C strand over the B strand and stabilization of the AB-loop conformation, completely abolished fucose attachment (Figure 5C).

Rational design of a minimal POFUT2 substrate resulted in the artificial mini-TSR molecule which we predicted to contain all necessary structural features for folding into the correct TSR fold needed for productive fucosylation. We found mini-TSR to be modified with the common glycan structures known from wild-type TSR, thereby confirming that it indeed can fold into the correct 3D AB-loop TSR structure and act as a POFUT2 substrate. This result not only defines a minimal POFUT2 substrate and validates our proposed binding mode but also brings new insight into folding and stability of TSR molecules. It shows that the correct disulphide bond pattern needed for the proper AB-loop conformation can be established with a minimal side-chain stacking unit composed of one tryptophan and one arginine residue only.

Having realized that POFUT2 substrate recognition is likely to be driven by the conserved TSR residues responsible for the unique layered TSR fold, we wondered how the substrate binding site is able to accept the large charge and size variation of amino acids on the remaining ~40 sequence positions. Strikingly, we found in our model of the POFUT2 TSR complex that out of the ~30 TSR residues building the upper half of the TSR (predicted to interact with POFUT2) 10 are conserved determining the TSR fold (in the central layer of stacking residues and SS-bonds) or are part of the consensus motif for fucosylation. Another nine residues are solvent exposed and likely not involved in POFUT2 interaction. At the 11 remaining TSR positions where wide sequence diversity is present we find large cavities in POFUT2, ready to accommodate side chains of different lengths or with different physicochemical properties (Figure 7B).
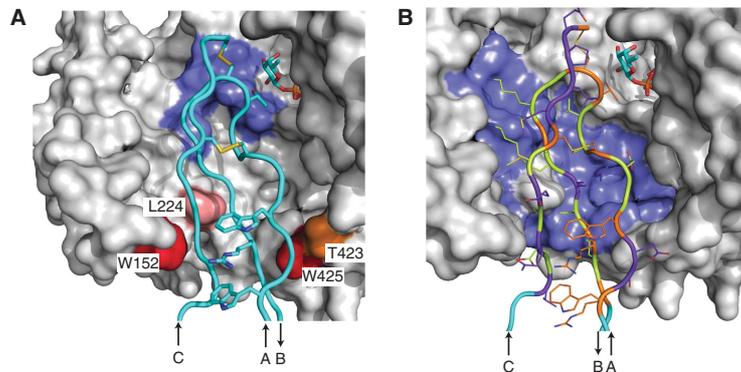
Figure 7 Structural details of the proposed TSR binding mode in human POFUT2. (**A**) Surface representation (grey) of POFUT2 with bound TSR4 (F-spondin) as ribbons (cyan) and the GDP-fucose substrate as sticks. POFUT2 residues predicted to recognize key elements of the TSR fold are coloured: W152 and W425 (red) scan the width of the substrate; T423 (orange) senses the bulge in the A strand and acts as a ruler; L224 (pink) scans the crossing over of the C strand; N51, P52, P53, E54, L58, D61, E221 (all in blue) recognize the correct conformation of the AB loop. Key structural residues of the TSR fold are displayed as sticks in cyan (atom colours). (**B**) The wide TSR sequence diversity of POFUT2 substrates can be explained by the proposed TSR binding mode. POFUT2 provides large cavities (solid surface, blue) for highly variable TSR residues (green) thereby tolerating side chains of different size and physicochemical properties at these positions. TSR side chains can also vary at positions predicted to be solvent exposed (purple) while TSR residues important for the fold and the fucosylation motif are strictly conserved (orange).
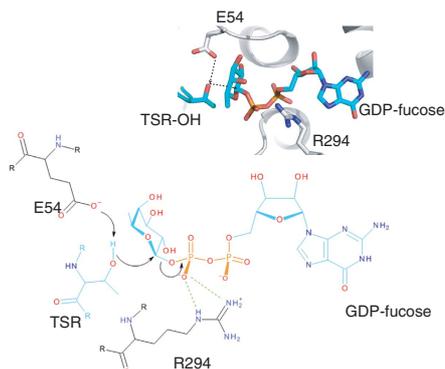

Figure 8 Proposed enzymatic mechanism for POFUT2-dependent TSR fucosylation. Chemical drawing of the enzymatic reaction. The carboxylate side chain of E54 deprotonates the TSR S/T hydroxyl group that attacks as a nucleophile the anomeric carbon of the GDP-fucose sugar donor. R294 activates the labile bond by forming a salt bridge with the β-phosphate group. The arrangement of critical catalytic residues in the model of the ternary POFUT2 GDP-fucose TSR complex is shown. Distances: E54 carboxylate to TSR-OH 3.1 Å, TSR-OH to GDP-fucose anomeric carbon 3.3 Å.

Our model of TSR–POFUT2 interaction is also compatible with tryptophan *C*-mannosylation present on many TSR (Hofsteenge *et al*, 2001; Tan *et al*, 2002) as the CWR layer with a potentially attached α-mannosyl residue is facing solvent and would therefore not be involved in POFUT2 interaction. Finally, our structural data also explain why POFUT2 is specific for TSR modules and why EGF repeats, the other known protein module to contain *O*-fucose modifications, are not accepted as substrate. EGF repeats simply do not have the critical 3D TSR elements (e.g., AB-loop conformation) needed for binding to the POFUT2 active site.

Catalytic mechanism of POFUT2

Structural and biochemical data of wild-type and mutant POFUT2 allowed us to suggest a catalytic mechanism for the fucosyltransferase reaction (Figure 8). POFUT2 belongs to the GT-B family of inverting glycosyltransferases where the key catalytic residue acts as a general base responsible for deprotonation of the nucleophile functional group of the sugar acceptor (Lairson *et al*, 2008). Only the fully conserved residues E54 and D297 are in close proximity of the TSR S/T hydroxyls that undergo fucose attachment. Both residues are located right at the entry of the GDP-fucose binding site, E54 on a surface exposed loop of the N-terminal domain, and D297 in a long loop in the C-terminal domain. The E54 carboxylate side chain is closer to the anomeric carbon and is freely accessible while D297 is located slightly further away and its side chain is sandwiched between the two guanidinium groups of R294 and R303 in the native POFUT2 structure reducing its mobility and, importantly, lowering its pKa. Complete loss of enzymatic activity for the E54A POFUT2 mutant (the D297A mutant retained ~16% activity), and the fact that we were able to obtain a structure of the non-hydrolysed sugar donor complex with the same catalytically inactive E54A mutant, are strong arguments for E54 to be the catalytic base of POFUT2. D297, on the other hand, seems to be needed to correctly orient the R294 side chain for binding of the GDP-fucose diphosphate group. Positioning of aspartate, glutamate, or histidine as the catalytic base on the N-terminal domain facing the sugar donor binding site (as seen in POFUT2) is known from other inverting GT-B family members like WaaC (Grizot *et al*, 2006), T4 phage glucosyltransferase BGT (Larivière *et al*, 2003), *H. pylori* fucosyltransferase α1,3-FucT (Sun *et al*, 2007), *E. coli* MurG (Hu *et al*, 2003), or *Vv*GT1 (Offen *et al*, 2006) of which structures have been solved and the catalytic base residue has been identified (Supplementary Figure S6). During E54-dependent deprotonation, the TSR S/T nucleophile can attack the anomeric

carbon of the GDP-fucose to form the new glycosidic bond with inverted configuration followed by release of the two products. The fucose ring has been modelled in the 1C4 conformation resulting in a better fit to the 3.4 Å electron density if compared with the 4C1 conformation. Interestingly, the two conformations differ only in the ring flip and an ∼38° conformational change around the P1-O2P bond. Both ring conformations would need considerable distortion for in-line nucleophilic attack geometry seen in other inverting glycosyltransferases such as *Vv*GT1 (Offen *et al*, 2006). Therefore, the reactive fucose ring conformation will most likely be established upon binding of the TSR sugar acceptor. Activation of the labile phosphoester bond of the sugar donor is mainly achieved by the charged residue R294 that binds the diphosphate group and by the positive dipoles of two helices near the diphospho and the pentose moiety. Mutation of R294 to alanine in the human protein as well as the corresponding mutation in the *C. elegans* orthologue (Canevascini *et al*, 2006) resulted in complete loss of enzymatic activity consistent with its role in directly binding the GDP-fucose diphosphates. A multiple sequence alignment of different fucosyltransferases reveals that R294 is also conserved in POFUT1 and FUT8 (Martinez-Duncker *et al*, 2003) where mutation of this residue also abolishes enzymatic activity (Takahashi *et al*, 2000; Okajima *et al*, 2005). While the vast majority of inverting GT-B family glycosyltransferases are metal-ion independent there are three family members for which metal ions significantly enhance activity: T4 phage BGT (Moréra *et al*, 2001), hamster POFUT1 (Wang and Spellman, 1998), as well as human POFUT2 which we described here (Figure 3C). Despite many soaking and co-crystallization trials with Mn^{2+} we were not able to localize the cation in anomalous difference Fourier electron density maps in the POFUT2 apo structure. In addition, we only obtained crystals of the GDP-fucose complex when we added EDTA to the crystallization buffer to remove all remaining metal ions from the protein. These findings, together with the fact that the enzyme retains ∼5% enzymatic activity in the presence of EDTA, point to a role of metal ions in product release. This is known from other GT-B fold glycosyltransferases like T4 phage BGT where Mn^{2+} complexes the pyrophosphate group of the UDP product at the place occupied by the sugar moiety in the UDP-Glc complex structure (Moréra *et al*, 2001). Similarly, the Mn^{2+} binding site has been identified in the crystal structure of a *C. elegans* POFUT1 GDP Mn^{2+} complex where the ion also binds the pyrophosphate of the GDP product exactly where the fucose moiety is placed in the GDP-fucose structure. However, in the latter case the authors do not relate this finding to the metal-dependent enzymatic activity (Lira-Navarrete *et al*, 2011). Other GT-B fold glycosyltransferases like FUT8 seem to have metal-independent ways of nucleotide diphosphate release and do not need divalent cations to reach full activity (Ihara *et al*, 2006). Kinetic experiments yielded a k_{cat} of 144 per minute for the POFUT2 enzyme, a value that is comparable with published results from other glycosyltransferases (Ihara *et al*, 2006; Sun *et al*, 2007). K_M values on the other hand differ with 9.8 μM for GDP-fucose and 29.5 μM for TSR4. A K_M in the low micromolar range for the sugar donor is common for GT-B fold glycosyltransferases (Jeanneau *et al*, 2004; Grizot *et al*, 2006; Ihara *et al*, 2006) while the K_M value for TSR4 is rather low.

In summary, POFUT2 seems to utilize a well-established catalytic mechanism for GT-B fold inverting glycosyltransferases with E54 acting as general base. This is in contrast to the suggested mechanism in *C. elegans* POFUT1 where no residue acting as catalytic base could be identified and the reaction after cleavage of the glycosidic bond (facilitated by R240) proceeds *via* an oxocarbenium-phosphate ion pair transition state and subsequent attack of the acceptor OH group at the anomeric carbon (Lira-Navarrete *et al*, 2011).

Substrate specificity of protein glycosyltransferases
The mechanism of glycan transfer to a protein or peptide acceptor has for a long time been poorly understood. It was largely unknown how short sequence motifs present in polypeptides of wide sequence diversity can be modified by a position-specific enzyme. It was only recently that crystal structures of glycosyltransferases in complex with acceptor peptides gave insight into substrate specificity and how a few key elements present in the recognition sequons enable glycosylation of specific residues. Structures of acceptor peptide complexes are now available for all glycosyltransferase families and reveal surprising similarities: GALNT10 (GT-A) (Fritz *et al*, 2006), OGT (GT-B) (Lazarus *et al*, 2011), and PglB (GT-C) (Lizak *et al*, 2011) all recognize glycosylation sequons in flexible unstructured protein regions and bind the substrate peptide mainly *via* backbone interactions. Many structured water molecules are present providing an adaptable protein interface ready to accommodate a wide range of polypeptides with side chains of different size, charge, and polarity. Sequon specificity is most clearly defined in PglB (Asn-X-Ser/Thr) where a WWD protein motif binds the Ser/Thr residue side chain at the +2 position and thereby positions the asparagine correctly for N-linked glycosylation. A similar mode is used in GALNT2 where the proline at the +3 position is specifically bound to position the Ser/Thr correctly in the active site. For OGT on the other hand, no *O*-GlcNAcylation motif has been identified so far but a preference for residues that form an extended peptide conformation near the glycosylation site can be explained by the binding mode of the peptide as seen in the crystal structures.

Here, we present a completely novel mode of substrate recognition for protein glycosyltransferases that explains why the specific fucosylation consensus motif $CX_{2-3}(S/T)CX_2G$ (Hofsteenge *et al*, 2001) can only be modified in the context of a properly folded TSR protein domain and how these structural constraints are not in conflict with the wide sequence diversity present on fucosylated TSR. POFUT2 has evolved to specifically recognize unique 3D structural TSR elements, which are defined by a few strictly conserved residues and the consensus motif itself. This allows for wide sequence diversity at all the other TSR positions, probably reflecting the diverse biological functions of proteins containing the TSR module.

Materials and methods

Crystallization, data collection and structure determination
POFUT2 crystals were grown at 4°C by the vapour diffusion method in 96-well crystallization plates by mixing 0.1 μl of POFUT2 protein solution (7.5 mg/ml) with 0.1 μl of crystallization buffer (20 mM Tris–HCl, pH 8.5, 12% PEG 20000). For native data collection, crystals were soaked in mother liquor containing 25% ethylene glycol and frozen in liquid nitrogen. For heavy atom derivatization, crystals were soaked in mother liquor containing 5 mM of K_2PtCl_4 for 6 min. Diffraction data were collected at beamlines X06DA and X10SA at the Swiss Light Source synchrotron in Villigen, Switzerland. Diffraction images were processed and scaled with HKL-2000 (Otwinowski *et al*, 1997). The structure of POFUT2 was solved by the SIRAS method using two platinum sites per molecule identified in SHELXD (Sheldrick, 2008). Heavy atom sites were used for phase calculation and refinement of sites in Sharp (Bricogne *et al*, 2003) followed by density modification using Solomon (Abrahams and Leslie, 1996). Phases from density modification were then used for automatic model building in PHENIX (Adams *et al*, 2010) and in BUCCANEER (Cowtan, 2006) followed by manual completion of the model using COOT (Emsley *et al*, 2010). Structures were refined by the crystallographic simulated annealing routine followed by individual *B*-factor refinement in PHENIX including NCS restraints.

For crystals of the POFUT2 GDP-fucose complex, 13.6 mg/ml of high mannose type E54A POFUT2 was incubated in protein buffer containing 3.5 mM EDTA and 1 mM GDP-fucose (Sigma) for 30 min on ice before setting up the crystallization experiment. Crystals were grown at 20°C by the vapour diffusion method in 96-well crystallization plates by mixing 0.25 μl of protein solution with 0.25 μl of crystallization buffer (20% PEG 3350, 0.2 M NaSCN). Crystals of the complex were cryoprotected and frozen as described for native crystals. The structure of the POFUT2 GDP-fucose complex was solved by molecular replacement using PHASER (McCoy *et al*, 2007) with the native POFUT2 structure as search model and subsequent refinement in PHENIX. Clear mF_O-DF_C difference electron density for the missing GDP-fucose moieties was visible in the active sites for all four molecules in the a.u. The structures of native POFUT2 and of the E54A-POFUT2 GDP-fucose complex were refined by several rounds of manual rebuilding in COOT followed by refinement in PHENIX. Of the four E54A-POFUT2 molecules present in the a.u. only chains A, B, and C have full occupancy whereas chain D is partially occupied (or has high mobility). This results in less well-defined $2mF_O$-DF_C electron density for chain D. All crystal structures were validated using MolProbity (Chen *et al*, 2010) and COOT. Structural images for figures were prepared with PyMOL (http://pymol.sourceforge.net/). Atomic coordinates and structure factors have been deposited in the PDB with entry codes 4AP5 (apo) and 4AP6 (GDP-fucose complex).

Supplementary data
Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abrahams JP, Leslie AGW (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52:** 30–42

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221

Akiyama M, Takeda S, Kokame K, Takagi J, Miyata T (2009) Crystal structures of the noncatalytic domains of ADAMTS13 reveal multiple discontinuous exosites for von Willebrand factor. *Proc Natl Acad Sci USA* **106:** 19274–19279

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

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98:** 10037–10041

Bjoern S, Foster DC, Thim L, Wiberg FC, Christensen M, Komiyama Y, Pedersen AH, Kisiel W (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:** 11051–11057

Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* **59:** 2023–2030

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54:** 905–921

Brzezinski K, Stepkowski T, Panjikar S, Bujacz G, Jaskolski M (2007) High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor. *Acta Biochim Pol* **54:** 537–549

Buko AM, Kentzer EJ, Petros A, Menon G, Zuiderweg ER, Sarin VK (1991) Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator. *Proc Natl Acad Sci USA* **88:** 3992–3996

Canevascini S, Kozma K, Grob M, Althaus J, Klein D, Chiquet-Ehrismann R, Hofsteenge J (2006) *Protein O-Fucosylation is Important for Normal Distal Tip Migration* In:European Worm Meeting, Hersonissos, Crete, Greece

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37:** D233–D238

Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12–21

Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62:** 1002–1011

Du J, Takeuchi H, Leonhard-Melief C, Shroyer KR, Dlugosz M, Haltiwanger RS, Holdener BC (2010) O-fucosylation of thrombospondin type 1 repeats restricts epithelial to mesenchymal transition (EMT) and maintains epiblast pluripotency during mouse gastrulation. *Dev Biol* **346:** 25–38

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of coot. *Acta Crystallogr D Biol Crystallogr* **66:** 486–501

Freeze HH (2007) Congenital disorders of glycosylation: CDG-I, CDG-II, and beyond. *Curr Mol Med* **7:** 389–396

Fritz TA, Raman J, Tabak LA (2006) Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:Polypeptide α-N-acetylgalactosaminyltransferase-2. *J Biol Chem* **281:** 8613–8619

Gonzalez de Peredo A, Klein D, Macek B, Hess D, Peter-Katalinic J, Hofsteenge J (2002) C-mannosylation and *O*-fucosylation of thrombospondin type 1 repeats. *Mol Cell Proteomics* **1:** 11–18

Grizot S, Salem M, Vongsouthi V, Durand L, Moreau F, Dohi H, Vincent S, Escaich S, Ducruix A (2006) Structure of the *Escherichia coli* heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. *J Mol Biol* **363:** 383–394

Harris RJ, Ling VT, Spellman MW (1992) O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. *J Biol Chem* **267:** 5102–5107

Harris RJ, Spellman MW (1993) O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3:** 219–224

Hess D, Keusch JJ, Oberstein SA, Hennekam RC, Hofsteenge J (2008) Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective O-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* **283:** 7354–7360

Hofsteenge J, Huwiler KG, Macek B, Hess D, Lawler J, Mosher DF, Peter-Katalinic J (2001) C-mannosylation and O-fucosylation of the Thrombospondin Type 1 Module. *J Biol Chem* **276:** 6485–6498

Hofsteenge J, Mueller DR, de Beer T, Loeffler A, Richter WJ, Vliegenthart JFG (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* **33:** 13524–13530

Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38:** W545–W549

Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc Natl Acad Sci USA* **100:** 845–849

Ihara H, Ikeda Y, Taniguchi N (2006) Reaction mechanism and substrate specificity for nucleotide sugar of mammalian α1,6-fucosyltransferase–a large-scale preparation and characterization of recombinant human FUT8. *Glycobiology* **16:** 333–342

Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, Nakagawa A, Taniguchi N (2007) Crystal structure of mammalian α1,6-fucosyltransferase, FUT8. *Glycobiology* **17:** 455–466

Jaeken J, Matthijs G (2007) Congenital disorders of glycosylation: a rapidly expanding disease family. *Annu Rev Genomics Hum Genet* **8:** 261–278

Jeanneau C, Chazalet V, Augé C, Soumpasis DM, Harduin-Lepers A, Delannoy P, Imberty A, Breton C (2004) Structure-function analysis of the human sialyltransferase ST3Gal I. *J Biol Chem* **279:** 13461–13468

Kozma K, Keusch JJ, Hegemann B, Luther KB, Klein D, Hess D, Haltiwanger RS, Hofsteenge J (2006) Identification and characterization of a β1,3-glucosyltransferase that synthesizes the Glc-β1,3-Fuc disaccharide on thrombospondin type 1 repeats. *J Biol Chem* **281:** 36742–36751

Kubota T, Shiba T, Sugioka S, Furukawa S, Sawaki H, Kato R, Wakatsuki S, Narimatsu H (2006) Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide α-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol* **359:** 708–727

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77:** 521–555

Larivière L, Gueguen-Chaignon V, Moréra S (2003) Crystal structures of the T4 phage β-glucosyltransferase and the D100A mutant in complex with UDP-glucose: glucose binding and identification of the catalytic base for a direct displacement mechanism. *J Mol Biol* **330:** 1077–1086

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948

Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469:** 564–567

Lesnik Oberstein SA, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, Breuning MH, Hennekam RC (2006) Peters Plus syndrome is caused by mutations in β3GALTL, a putative glycosyltransferase. *Am J Hum Genet* **79:** 562–566

Lira-Navarrete E, Valero-González J, Villanueva R, Martínez-Júlvez M, Tejero T, Merino P, Panjikar S, Hurtado-Guerrero R (2011) Structural insights into the mechanism of protein O-fucosylation. *PLoS One* **6:** e25365

Lizak C, Gerber S, Numao S, Aebi M, Locher KP (2011) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474:** 350–355

Luo Y, Koles K, Vorndam W, Haltiwanger RS, Panin VM (2006a) Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *J Biol Chem* **281:** 9393–9399

Luo Y, Nita-Lazar A, Haltiwanger RS (2006b) Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J Biol Chem* **281:** 9385–9392

Luther KB, Haltiwanger RS (2009) Role of unusual O-glycans in intercellular signaling. *Int J Biochem Cell Biol* **41:** 1011–1024

Maita N, Nyirenda J, Igura M, Kamishikiryo J, Kohda D (2010) Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J Biol Chem* **285:** 4941–4950

Martinez-Duncker I, Mollicone R, Candelier JJ, Breton C, Oriol R (2003) A new superfamily of protein-O-fucosyltransferases, α2-fucosyltransferases, and α6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. *Glycobiology* **13:** 1C–5C

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658–674

Moréra S, Larivière L, Kurzeck J, Aschke-Sonnenborn U, Freemont PS, Janin J, Rüger W (2001) High resolution crystal structures of T4 phage β-glucosyltransferase: induced fit and effect of substrate and metal binding. *J Mol Biol* **311:** 569–577

Nishimura H, Takao T, Hase S, Shimonishi Y, Iwanaga S (1992) Human factor IX has a tetrasaccharide O-glycosidically linked to serine 61 through the fucose residue. *J Biol Chem* **267:** 17520–17525

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J* **25:** 1396–1405

Okajima T, Xu A, Lei L, Irvine KD (2005) Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding. *Science* **307:** 1599–1603

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology, Vol. 276: Macromolecular Crystallography*, Carter Jr CW, Sweet RM (eds) Part A, pp 307–326. New York: Academic Press

Pääkkönen K, Tossavainen H, Permi P, Rakkolainen H, Rauvala H, Raulo E, Kilpeläinen I, Güntert P (2006) Solution structures of the first and fourth TSR domains of F-spondin. *Proteins* **64:** 665–672

Rana NA, Haltiwanger RS (2011) Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Curr Opin Struct Biol* **21:** 583–589

Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, Majerus EM (2007) O-fucosylation is required for ADAMTS13 secretion. *J Biol Chem* **282:** 17014–17023

Sato T, Sato M, Kiyohara K, Sogabe M, Shikanai T, Kikuchi N, Togayachi A, Ishida H, Ito H, Kameyama A, Gotoh M, Narimatsu H (2006) Molecular cloning and characterization of a novel human β1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain. *Glycobiology* **16:** 1194–1206

Sheldrick G (2008) A short history of SHELX. *Acta Crystallogr A* **64:** 112–122

Stahl M, Uemura K, Ge C, Shi S, Tashima Y, Stanley P (2008) Roles of Pofut1 and O-fucose in mammalian Notch signaling. *J Biol Chem* **283:** 13638–13651

Stanley P (2007) Regulation of Notch signaling by glycosylation. *Curr Opin Struct Biol* **17:** 530–535

Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin CN, Wang AH, Lin CH (2007) Structure and mechanism of Helicobacter pylori fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem* **282:** 9973–9982

Takahashi T, Ikeda Y, Tateishi A, Yamaguchi Y, Ishikawa M, Taniguchi N (2000) A sequence motif involved in the donor substrate binding by α1,6-fucosyltransferase: the role of the conserved arginine residues. *Glycobiology* **10:** 503–510

Tan K, Duquette M, Liu JH, Dong Y, Zhang R, Joachimiak A, Lawler J, Wang JH (2002) Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. *J Cell Biol* **159:** 373–382

Tossavainen H, Pihlajamaa T, Huttunen TK, Raulo E, Rauvala H, Permi P, Kilpeläinen I (2006) The layered fold of the TSR domain of P. falciparum TRAP contains a heparin binding site. *Protein Sci* **15:** 1760–1768

Wang LW, Dlugosz M, Somerville RPT, Raed M, Haltiwanger RS, Apte SS (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:** 17024–17031

Wang Y, Spellman MW (1998) Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *J Biol Chem* **273:** 8112–8118

Supplementary Information to:

Structure of Human POFUT2: Insights into Thrombospondin Type 1 Repeat Fold and *O*-Fucosylation

Chun-I Chen, Jeremy J. Keusch, Dominique Klein, Daniel Hess, Jan Hofsteenge, Heinz Gut

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

Supplementary materials and methods

Expression and purification of wild-type human POFUT2

Full length (lacking the 21 residue long N-terminal signal peptide) human POFUT2 was cloned into pSecTagB vector (Invitrogen) with an N-terminal FLAG tag for purification. HEK293T cells were cultured as an adherent monolayer in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS, Sigma) and 2 mM glutamine. At 80% confluency, cells were transfected with polyethyleneimine according to the manufacturer's instructions (jetPEI; Polyplus-transfection Inc.). After 48 hours, selection of stable lines was started by passaging cells into medium containing 100 μg/ml of Zeocin (Invitrogen). Resistant single clones were selected and expanded. The clone with highest expression of POFUT2 was selected for large-scale protein expression.

For a large-scale protein expression, 20 dishes (150 x 15mm, BD Falcon) of confluent stable HEK293T cells expressing FLAG-Δ21-POFUT2 were harvested and split into 200 dishes with the culture medium containing additional 5 μM of kifunensine (Chang et al, 2007). After seven days, the conditioned medium was harvested and filtered through a 0.22 μm filter (Stericup, Millipore). The conditioned medium and anti-FLAG M2 resin (Sigma) were mixed in a two liter roller bottle (Greiner Bio-One) overnight at 4°C. M2 resin was collected in a 2.5 x 10 cm glass column (Bio-Rad) and washed with 50 ml of TBS buffer (20 mM Tris, pH 7.4, 150 mM NaCl), followed by 50 ml of 20 mM Tris-HCl, pH 7.4, 1 M NaCl and re-equilibrated in 50 ml of TBS. FLAG-Δ21-POFUT2 was eluted by 100 μg/ml of three times FLAG peptide in 20 mM Tris-HCl, pH 7.4, 300 mM NaCl.

1

To remove N-glycans, 30 kU of Endo H_f (NEB) was added to 1 mg of eluted protein for 3 hours at 37ºC. The FLAG tag was removed with enterokinase (NEB) at 23ºC for 16 hours. EndoH$_f$ was removed by amylose resin (NEB) according to the manufacturer's instructions, and the flow through was loaded onto a Superdex 200 16/60 column, which was equilibrated in 10 mM Tris-HCl, pH 8.8, 300 mM NaCl, 10 mM $MgCl_2$. Fractions with POFUT2 were pooled and concentrated by centrifugation using an Amicon Ultra-4 PL-10 device (Millipore) to 6.5 - 8 mg/ml, aliquoted and stored at -80ºC.

Expression and purification of human E54A POFUT2

The stable HEK293T cell line expressing E54A-POFUT2 was established as for the wild-type POFUT2 except that the transfection was performed with Lipofectamine (Invitrogen). For large-scale expression, 20 dishes (150 x 15mm, BD Falcon) of confluent HEK293T cells expressing His$_6$-FLAG-3C-Δ36-POFUT2 were harvested and split into 200 dishes with DMEM supplemented with 10% FCS and 2 mM glutamine. When the dishes were confluent, the medium was switched to Pro293a medium (Lonza) supplemented with 5 μM kifunensine and 2 mM glutamine. After 96 hours expression medium was harvested and filtered (Stericup, Millipore). The medium was incubated with Ni-NTA agarose (Qiagen) in a two liter roller bottle (Greiner Bio-One) overnight at 4ºC. The resin was collected by centrifugation at 100 x g for 5 minutes at 4ºC. After washing with three gel volumes of 20 mM Tris-HCl, pH 8.5, 500 mM NaCl, 20 mM imidazole, the protein was eluted in the same buffer but containing 500 mM imidazole. Eluate was loaded onto a Superdex 200 16/60 column, which was equilibrated in 10 mM Tris-HCl pH 8.5, 200 mM NaCl. Fractions with target protein were pooled and buffer exchanged into 10 mM Tris-HCl pH 8.5, 50 mM NaCl by gel filtration on a HiPrep 16/10 desalting column. The sample was further purified by Mono Q 5/50 anion exchange with a gradient from 0% buffer B to 50% buffer B in 30 minutes (buffer A: 10 mM Tris-HCl, pH 8.5, 50 mM NaCl, Buffer B: 10 mM Tris-HCl, pH 8.5, 1 M NaCl). The protein was concentrated to 20 mg/ml using a Amicon Ultra-4 PL-10 device (Millipore) in 20 mM Tris pH 8.5, 225 mM NaCl and stored at -80ºC.

Site-directed mutagenesis of POFUT2

POFUT2 mutants were generated by the QuickChange method (Stratagene). Wild-type and mutant POFUT2 were cloned into a pSecTag vector with N-terminal His$_6$-FLAG-3C

2

tag. HEK293T cells were cultured in DMEM supplemented with 10% FCS and 2 mM glutamine to 80% confluency before transfection. The cells were transfected by Lipofectamine (Invitrogen) using the manufacturer's protocol. Eight hours after transfection, the medium was changed to Pro293a medium (Lonza) supplemented with 5 μM kifunensine and 2 mM glutamine. Ninety-six hours later, the medium was collected and incubated with anti-FLAG M2 resin overnight on roller at 4ºC. Proteins were eluted by 100 μg/ml of three times FLAG peptide in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl. POFUT2 mutants were aliquoted and stored at -80ºC before testing the enzymatic activity in the *in vitro* POFUT2 fucosyltransferase assay.

Expression and Purification of TSR4 from F-spondin

Rat F-spondin TSR4 containing a cleavable N-terminal His-tag was cloned into a pOPINF vector (Berrow et al, 2007) and used to transform Rosetta2 expression competent *E. coli* cells. A single colony of *E. coli* (Rosetta 2) cells was grown overnight in 20 ml of GS96 medium containing 1% glucose. The overnight culture was diluted 1:100 into 0.5 L of auto-induction medium (Studier, 2005) containing 34 μg/ml chloramphenicol and 50 μg/ml carbenicillin. Cells were grown at 37ºC, 225 rpm for 4 hours, followed by reduction of the temperature to 20ºC for another 20 hours. Cells were harvested by centrifugation at 6500 x g for 30 minutes. The cell pellet was resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.2% Tween20, Complete Protease Inhibitor) and lysed using a high pressure cell disruptor. The lysate was cleared by centrifugation and mixed with Ni-NTA agarose (Qiagen). The resin was washed with five gel volumes of lysis buffer and the TSR4 was eluted with 50 mM Tris-HCl, pH 7.5, 500 mM NaCl and 500 mM of imidazole. The eluate was loaded onto a Superdex-75 16/60 PG column (GE Healthcare) in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. Fractions containing TSR4 were pooled and buffer exchanged into 20 mM Tris-HCl, pH 7.5 10 mM NaCl by gel filtration on a HiPrep 26/10 desalting column. TSR4 was further purified by ion-exchange chromatography on a MonoQ 5/50 column (GE Healthcare) and eluted in a linear gradient of 0-500 mM NaCl over 50 minutes. Pure TSR4 was aliquoted and stored at -80ºC in 20 mM Tris, pH 7.5, 110 mM NaCl.

3

In vitro **LC-MS POFUT2 activity assay**

The reaction mixture for the POFUT2 activity assay contained 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM MnCl$_2$, 10 μM GDP-fucose, 10 μM TSR4 and 1 ng of recombinantly expressed pure POFUT2 in a volume of 10 μl. After 10 minutes (corresponding to less than 25% substrate consumption) incubation in 0.5 ml protein LoBind tubes (Eppendorf) at 37ºC, the reaction was quenched by acidification with 40 μl of cold 0.2% trifluoroacetic acid (TFA) and 2% acetonitrile. Fucosylated and non-fucosylated TSR4 was separated by HPLC using a C18 column with a linear gradient from 37% buffer B to 43% buffer B in 15 minutes (buffer A: 2% CH$_3$CN, 0.1% TFA; buffer B: 80% CH$_3$CN, 0.1% TFA). The substrate and product of the reaction were quantified by integrating the peak area at 280 nm or by ion count from mass spectrometry (4000 Q TRAP, Applied Biosystems) in positive ion mode. Both approaches gave the same result.

In vitro LC-MS POFUT2 activity assay for kinetics

Steady-state kinetic parameters of POFUT2 at different concentrations of substrates were determined in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM MnCl$_2$ and 2 ng of recombinantly expressed POFUT2 in a volume of 10 μl for 20 minutes. The turnover was first measured at fixed TSR4 concentrations ranging from 3.3, 10, 20, 40, 80 to 200 μM with a stepwise increase of the GDP-fucose concentration from 2.5, 5, 10, 20, 40, 80 to 160 μM (Fig. S2B). Then, the turnover was determined with fixed concentration of GDP-fucose at 2.5, 5, 10, 20, 40, 80 and 160 μM with the stepwise increased concentration of TSR4 from 3.3, 10, 20, 40, 80, and 200 μM (Fig. S2C). Quantification and analyses of the reaction product by mass-spectrometry is described above. Kinetic parameters were computed with the GraphPad Prism software by fitting the data to the Michaelis–Menten equation.

HEK293T cell POFUT2 activity assay using wild-type and mutant TSR4 substrates

Wild-type and mutant TSR4 of rat F-spondin were cloned into pSecTagB vector (Invitrogen) with TEV-myc-His$_6$ at the C-terminus. HEK293-T cells were grown in DMEM supplemented with 10% FCS and 2 mM glutamine to 80 % confluence before transfection. Cells were transfected with Lipofectamine (Invitrogen) according to the manufacturer's protocol. After 60 hours of expression in Opti-MEM (Invitrogen), the

4

medium was harvested and dialyzed against 20 mM Tri-HCl, pH 8.0, 100 mM NaCl. For purification, 200 µl of Co^{2+} TALON resin (Clontech) was incubated with 50 ml of medium overnight on a roller at 4ºC. The resin was washed three times with the dialysis buffer containing 4 mM imidazole before elution with the same buffer containing 300 mM imidazole. The eluate was dialyzed against 50 mM ammonium carbonate using a Slide-A-Lyzer device (3.5 KDa MWCO). Samples were reduced, carboxymethylated (Krieg et al, 1998) and digested with Lys-C or Lys-C/Asp-N (for mutant C(13,46)A and C(13,46,17,51)A). The digests were injected onto a Vydac column C18 (1 x 250 mm) for high pressure liquid chromatography-mass spectrometry (HPLC-MS) analysis. The analysis was performed on an 1100 CapLC system (Agilent Technologies) that was connected to a API 300 or a 4000 QTRAP (AB sciex) mass spectrometer. The peptides were separated in 0%-70% of buffer B over 120 minutes (buffer A: 2% CH_3CN, 0.1% TFA; buffer B: 80% CH_3CN, 0.1% TFA) and quantification was done by integration of peak size in the UV of 280 nm or 214 nm. *O*-fucosylated peptides were identified and confirmed by their mass and neutral losses (Hess et al, 2008; Hofsteenge et al, 2001).

In vitro POFUT2 activity assay using mini-TSR substrate

The mini-TSR DNA was synthesized by Sigma and cloned into a pOPINJ vector with an N-terminal His_6-GST-3C tag. After transformation and plating out on LB Agar, a single colony of *E. coli* (BL21 DE3) was picked and inoculated into a 96-well deep-well block with 1.5 ml of GS96 medium with 1% glucose, 0.05% glycerol and the culture was incubated at 37ºC (225 rpm) overnight. The overnight culture was diluted 1:50 into 2.5 ml of the same fresh medium in a 24-well deep-well block and incubated at 37ºC with shaking at 225 rpm. At OD_{600} 0.6, the culture was cooled to 20ºC before IPTG (0.5 mM) induction. Following induction, cells were grown for 20 hours at 225 rpm at 20ºC. Cells were harvested by centrifugation and the cell pellet was lysed with 1 mg/ml of lysozyme and 3 unit/ml of Benzonase. After centrifugation, the supernatant was incubated with Ni-NTA agarose and the protein eluted with 50 mM Tris-HCl, pH 7.5, 500 mM NaCl and 300 mM imidazole. The eluate was incubated overnight at 4ºC with rhinovirus 3C protease for cleavage of the His_6-GST tag. The sample was diluted ten times and incubated with Ni-NTA agarose and the flow through was collected for the *in vitro* fucosylation assay. 38 µl of wild-type TSR4 and mini-TSR4 from the flow through were incubated with 250 µM of GDP-fucose and 5 ng of Δ36-POFUT2 for 60 minutes at 37ºC.

5

The reaction was quenched by acidification with TFA to a final pH of 2. The mixture was analyzed by NanoLC-MS using a Magic C18 HPLC column (75 μm × 10 cm; Swiss BioAnalytics) with a 1200 NanoLC system (Agilent Technologies) connected to a LTQ Orbitrap Velos (Thermo Fisher Scientific). The peptides were loaded onto a peptide captrap (Michrom BioResources) at a flow rate of 10 μl/min for 5 minutes in 10% A. They were eluted at a flow rate of 400 nl/min in 30 min with a linear gradient of 10–80% B in A (A: 0.1% formic acid, 0.005% TFA, 2% acetonitrile in H_2O; buffer B : 0.1% formic acid, 0.005% TFA, 80% acetonitrile in H_2O. The MS were acquired in positive ion mode with a resolution of 60000 at m/z 400 according to the manufacturer's recommendation.

HEK293T cell POFUT2 activity assay using mini-TSR substrate

Mini-TSR DNA was cloned into a pSecTag vector with an N-terminal His_6-FLAG-3C tag. HEK293T cells were transfected as described for E54A-POFUT2. Eight hours after transfection, the medium was changed to Pro293a medium (Lonza) supplemented with 2 mM glutamine. Ninety-six hours after transfection the medium was harvested and filtered. Conditioned medium was incubated with Ni-NTA agarose and the mini-TSR protein eluted in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 300 mM imidazole. The mass of the intact mini-TSR was determined by NanoLC-MS as described above. The mini-TSR4 peptide (GPWSDCSVTCGK) for glycan analysis was obtained by proteolytic digestion of His_6-FLAG-mini-TSR with rhinovirus 3C protease (removal of the N-terminal tag) and Lys-C protease, followed by reduction/alkylation. The GPWSDCSVTCGK peptide was analyzed on the same LC-MS system but using a gradient from 0%B-45%B in 30 min. The 2+ ions of this peptide with m/z 912.36 were fragmented with HCD using an isolation width of the precursor of m/z 2.0 and normalized collision energies of 10%, 15%, 20%, 25% and 30% respectively. All fragment spectra containing signals for the peptide were summed up (Fig. 6E).

Influence of glycan composition, metal ions, and N-terminal truncations on POFUT2 activity

Enzymatic activity of different POFUT2 N-glycosylation types was assessed using a modified form of the assay developed by Wang and colleagues (Wang et al, 1996). Reaction mixtures for assaying enzymatic activity contained 100 mM imidazole, pH 6.8,

6

5 mM MnCl$_2$, protease inhibitor cocktail (Roche), 20 μM TSR4, 20 μM GDP-fucose, GDP-[2-^{3}H]fucose (300 nCi/μl) and 0.1 ng of different glycoforms of Δ36-POFUT2 in a volume of 25 μl. After incubation for 1h at 37ºC, the reaction was terminated by adding 200 μl of 0.2% ice-cold trifluoroacetic acid. Radioactive substrate and product were separated on a C8 reverse phase Sep-Pak Vac cartridge as described (Kozma et al, 2006). Incorporation of [^{3}H]fucose into the acceptor substrate was determined by scintillation counting.

For measurements of POFUT2 enzymatic activity in the presence of different metal ions, purified Δ36-POFUT2 with complex type N-glycans was pre-incubated with 2 mM EDTA to remove remaining metal ions. The reactions was started by adding GDP-fucose to the reaction mixture containing 1 ng POFUT2, 20 μM TSR4, 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA and 20 mM of either MnCl$_2$, MgCl$_2$, ZnCl$_2$, or CaCl$_2$. The reaction was terminated by adding 37.5 μl of 0.2% trifluoroacetic acid, 2% CH$_3$CN and the formation of fucosylated TSR4 was analysed by LC-MS as described above.

For POFUT2 activity with different N-terminal truncations, Δ36-POFUT2 or Δ21-POFUT2 with complex type N-glycans was incubated with 20 μM TSR4, 20 μM GDP-fucose, 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 5 mM MgCl$_2$. Reactions were terminated by adding 37.5 μl of 0.2% trifluoroacetic acid, 2% CH$_3$CN and the formation of fucosylated TSR4 was analysed with LC-MS as described above.

Supplementary references

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **35:** e45

Brzezinski K, Stepkowski T, Panjikar S, Bujacz G, Jaskolski M (2007) High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor. *Acta Biochim Pol* **54:** 537-549

Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI, Dwek RA, Jones EY, Owens RJ, Davis SJ (2007) Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15:** 267-273

Grizot S, Salem M, Vongsouthi V, Durand L, Moreau F, Dohi H, Vincent S, Escaich S, Ducruix A (2006) Structure of the Escherichia coli heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. *J Mol Biol* **363:** 383-394

7

Hess D, Keusch JJ, Oberstein SA, Hennekam RC, Hofsteenge J (2008) Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective O-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* **283:** 7354-7360

Hofsteenge J, Huwiler KG, Macek B, Hess D, Lawler J, Mosher DF, Peter-Katalinic J (2001) C-Mannosylation and O-Fucosylation of the Thrombospondin Type 1 Module. *J Biol Chem* **276:** 6485-6498

Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, Nakagawa A, Taniguchi N (2007) Crystal structure of mammalian α1,6-fucosyltransferase, FUT8. *Glycobiology* **17:** 455-466

Kozma K, Keusch JJ, Hegemann B, Luther KB, Klein D, Hess D, Haltiwanger RS, Hofsteenge J (2006) Identification and characterization of a β1,3-glucosyltransferase that synthesizes the Glc-β1,3-Fuc disaccharide on thrombospondin type 1 repeats. *J Biol Chem* **281:** 36742-36751

Krieg J, Hartmann S, Vicentini A, Gläsner W, Hess D, Hofsteenge J (1998) Recognition Signal for C-Mannosylation of Trp-7 in RNase 2 Consists of Sequence Trp-x-x-Trp. *Mol Biol Cell* **9:** 301-309

Lira-Navarrete E, Valero-González J, Villanueva R, Martínez-Júlvez M, Tejero T, Merino P, Panjikar S, Hurtado-Guerrero R (2011) Structural Insights into the Mechanism of Protein O-Fucosylation. *PLoS ONE* **6:** e25365

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* **41:** 207-234

Wang Y, Lee GF, Kelley RF, Spellman MW (1996) Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains. *Glycobiology* **6:** 837-842

Supplementary figure legends

Figure S1

Comparison of the nearest structural neighbours of human POFUT2. The orientation is kept the same for all molecules to facilitate the structural comparison. (**A**) Human POFUT2. (**B**) *C. elegans* POFUT1. (**C**) *B. japonicum* nodulation fucosyltransferase NodZ. (**D**) *E. coli* lipopolysaccharide heptosyltransferase WaaC. (**E**) Human α-1,6-fucosyltransferase FUT8.

8

Figure S2

Michaelis-Menten enzyme kinetics for POFUT2-dependent fucosylation of TSR4 from rat F-spondin. (**A**) Chromatogram showing the separation of the fucosylated TSR4 product from the TSR4 substrate by LC-MS. (**B**) Formation of fucosylated TSR4 reaction product at varying GDP-fucose concentrations and different TSR4 concentrations (3.3-200 μM). (**C**) As in (B) but with keeping the GDP-fucose concentration constant (2.5-160 μM) and varying the TSR4 concentration.

Figure S3

Superposition of the non-crystallographic-symmetry dimer of the POFUT2 apo structure (light green, cartoon model) with a crystallographic dimer of the POFUT2 GDP-fucose complex (in cyan with GDP-fucose as sticks in atom colours). Dimers were superimposed only on one chain to display the different packing. A rotational and translational shift of the symmetry related dimer (~ 10 Å) is clearly visible in the POFUT2 GDP-fucose complex. The calculated interface area between symmetry related chains amounts to 1670 $Å^2$ for the apo structure and is reduced to 1315 $Å^2$ for the POFUT2 GDP-fucose complex revealing the opening of the interchain cavity and enabling access to the sugar donor binding site.

Figure S4

Stereo view of the superposition of refined GDP-fucose POFUT2 complex structures with GDP-fucose once in the 1C4 (cyan) and once in the 4C1 conformation (gray) as sticks in atom colours. A better fit of the fucose moiety to the electron density with a real space correlation coefficient of 0.952 (calculated for chain C) was achieved with the 1C4 conformation versus 0.924 for the 4C1 conformation (F_c versus $2mF_o$-DF_c calculated with PHENIX). Selected GDP-fucose binding residues are displayed as fine gray lines in atom colours while the E54 side chain from the apo structure is superimposed (sticks in gray).

Figure S5

Profile hidden Markov model (HMM) of the Pfam entry of the thrombospondin type 1 domain (PF00090, http://pfam.sanger.ac.uk) shown as sequence logo image. Sequence stretches present in the mini-TSR are indicated as solid lines.

9

Figure S6

Cartoon models of GT-B fold inverting glycosyltransferases of known structure where the catalytic base (either aspartate, glutamate, or histidine) has been described and its position mapped. (**A**) Human POFUT2. (**B**) *E. coli* lipopolysaccharide heptosyltransferase WaaC. (**C**) T4 phage β-glucosyltransferase BGT. (**D**) *H. pylori* fucosyltransferase α1,3-FucT. (**E**) *E. coli* MurG. (**F**) plant *Vv*GT1. Sugar donors and the catalytic base are displayed as sticks in atom colours. In the structure of BGT in complex with UDP-glucose the catalytic aspartate (D100) was mutated to alanine.

10

Supplementary Table

Table S1: DALI searches

search template	PDB entries	Z score	rmsd	alignment/ residue	sequence identity	Protein	Reference
POFUT2 (42-429)	3ZY2	22.9	2.9	281/350	21%	POFUT1	(Lira-Navarrete et al, 2011)
	2HHC	16.1	3.8	251/291	17%	NODZ	(Brzezinski et al, 2007)
	2H1H	16	3.9	250/286	18%	WaaC	(Grizot et al, 2006)
	2DE0	15	3.7	235/460	14%	FUT8	(Ihara et al, 2007)
POFUT2 (42-242)	3ZY3	12.8	3.0	156/352	14%	POFUT1	(Lira-Navarrete et al, 2011)
	2HHC	7.4	3.5	122/291	14%	NODZ	(Brzezinski et al, 2007)
POFUT2 (243-429)	3ZY5	13.6	1.9	123/356	28%	POFUT1	(Lira-Navarrete et al, 2011)
	2H1H	12	2.8	124/286	20%	WaaC	(Grizot et al, 2006)
	2HHC	11.8	2.9	125/291	20%	NODZ	(Brzezinski et al, 2007)
	2DE0	8.3	2.8	135/460	19%	FUT8	(Ihara et al, 2007)

Figure S1

A
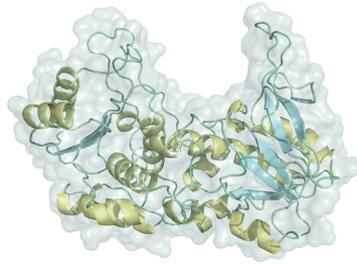
hPOFUT2

B
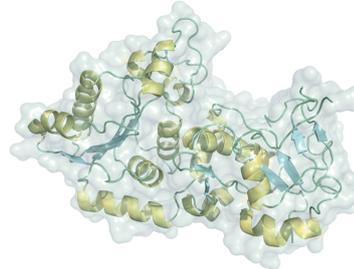
cePOFUT1: 3ZY2

C
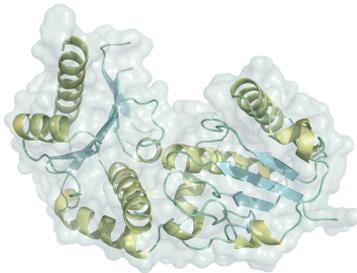
NodZ: 2HHC

D
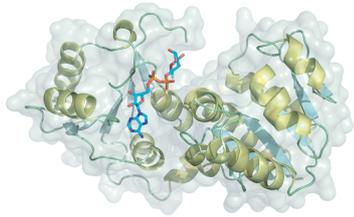
Waac: 2H1H

E
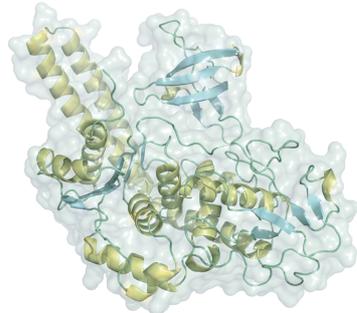
FUT8: 2DE0

Figure S2

A

B

C
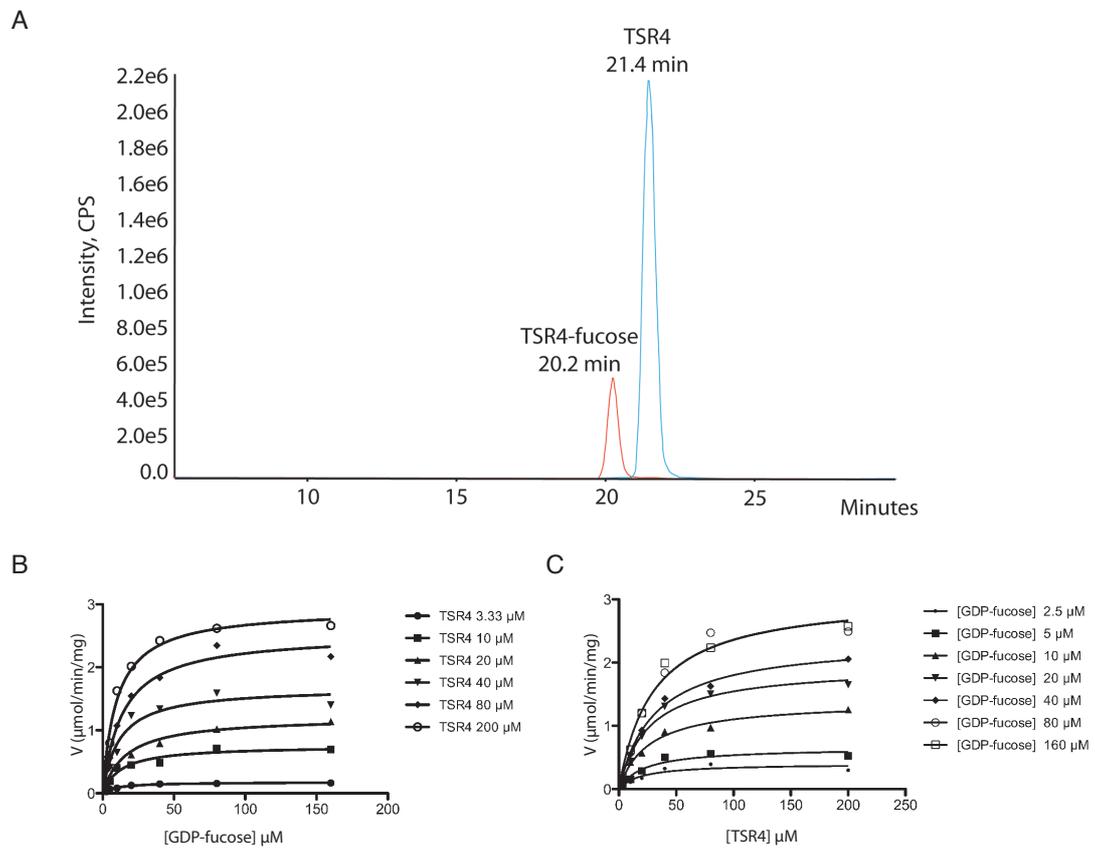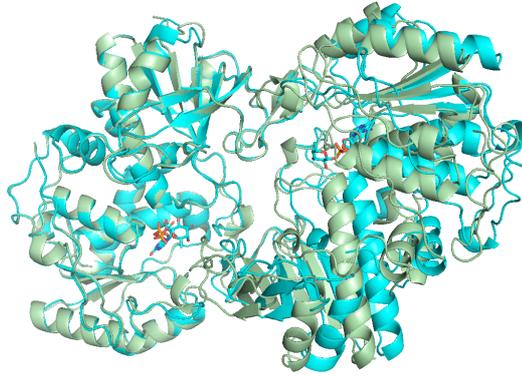
Figure S3

Figure S4

Figure S5

Figure S6

A

hPOFUT2

B

WaaC: 2H1H

C

BGT: 1NZD

D

α1,3-FucT: 2NZY

E

MurG: 1NLM

F

VvGT1: 2C1Z
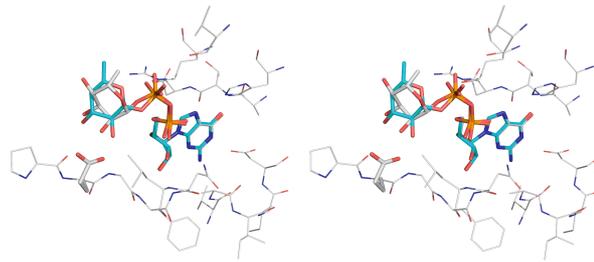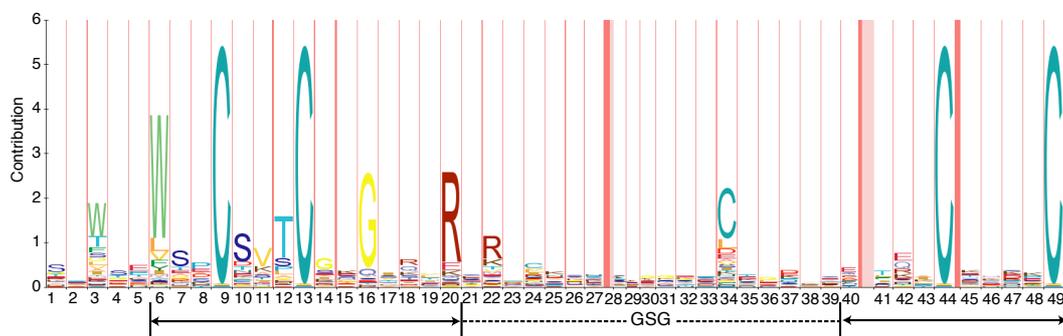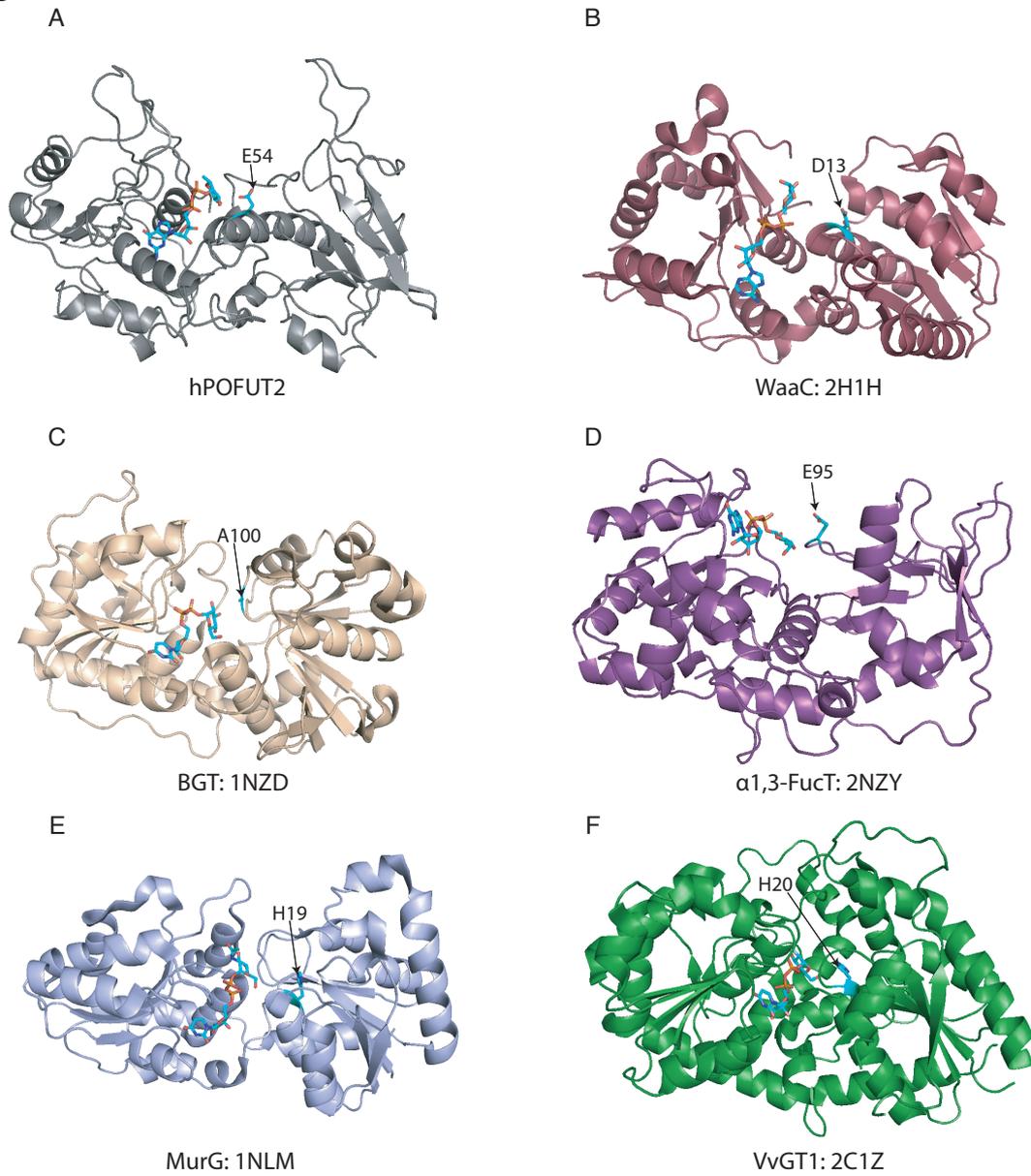
6.3 Tag-free purification of glycoproteins

Introduction

A major obstacle in protein crystallography is the difficulty of expressing and purifying proteins of high quality in large quantities. Since glycoproteins usually cannot be functionally expressed in bacterial expression systems, cell expression systems possessing a glycosylation machinery such as yeast, insect, or mammalian systems are usually required. Glycoproteins typically have more than one glycosylation site, which can be linked to asparagine, serine, or threonine. The structure of these glycans may be protein-specific and/or site-specific, which can also depend on the expression organism. For example, the *N*-glycan structure from HEK293T cells is usually a mixture of high-mannose, hybrid, and complex types with multi-antenna *N*-glycan and negatively charged sialic acid on the terminus (Chang et al., 2007, Chapter 6.1). To improve the diffracting quality of the protein crystal, it is usually necessary to simplify the heterogeneity of the *N*-glycan structure by treating the cell with chemicals such as kifunensine, which will interfere with the glycosylation pathway or to enzymatically cleave the *N*-glycans to a single GlcNAc by treatment with endoglycosidase H (Chang et al, 2007).

Purification of recombinant proteins with artificial peptides or protein tags is a powerful tool. Purification with artificial tags can usually provide hundred-fold increases in purity compared to the crude cell lysate extract. The mild and specific elution conditions for affinity purification also make it a practical approach for maintaining the activity of the target protein or purifying the protein in complex with other components. However, there are some drawbacks in the application of affinity tags. For example, the position of the tag may affect the expression level or accessibility to the resins. Small-scale test expression with different tags at different positions is necessary to select the best construct. Furthermore, the tags usually need to be removed to exclude the effect of the tag on the structure or function of the target protein. However, proteolytic cleavage of the affinity tag is not always practical and can be a time-consuming and laborious process.

To improve the purification process of glycoproteins, we have developed a method to purify these proteins without using artificial peptides or protein tags based on lectin affinity chromatography. Lectins are a group of carbohydrate-binding proteins that

bind to specific epitopes on carbohydrates. Since the discovery of the first lectin over a hundred years ago, many lectins have been identified in bacteria, plants, and animals. Lectin column chromatography has been extensively used in the fractionation of oligosaccharides, glycopeptides, and glycoproteins (Freeze, 2001; Fukuda & Kobata, 1993). In contrast to other purification techniques such as size-exclusion, ion-exchange, reverse-phase, and normal-phase chromatography, lectin affinity chromatography is based on its specificity for carbohydrates; therefore, the purification principle is more similar to that of affinity purification of recombinant proteins with an artificial tag.

One of the most well-characterized and widely used lectins is concanavalin A (Con A), because it is commercially available and has defined specificity. Con A displays high affinity for high-mannose type and hybrid-type *N*-glycans. It recognizes the tri-mannosyl core structure on the *N*-glycan (the boxed structure on the right); modification on mannose (Man) with GlcNAc will reduce the affinity for Con A, but the affinity will not increase if R is mannose (Baenziger & Fiete, 1979). Con A also weakly binds α-methylmannoside, which is very helpful feature because the bound protein can be competitively eluted from Con A by using high concentration of α-methylmannoside. Here, we present a purification method for *N*-glycosylated glycoproteins without using an artificial tag. The method has high specificity and high binding capacity; >50 mg of pure human POFUT1 was successfully purified by a single step of Con A purification.
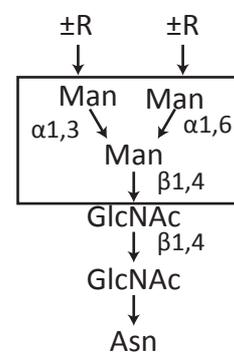
Materials and Methods

Cell culture

The method for creating a stable HEK293T cell line expressing Δ21-POFUT2 was described in Chapter 6.3. The stable line expressing PAD2 (the POFUT2 orthologue in *C. elegans*) was established with the same method. The HEK293T cell line expressing Δ23-POFUT1 was kindly provided by Dr. Jeremy Keusch. HEK293T cell lines expressing different proteins were all cultured under following conditions. Cells were first grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, penicillin, streptomycin, and Zeozin until the cells reached ~90% confluency. The DMEM medium was replaced by

Pro293a medium (Lonza) supplemented with 2 mM glutamine and 5 μM kifunensine for protein expression. The medium was harvested for purification after 72 h of cultivation.

Reducing volume of conditioned medium by Vivaflow 200

The Vivaflow 200 (10 kDa, PES, Sartorius) was connected to a peristaltic pump, according to manufacturer's instructions. The module was washed first with water and then with 400 ml of 20 mM Tris-HCl pH 7.4, 150 mM NaCl, and 1 mM $CaCl_2$. The sample was circulated through the system at a flow rate of 250 ml/min. The medium (2.5 l) was concentrated to 250 ml in 2 h.

Western blot

Proteins were separated on a 4%–12% Bis-Tris gel (NuPAGE, Invitrogen) and transferred to a nitrocellulose membrane. The primary antibody was rabbit anti-human POFUT2 (amino acids 44–58). The secondary antibody was anti-rabbit IgG HRP (GE Healthcare). The primary antibody was diluted 1:7500 in 5% low-fat milk powder in 20mM Tris-HCl, 150 mM NaCl and 0.05% Tween 20 (TBST), and the secondary antibody was diluted 1:10,000 in TBST. The blots were developed using the enhanced chemiluminescence detection kit (GE Healthcare).

Con A affinity chromatography

A 5-ml HiTrap Con A 4B column (GE Healthcare) was connected to an ÄKTA purifier and washed with Buffer A (20 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM $MnCl_2$, and 2 mM $CaCl_2$) before loading the sample. The concentration of $CaCl_2$ and $MnCl_2$ in the conditioned medium was adjusted to 2 mM and 1 mM, respectively. The medium was loaded and circulated within the column overnight at a flow rate of 1 ml/min. The medium was collected the next morning, and the column was washed with Buffer A until the A280 was stable. The protein was eluted by the gradient of 0%–50% Buffer B (1 M mannosyl pyranose in Buffer A) for 25 min; the flow was paused for 10 min, and then, restarted to allow 5 ml to flow. The eluted fractions were analyzed by SDS-PAGE, and the fractions with target protein were pooled for further processing. POFUT2 and PAD2 were purified by the same method, except that a 1-ml Con A column was used.

Endo Hf digestion of POFUT1

The POFUT1 fraction was buffer exchanged into 50 mM sodium citrate pH 5.5, 300 mM NaCl, 2 mM $CaCl_2$, and 1 mM $MnCl_2$ by using a HiTrap desalting column (GE

Healthcare). The concentration of the protein was adjusted to 0.15 mg/ml. POFUT1 (1 μg) was digested with 1 U of Endo Hf (New England BioLabs) for 3 h at 37 ºC. The digested sample was incubated with Amylose resin. Endo Hf was retained on the Amylose resin; the flow through was collected; buffer exchanged into Tris-HCl pH 8.5, 100 mM NaCl, and 1 mM MnCl$_2$ by using a HiTrap Desalting column; and concentrated to 7 mg/mL and 11.45 mg/mL.

Size-exclusion chromatography

Size-exclusion chromatography was used to purify the POFUT2 and PAD2 samples after Con A affinity chromatography. Samples were loaded onto a Superdex 200 16/60 column (GE Healthcare) in 20 mM Tris-HCl pH 7.4, 150 mM NaCl. The fractions were checked by SDS-PAGE, and the fractions with target protein were pooled, concentrated, and stored.

Results

Expression levels in Pro293a

To purify glycoproteins from conditioned medium of mammalian cells by Con A affinity chromatography, the first obstacle is that the serum contains high amounts of glycoproteins, which will compete with the recombinant protein for Con A binding and contaminate the eluted sample. All our stable HEK293T cell lines were established in DMEM medium supplemented with 10% FCS. To avoid the serum in the culture medium, we tested POFUT2 expression in Pro293a medium, which is a chemically defined, protein-free medium optimized for adherent HEK293 cell lines. We first tried to adapt the stable HEK293T cell line to Pro293a medium, but it was impractical because of the low growth rate of the cells, which results in low protein yields. Cells grown in serum-free medium cannot adhere to the culture dish, which reduces their growth rate.
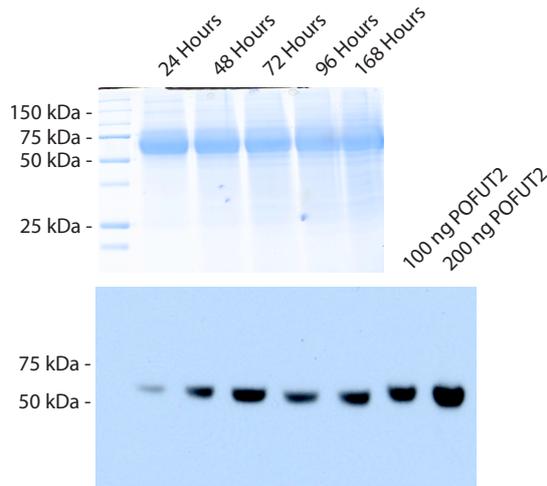
Figure 6.3.1 Expression level of POFUT2 in Pro293a medium at different time points

The expression level of POFUT2 from HEK293T cell in serum-free medium was quantified by SDS-PAGE and western blotting. No obvious difference was observed between the signals from the medium harvested at different time points, based on Coomassie Blue staining, because the major protein in the medium was BSA, which also runs at a similar molecular weight on the SDS-PAGE. On the western blot (1°Ab: anti-hPOFUT2:44-58), the intensity of the band increased from 24 h to 72 h. The expression level of POFUT2, compared to the standard purified POFUT2, was about 15 mg/l of medium after 72 h of cultivation in Pro293a medium.

To circumvent the growth-related problem, we cultured the cells in DMEM medium supplemented with 10% FCS until the cells reached 90% confluency. The medium was then switched to Pro293a medium supplemented with 2 mM glutamine and 5 μM kifunensine to produce high-mannose type POFUT2. The conditioned medium was harvested at 24 h, 48 h, 72 h, 96 h, and 168 h after exchanging the cells into the medium. The same volume of the conditioned medium from each time point was analyzed by Coomassie Blue staining and western blotting by using an antibody specific to the amino acids 44–58 of human POFUT2 (Fig. 6.3.1). From the results of Coomassie Blue staining, no obvious difference in protein expression was observed between the different time points. An intense band between 50 kDa and 75 kDa could be the residual bovine serum albumin (66 kDa) from the complete medium. Although POFUT2 could not be identified by Coomassie Blue staining of the SDS-PAGE, there was a clear signal on the immunoblot obtained using an antibody against the N-terminus of POFUT2 (Fig. 6.3.1). After 24 h in Pro293a medium, a faint but clear band was observed. The intensity of the signal increased after 48-h and 72-h cultivation in the Pro293a medium. We observed a decrease in signal intensity after 96 h of cultivation. By comparing the signal intensity of 100 ng and 200 ng POFUT2, we estimated that the expression level of POFUT2 after 72 h of cultivation in Pro293a medium was ~15 mg/l, which is higher than the expression level in the same stable

cell line in complete medium (~7 mg/mL). The expression level testing of POFUT2 in Pro293a medium suggests that the serum-free and protein-free properties of this medium may aid in downstream purification and that it did not compromise protein expression levels.
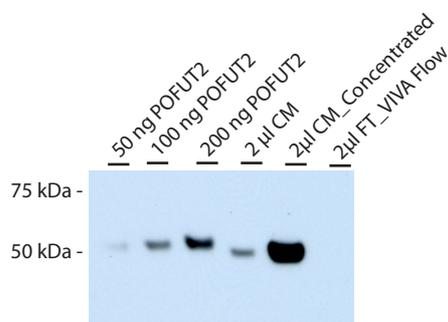

Figure 6.3.2 Reduction of sample volumes by Vivaflow 200

Western blotting (1°Ab: anti-hPOFUT2:44-58) was used to detect the expression level of POFUT2 in Pro293a medium and the concentration efficiency of Vivaflow 200. Conditioned medium (1 l) was efficiently concentrated to 50 ml in 30 min without significant protein loss.

The next problem we encountered in the purification of proteins from conditioned medium is the large sample volume. A batch of large-scale purification in our laboratory contains 4–5 l of conditioned medium. The loading of the medium onto the HiTrap Con A column is time-consuming because the maximum flow rate is 4 ml/min. Reducing the sample volume during the early steps of purification would increase the efficiency of the purification method (i.e., time and sample handling). We tested the concentrating efficiency of cross-flow filtration (Vivaflow 200 PES, 10 kDa). Conditioned medium (1 l) with POFUT2 after 72 h of cultivation in Pro293a was harvested and concentrated by Vivaflow 200. The medium was concentrated to 50 ml in 30 min at 4 °C. Importantly, we did not observe POFUT2 in the flow-through fraction, and the intensity of the signal in the concentrated fraction was stronger than that in the original conditioned medium, as determined on the basis of the data from western blotting by using the antibody against the N-terminus of POFUT2 (Fig 6.3.2). These data indicate that cross-filtration, as an early purification step, efficiently reduced the sample volume without significant sample loss, which can save time and effort in downstream purification steps.

Large-scale purification of POFUT1 by using the tag-free purification approach

Pro293a medium (2.5 l) after 72 h of cultivation was first concentrated 10-fold by using Vivaflow 200. A 5-ml Hitrap Con A column was connected to the ÄKTA purifier, and the sample was circulated overnight through the Con A column at a flow rate of

1.0 ml/min. The unbound fraction was collected in the next morning. After washing the column, we first eluted the sample with a linear gradient from 0% B to 50% B buffer over 25 min (Fig. 6.3.3A). However, we did not observe a clear elution peak by using this gradient, based on the UV A_{280} signal. It has been reported that some proteins bind tightly to the Con A column, and elution of tightly bound proteins can be achieved by introducing a short pause during elution (Soper & Aird, 2007). After we introduced several pauses during the elution, we observed peaks immediately after we restarted the flow. Increasing the concentration of mannosyl-pyranose, as recommended by Soper & Aird (2007), did not improve the elution efficiency.
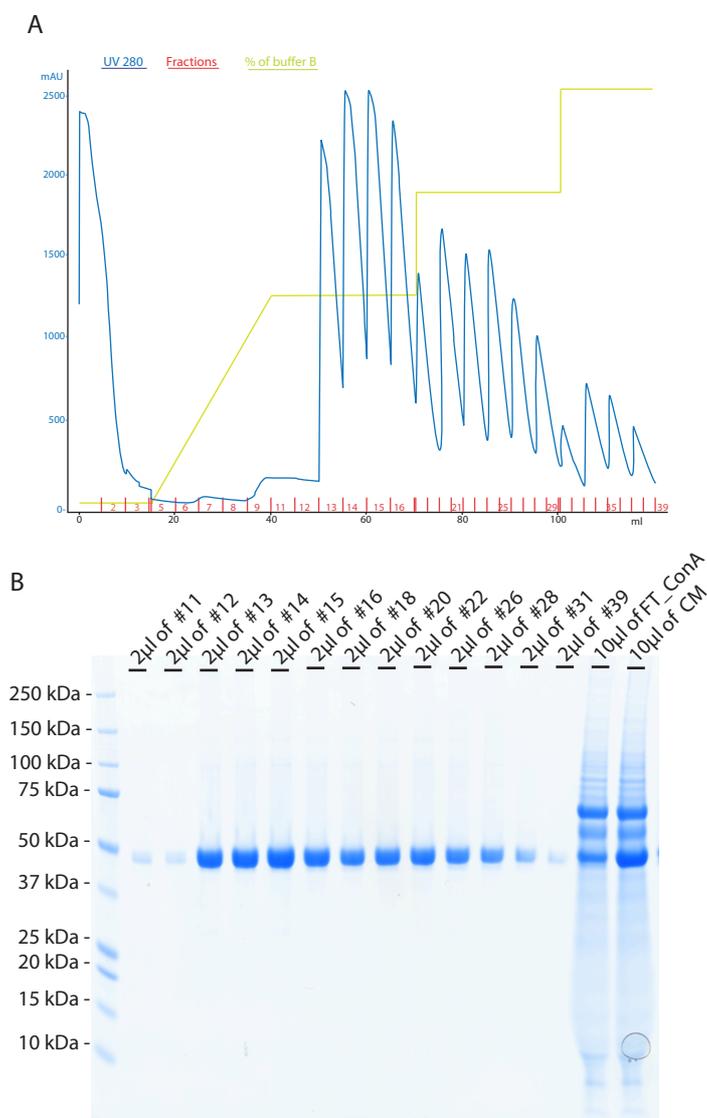
Figure 6.3.3 Purification of hPOFUT1 by using the Con A column

(A) The Con A column was connected to an ÄKTA purifier, and the elution fractions were monitored by reading the A280 (blue). The gradient of buffer B (1 M mannosyl pyranose) is shown by the green line, and the fraction number is given on the bottom in red. The fraction size was 5 ml. (B) SDS-PAGE analysis of the samples before and

after purification on the Con A column. Different fractions eluted from the Con A column were loaded into the gel.

The different column elution fractions from the Con A affinity chromatography column were analyzed by SDS-PAGE (Fig 6.3.3B). We observed a weak band at the predicted molecular weight for fractions #11 and #12. The intensity of the band increased when we introduced pauses in the elution procedure. Although only 0.04% of the sample from each fraction was loaded, we still observed a strong band on the gel. We also observed a band corresponding to POFUT1 in the flow-through (FT) fraction from the Con A column, although the intensity was weaker compared to the same band from the conditioned medium. These data suggest that the expression level of POFUT1 was high such that loading 2.5 l of conditioned medium saturated the column, and some POFUT1 flowed through the Con A column. We pooled fractions #11 to #39 and estimated the total amount of protein to be approximately 50 mg.

It is known from POFUT2 data that structurally flexible and heterogeneous *N*-glycans on the protein surface can interfere with the crystallization process and decrease the diffracting ability of the crystal. Based on the sequence of POFUT1, we know that it has two potential *N*-glycosylation sites: Asn62 and Asn160. Since one of our goals was to solve the crystal structure of POFUT1 and compare the catalytic mechanism between the two *O*-fucosyltransferases, we used Endo Hf to remove *N*-glycans on the surface of hPOFUT1.
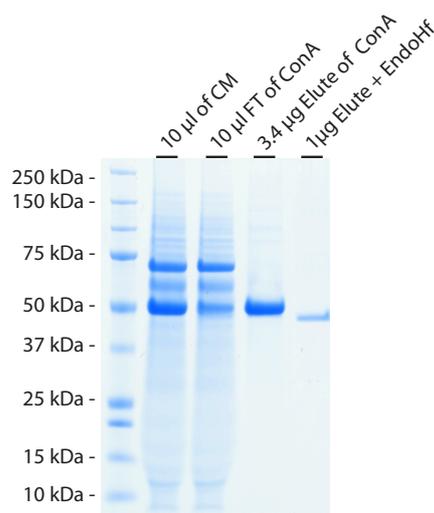

Figure 6.3.4 Purification of hPOFUT1 with a Con A affinity column

The purification of high-mannose type POFUT1 can be achieved by single-step Con A affinity purification. The *N*-glycans on POFUT1 can be removed by Endoglycosidase Hf treatment.

After Endo Hf treatment, we observed a band shift compared to the untreated sample by SDS-PAGE (Fig. 6.3.4). However, without *N*-glycans, POFUT1 tends to precipitate at an acidic pH (pH 5.5 for Endo Hf cleavage), which leads to significant sample loss. As such, POFUT1 was immediately buffer exchanged into 20 mM Tris-HCl pH 8.5, 100 mM NaCl, and 1 mM $MnCl_2$, and then concentrated to 7.9 mg/ml. Overall, we isolated 15 mg of purified, de-glycosylated POFUT1.

Purification of POFUT2 by Con A affinity chromatography

We have shown that tag-free purification of human POFUT1, which has two potential *N*-glycosylation sites, was successful. The *N*-glycan in medium supplemented with kifunensine is high-mannose type, which is a strong ligand for Con A. Different glycoproteins may have a different number of *N*-glycosylation sites; therefore, we wished to test whether these proteins could also be purified by the Con A-based affinity chromatography method without using an artificial tag.

Human POFUT2 has three potential *N*-glycosylation sites, which were shown to be fully occupied by electron density map data. Similar to the expression approach for human POFUT1, we first cultivated the cells in the DMEM medium supplemented with 10% FCS, and the medium was replaced with Pro293a medium supplemented with 5 mM kifunensine and 2 mM glutamine when cells reached 90% confluency. The medium from twenty 15 cm dishes (~370 ml) was harvested, filtered, and purified using a 1-ml HiTrap Con A column. We also used the pause/restart flow approach to improve the elution of tightly bound protein from the Con A column. Coomassie Blue staining of the elution fractions from the Con A column, (lane 1 in Fig 6.3.5B) showed that POFUT2 was contaminated with other proteins, which were removed by size-exclusion column chromatography (Superdex 200 16/60 column); the chromatogram is shown in Figure 6.3.5A. From the signal at A_{280}, we observed two main peaks. The first one was eluted around 50 ml, which ran at 200 kDa on the SDS-PAGE gel; POFUT2 was eluted in the second peak at ~78 ml from the Superdex 200 column. Therefore, compared to the purification procedure of POFUT1, only one additional step of size-exclusion chromatography was required to obtain a pure protein. Importantly, we could purify a protein that has three *N*-glycans from medium using an artificial tag.

A

B
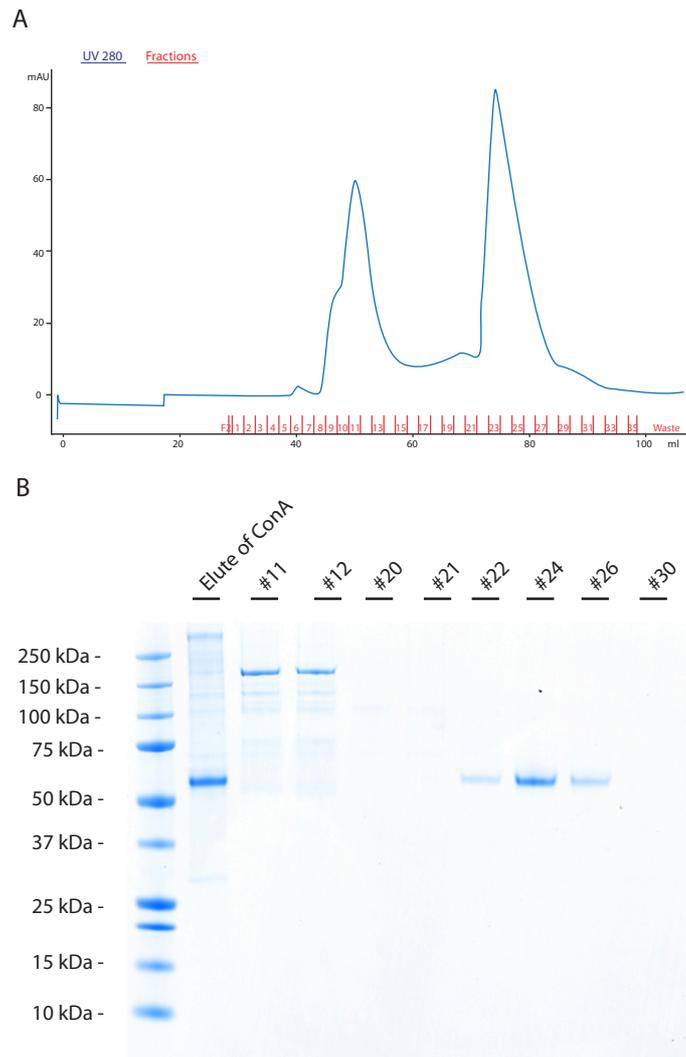
Figure 6.3.5 Purification of POFUT2 by Con A affinity chromatography and size-exclusion chromatography (SEC)

POFUT2 was expressed and purified from Pro293a medium in two steps. The elution fractions from the Con A column were pooled and purified by SEC. The chromatogram from the SEC in (A) and SDS-PAGE analysis of the fractions from the SEC in (B) show that POFUT2 can be purified to a single band in two sequential chromatographic steps.

Purification of cePOFUT2 (PAD2) by Con A affinity chromatography

POFUT1 and POFUT2 have two and three *N*-glycans, respectively, on the protein surface. The number of *N*-glycosylation sites on the protein may affect the binding affinity of Con A. Therefore, we investigated the application of this tag-free purification approach to purify a protein with only one *N*-glycosylation site from the conditioned medium.

A

B
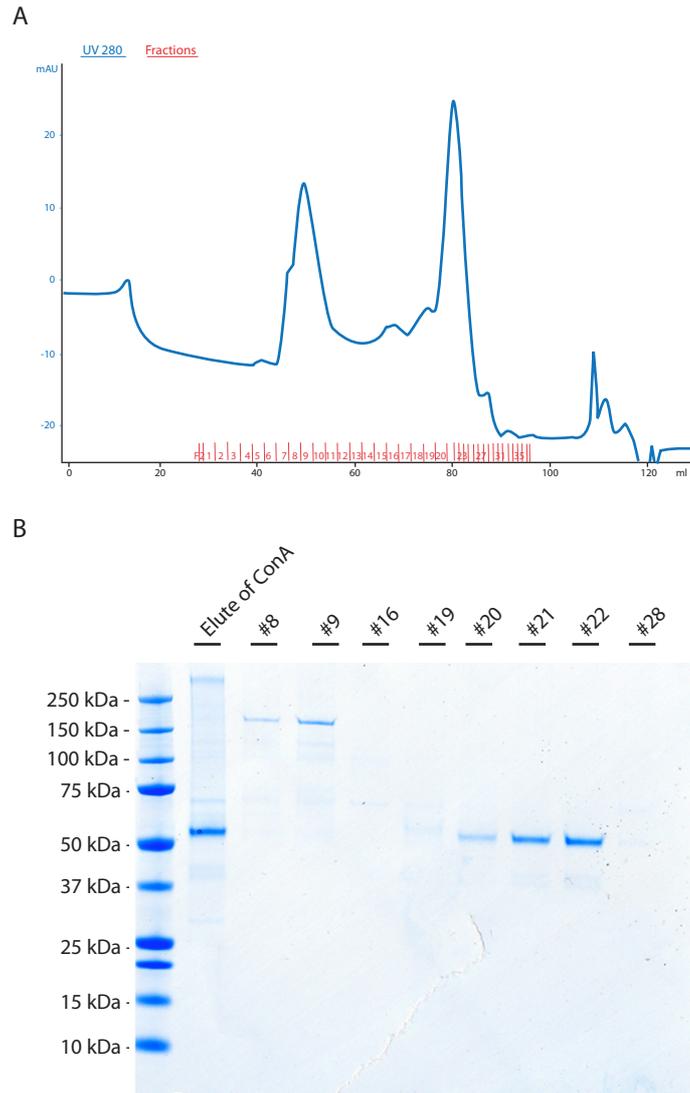
Figure 6.3.6 Purification of PAD2 by Con A affinity chromatography and size-exclusion chromatography

PAD2, which only has one *N*-glycosylation site, was purified sequentially by Con A affinity chromatography and size-exclusion chromatography. The chromatogram of SEC is displayed in (A), and the samples after Con A purification and SEC were analyzed by SDS-PAGE (B).

We expressed PAD2, which has only one potential *N*-glycosylation site on Asn205. The conditioned medium from the HEK293T stable line expressing PAD2 was harvested and circulated in the 1-ml HiTrap Con A column at flow rate of 1 ml/min. PAD2 was not eluted from the Con A column with the linear gradient; therefore, it was also eluted with the pause/restart approach with 0.5 M mannosyl pyranose. Similar to what we observed with POFUT2, PAD2 purified from the Con A affinity chromatography column was contaminated with other proteins (lane 2 of Fig. 6.3.6 B). An additional step of size-exclusion chromatography successfully removed most of the contaminants (Fig. 6.3.5), resulting in a single band in the fractions #20 to #22.

These data indicate the binding provided by a single high-mannose type *N*-glycan is sufficient for the tag-free purification of proteins with only one *N*-glycosylation site by Con A affinity chromatography.

Discussion

Production of a large quantity of active, pure protein is crucial to generate high-quality crystals to elucidate the protein structure by X-ray crystallography; this is particularly challenging for proteins modified by *N*-glycosylation. *N*-glycosylated proteins can usually not be properly expressed in prokaryotic expression systems as they lack the glycosylation machinery. *N*-glycan is generally required for correct folding as well as to maintain protein solubility. Glycans on glycoproteins expressed in mammalian cells are usually heterogenous. With kifunensine, a compound that can inhibit the activity of mannosidase II in the *N*-glycan processing pathway, the high-mannose type *N*-glycan could still have a different number of mannose groups (from Man_5–Man_9). Although protein expression in mammalian cells has improved in the recent years, and the expression levels in an optimized case have reached up to 1 g/l (Backliwal et al., 2008), production of active glycoproteins for structural studies is still a challenging task. In this chapter, we described the production and purification of glycoproteins from HEK293T cells with a newly established tag-free purification system. The glycoproteins were successfully purified from a serum-free conditioned medium by using Con A affinity chromatography without the involvement of an artificial tag (illustrated in Fig. 6.3.7). We purified proteins with one, two, or three *N*-glycosylation sites; the target proteins were efficiently purified to a single band with a single step of Con A affinity chromatography or with an additional step of size-exclusion chromatography. Although this technique is very promising for the purification of glycoproteins, feasible improvements to this purification approach and the application of other lectin affinity chromatography methods are discussed below.


| Protein expressed with high-mannose type *N*-glycans | Con A affinity chromatography → Pure protein with *N*-glycans | Endo H digestion on N-glycans → Protein after Endo H treatment |
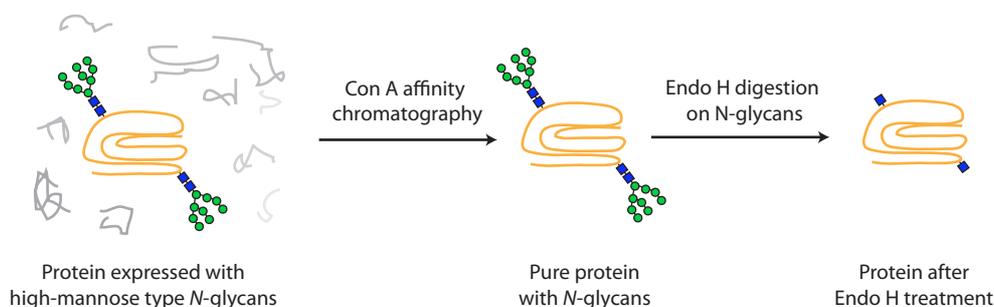
Figure 6.3.7 Con A purification scheme of human POFUT1

POFUT1 was purified from the conditioned medium by a single step of Con A affinity chromatography. Two *N*-glycans on the protein can be digested by Endoglycosidase H, which leaves a single *N*-acetylglucosamine on each *N*-glycosylation site.

Different cell lines for protein production

In this work, glycoproteins were produced in adherent HEK293T cells. However, culturing of adherent cells can be difficult to scale up. Growing the adherent cells involves trypsinization-based harvesting of the cells and reseeding the trypsinized cells in new culture dishes, which can become a tedious process when 5 l of cultured cells is required (i.e., 250 15 cm culture dishes are required). Moreover, the HEK293T cell line has not been adapted to the serum-free system. Cell culture for protein expression studies can be divided into two phases. The first phase is expansion in the serum-containing medium until the cells grow to 90% confluency. The second phase requires switching the medium to serum-free medium for production of proteins. A cell line that has the ability to grow in suspension and in serum-free medium would be very useful to establish a streamlined workflow.

Few mammalian cell lines have been engineered to overcome the problems associated with large-scale glycoprotein production. HEK.EBNA is a modified HEK 293 cell line featuring a stable integrated copy of the EBNA-1 gene of the Epstein-Barr virus. The function of EBNA-1 is to drive episomal replication of the expression plasmid carrying the OriP origin of replication derived from the EBV (Van Craenenbroeck et al., 2000). A protocol for the adaption of HEK.EBNA cells to grow in suspension and in serum-free medium has been described (Meissner et al., 2001). The FreeStyle 293 Expression System is a commercial system (Invitrogen) using a cell line derived from HEK cells, which is adapted to grow at high density in serum-free suspension culture. The FreeStyle 293 Expression System also includes specific medium and transfection reagents for high transfection efficiency and high protein yield. HEK293S is another cell line derived from HEK293 cells. It lacks *N*-acetylglucosaminyltransferase I (GnTI) activity, and consequently, the *N*-glycan structure is maintained as the high-mannose type (Reeves et al., 2002). Although no study has described their adaptation to serum-free medium, protein production from HEK293S cells, which has only Man_5 *N*-glycans, still deems it an attractive cell line for producing glycoproteins for structural studies.

Saturation of the Con A column can reduce non-specific binding

The POFUT1 elution fractions from the Con A affinity column were >95% homogeneous (Fig. 6.3.3). On the other hand, the POFUT2 and PAD2 fractions eluted from the Con A column were not as pure as POFUT1 (Fig. 6.3.5 & 6.3.6); an additional step of gel filtration was necessary to eliminate the contaminants.

Like POFUT1, POFUT2 and PAD2 were expressed in the same serum-free medium by the same parental cell line. Although the number of *N*-glycans differs between these proteins, which could affect binding to the Con A column, this should not contribute to the presence of impurities in the POFUT2 and PAD2 fractions because POFUT2 has the highest number of *N*-glycans. We hypothesize that the impurities are because of the amount of target protein loaded onto the Con A column. POFUT1 is highly expressed, and although a 5-ml Con A column was used (for 2.5 l of medium), some residual POFUT1 was still detected in the unbound fraction, indicating that we had exceeded the column binding capacity. By comparison, the expression of POFUT2 and PAD2 was much lower, and therefore, a 1-ml Con A column was used for the purification step (for 400 ml of medium). However, the presence of contaminants in the eluted protein fractions suggests the presence of unoccupied Con A binding sites on the column, resulting in non-specific binding. Indeed, it is known that a 5-ml Con A column can roughly bind 50 mg of POFUT1, which is equivalent to 2 mM of *N*-glycan. Based on this estimation, the same Con A column could bind to ~33 mg of POFUT2 (and a 1-ml Con A column could bind ~6.6 mg of POFUT2). Since the expression level of POFUT2 was ~7 mg/l, we could increase the ratio of target protein to Con A to decrease non-specific binding.

An alternative solution to reduce non-specific binding is to add a Con A ligand to the binding buffer. Con A has high affinity for the tri-mannosyl *N*-glycan core structure, but it can also bind to mannosyl pyranose. A low concentration (10 mM) of mannosyl pyranose added to the binding buffer may help reduce non-specific binding.

Purification of glycoproteins from conditioned medium with different lectins

Many lectins are commercially available in an immobilized form for lectin affinity purification (Freeze, 2001). Con A, the lectin used in this chapter, is one of the most commonly used lectins because of its availability and stability. Wheat germ agglutinin (WGA) is another popular lectin because of its broad specificity. It has high affinity to the terminal *N*-acetylglucosamine, but also binds to terminal sialic acid (NeuAcα2-3(6)-R) and linear polylactosamine (Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Gal-

4GlcNAcβ1-R) chains with lower affinity. If the purified protein is not for use in structural studies, the protein can be expressed without the supplementation of kifunensine, leaving the *N*-glycans on the protein as a complex with sialic acid and polylactosamines, in order to allow purification with WGA.

Here, we have described an approach for the large-scale expression and purification of recombinant glycoproteins without using an artificial protein or peptide tags. The inherent *N*-glycan on the glycoprotein provides a natural tag for purification using the Con A lectin. The capacity of the Con A affinity purification is quite high; a single Con A column can purify >50 mg of POFUT1 in a single purification step. These features make it an attractive technique for glycoprotein purification.

7. Summary and Discussion

In this study, we determined the crystal structure of native POFUT2 and enzymatically inactive E54A-POFUT2 in complex with GDP-fucose. We showed that POFUT2 belongs to the GT-B superfamily of glycosyltransferases with two Rossmann-like domains separated by an inter-domain cleft. GDP-fucose binds to the C-terminal domain and we proposed that TSR would bind to POFUT2 in the inter-domain cleft. Based on the structures and the enzymatic activity assay of wild-type and POFUT2 mutants, we proposed a catalytic mechanism of POFUT2. R294 forms a salt bridge with the β-phosphate group of GDP-fucose. The carboxylate side chain of E54 works as a catalytic base that deprotonates the hydroxyl group on the serine or threonine on the TSR, which attacks as nucleophile the anomeric carbon of GDP-fucose. We also observed that the substrate specificity of POFUT2 for TSR seems to be achieved by the structural complementarity of a part of the TSR with the POFUT2 binding site. The disruption of the unique compact three-strand structure of TSR by mutating key residues decreased the fucosylation efficiency. The *in vitro* and *in vivo* fucosylation of artificial mini-TSR support our model that POFUT2 recognizes properly folded TSR through the region around the AB loop.

Divalent metal ion in GT-B superfamily

Unlike glycosyltransferases with the GT-A fold, which have a short DXD motif involved in metal binding for catalysis, most of the glycosyltransferases in the GT-B superfamily use a metal-ion independent mechanism for catalysis. In the GT-A superfamily, the positively charged Mn^{2+} ion is clearly positioned to neutralize the negative charges of the phosphate group of the sugar donor (Ünligil & Rini, 2000). However, catalysis of a few glycosyltransferases in the GT-B superfamily is stimulated in presence of divalent metal ions. The best characterized glycosyltransferase that needs metal ions for full activity is DNA β-glucosyltransferase (BGT) from T4 phage (Moréra et al, 2001). It has been shown that without metal ions in the assay BGT still maintained 10% of activity. The presence of metal ions in the activity assay such as Mg^{2+} and Mn^{2+} significantly increase the activity. Comparing the structures of BGT in complex with GDP-Mn^{2+} and GDP-glucose revealed that the metal ion binds almost the same position as glucose in the GDP-glucose complex with the metal ion and the glucose C1 atom being only separated by less than 1 Å. In other words, the sugar moiety fills the

space occupied by the metal (Larivière et al, 2003). This observation suggested that a role of the metal ion in GT-B fold glycosyltransferases is incompatible with activating catalysis like in the GT-A superfamily. The authors proposed a secondary role in product release after glucose transfer, but the detailed mechanism remains to be elucidated.

Another glycosyltransferase of the GT-B superfamily that needs metal ions for full activity is POFUT1. In the activity assay of POFUT1 expressed in CHO cells, it was shown that POFUT1 remains active in the presence of EDTA and that divalent metal ions such as Mn^{2+} increase the activity 17-fold (Wang & Spellman, 1998). This observation was consistent with the role of metal ions in BGT where metal ions are needed for full activity.

Interestingly, in the recent structures of POFUT1 from *C. elegans* it was also observed that the metal ion occupies the location of the fucose moiety in the GDP-fucose complex (Lira-Navarrete et al, 2011). The structures of POFUT1 with GDP-fucose and GDP imposed nicely, which indicated no significant conformational change induced by binding of the different ligands (Fig. 7.1). The close-up image of the GDP-fucose binding site shows that metal ion cannot co-exist with the fucose moiety from GDP-fucose as the ligand binds at the same place.
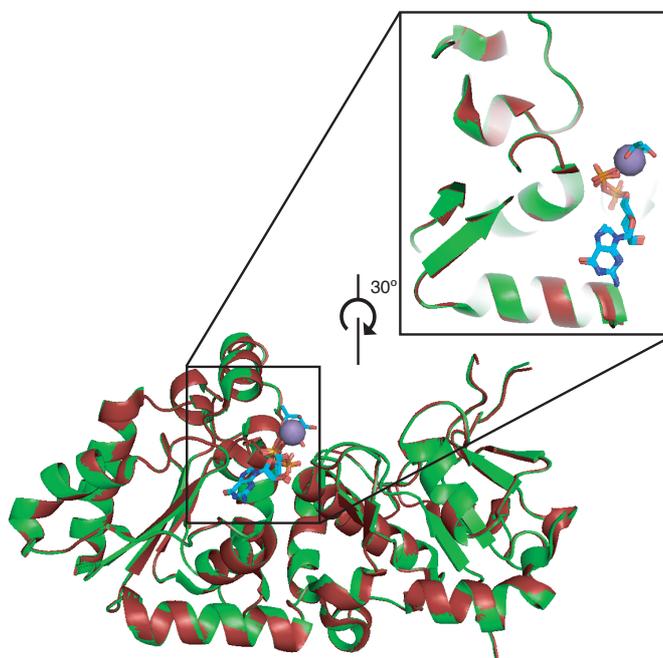
Figure 7.1 Structure of CePOFUT1 in complex with GDP-Mn^{2+} and GDP-fucose

Two POFUT1 structures are imposed and displayed as cartoon models. The structure in red is in complex with GDP-fucose (PDB 3ZY6) and structure in green is in complex with GDP and Mn^{2+} (PDB 3ZY2). GDP and GDP-fucose are shown as sticks and Mn^{2+}

is shown as a sphere. The GDP-fucose binding region is highlighted in the close-up image.

The role of the metal ion in the catalytic process of POFUT2 is also unclear. We showed that the basic activity of POFUT2 does not require metal ions but that Mg^{2+} or Mn^{2+} can increase the activity more than 10-fold (Fig. 6.2.3C). The lack of a GDP-Mn^{2+} POFUT2 structure makes it difficult to predict if a metal ion would locates in the position where fucose binds. Interestingly, crystals of E54A-POFUT2 in complex with GDP-fucose were grown in the presence of EDTA. Replacing the EDTA with 5 mM of $MnCl_2$ in the crystallization condition did not yield any crystals (unpublished data), implying that $MnCl_2$ may also interfere with the binding of GDP-fucose and affect the crystal packing. We proposed that the metal ion in POFUT2 plays a similar role as the metal ions in BGT and POFUT1 where these ions are not required for sugar transfer. Binding of the metal at the binding site of the sugar where it interacts with the diphosphate moiety of the hydrolyzed sugar donor may help the release of fucosylated product. Further experiments to crystallize POFUT2 with GDP and Mn^{2+} may help to elucidate the binding site of the ion in POFUT2.

POFUT2 *O*-fucosylation on TSRs-containing proteins

Most of the TSR does not exist as a single module in the protein; instead several TSRs may be linked together and form the multi-TSRs proteins (Fig.5.4.1). Although we used a single module TSR4 from rat F-spondin for testing the activity of POFUT2 and for building our POFUT2-GDP-fucose-TSR4 ternary model (Fig. 6.2.5), we did not exclude the possibility that the additional TSR module may affect fucosylation efficiency. The structure of TSR23 from TSP-1 was used to replace the TSR4 in the model (Fig. 6.1.12), which shows that POFUT2 could interact with two consecutive TSRs in the inter-domain cleft. Preliminary results of *O*-fucosylation of TSR23* show that POFUT2 can indeed fucosylate TSR2 on TSR23*, which suggests that the second module (TSR3*) would not interfere with the binding of TSR23* to POFUT2 (Chapter 6.1). Furthermore, the higher fucosylation efficiency of TSR23* compared to the single module TSR4 from F-spondin suggests that two consecutive TSR modules could act as a better substrate for *O*-fucosylation than a single TSR module (Chapter 6.1).

Interestingly, the recent structure of ADAMTS13-DTC, which contains only a single TSR module provides new insight into how POFUT2 recognizes and fucosylates TSR-containing proteins (Akiyama et al, 2009). ADATMTS13 is a multi-domain protein with a metalloproteinase domain, a disintegrin-like (D) domain, a TSR (T), a

cysteine-rich (C) region, a spacer (S), seven TSRs (T2–T8), and two CUB domains. The ADAMTS13-DTC encompasses the disintegrin-like (D) domain to the spacer (S), and only a single TSR locates between the disintegrin-like domain and the cysteine-rich domain (Fig. 7.2A). ADAMTS13-DTC was expressed in CHO cells and a Glcβ1,3Fuc disaccharide was found on the consensus sequence in the AB loop of TSR1 (Akiyama et al, 2009). When we superimposed the TSR1 of ADAMTS13-DTC onto the TSR in our model (Figure 6.2.5), we found that entire cysteine-rich region and parts of the spacer domain overlap with the N-terminal domain of POFUT2.
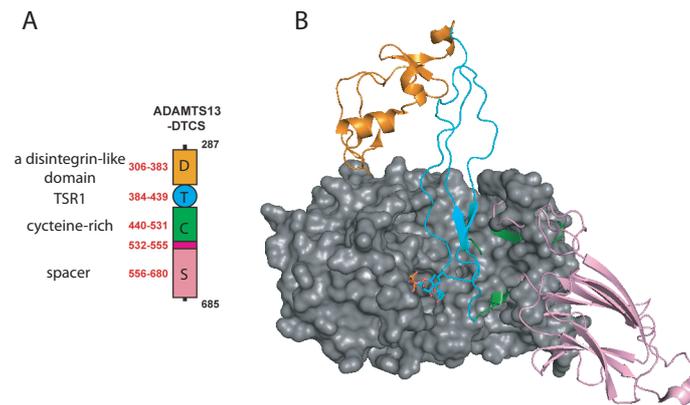
(A) Modified from (Akiyama et al, 2009)

Figure 7.2 The model of ADAMTS13-DTC in complex with POFUT2

The domain structure of ADAMTS13-DTC was illustrated in (A). Each domain is color-coded and the color of each domain is consistent with the structure in (B). POFUT2 is displayed as grey surface.

As TSR in the ADAMT13-DTC structure was found to be fucosylated, the POFUT2 in CHO cells must have a strategy to accept multidomain proteins as substrate. Currently, we hypothesize that POFUT2 binds and fucosylates TSR1 when the interdomain arrangement is not yet completed. Binding of POFUT2 to TSR not only transfers the fucose to TSR but may also delay the folding of the multidomain protein. After the fucose is added to the TSR, POFUT2 is released from ADAMT13 and interdomain contacts which the cysteine-rich region can be established.

In summary, we present the crystal structure of POFUT2 and of the POFUT2-GDP-fucose complex. We also propose an enzymatic mechanism of fucose transfer to TSR. Importantly, we elucidate that the substrate specificity relies on the unique three-stranded structure of TSR that precisely presents the fucosylation site to the catalytic region in the POFUT2 structure. Future studies of the POFUT2-TSRs complex and a comparison of binding affinities between single and two modules will uncover the underlying mechanism and function of this unique type of glycosylation.

8. List of Figures and Tables

9. References

Acar M, Jafar-Nejad H, Takeuchi H, Rajan A, Ibrani D, Rana NA, Pan H, Haltiwanger RS, Bellen HJ (2008) Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **132:** 247-258

Adams JC, Tucker RP (2000) The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development. *Dev Dyn* **218:** 280-299

Ahn J, Ludecke H-J, Lindow S, Horton WA, Lee B, Wagner MJ, Horsthemke B, Wells DE (1995) Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nature genetics* **11:** 137-143

Akiyama M, Takeda S, Kokame K, Takagi J, Miyata T (2009) Crystal structures of the noncatalytic domains of ADAMTS13 reveal multiple discontinuous exosites for von Willebrand factor. *Proc Nat Acad Sci USA* **106:** 19274-19279

Asch AS, Silbiger S, Heimer E, Nachman RL (1992) Thrombospondin sequence motif (CSVTCG) is responsible for CD36 binding. *Biochemical and biophysical research communications* **182:** 1208-1217

Baenziger JU, Fiete D (1979) Structural determinants of concanavalin A specificity for oligosaccharides. *J Biol Chem* **254:** 2400-2407

Baenziger NL, Brodie GN, Majerus PW (1971) A Thrombin-Sensitive Protein of Human Platelet Membranes. *Proc Nat Acad Sci USA* **68:** 240-243

Baker LH, Rowinsky EK, Mendelson D, Humerickhouse RA, Knight RA, Qian J, Carr RA, Gordon GB, Demetri GD (2008) Randomized, phase II study of the thrombospondin-1-mimetic angiogenesis inhibitor ABT-510 in patients with advanced soft tissue sarcoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26:** 5583-5588

Baldus SE, Zirbes TK, Monig SP, Engel S, Monaca E, Rafiqpoor K, Hanisch FG, Hanski C, Thiele J, Pichlmaier H, Dienes HP (1998) Histopathological subtypes and prognosis of gastric cancer are correlated with the expression of mucin-associated sialylated antigens: Sialosyl-Lewis(a), Sialosyl-Lewis(x) and sialosyl-Tn. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **19:** 445-453

Bolos V, Grego-Bessa J, de la Pompa JL (2007) Notch signaling in development and cancer. *Endocr Rev* **28:** 339-363

Brockhausen I, Schachter H, Stanley P (2009) O-GalNAc Glycans. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor NY: The Consortium of Glycobiology Editors, La Jolla, California

9. References

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

Campbell ID, Bork P (1993) Epidermal growth factor-like modules. *Curr Opin Struct Biol* **3:** 385-392

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37:** D233-238

Carlson CB, Lawler J, Mosher DF (2008) Structures of thrombospondins. *Cell Mol Life Sci* **65:** 672-686

Carraway KL, Fregien N, Carraway KL, 3rd, Carraway CA (1992) Tumor sialomucin complexes as tumor antigens and modulators of cellular interactions and proliferation. *J Cell Sci* **103 (Pt 2):** 299-307

Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI, Dwek RA, Jones EY, Owens RJ, Davis SJ (2007) Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15:** 267-273

Chen J, Moloney DJ, Stanley P (2001) Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1. *Proc Nat Acad Sci USA* **98:** 13716-13721

Cohen B, Bashirullah A, Dagnino L, Campbell C, Fisher WW, Leow CC, Whiting E, Ryan D, Zinyk D, Boulianne G, Hui CC, Gallie B, Phillips RA, Lipshitz HD, Egan SE (1997) Fringe boundaries coincide with Notch-dependent patterning centres in mammals and alter Notch-dependent development in Drosophila. *Nat Genet* **16:** 283-288

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328:** 307-317

Crawford SE, Stellmach V, Murphy-Ullrich JE, Ribeiro SM, Lawler J, Hynes RO, Boivin GP, Bouck N (1998) Thrombospondin-1 is a major activator of TGF-beta1 in vivo. *Cell* **93:** 1159-1170

Dawson DW, Pearce SF, Zhong R, Silverstein RL, Frazier WA, Bouck NP (1997) CD36 mediates the In vitro inhibitory effects of thrombospondin-1 on endothelial cells. *The Journal of cell biology* **138:** 707-717

de Beer T, Vliegenthart JFG, Loeffler A, Hofsteenge J (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:** 11785-11789

de Lonlay P, Seta N (2009) The clinical spectrum of phosphomannose isomerase deficiency, with an evaluation of mannose treatment for CDG-Ib. *Biochim Biophys Acta* **1792:** 841-843

Doucey MA, Hess D, Cacan R, Hofsteenge J (1998) Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor. *Molecular biology of the cell* **9:** 291-300

Du J, Takeuchi H, Leonhard-Melief C, Shroyer KR, Dlugosz M, Haltiwanger RS, Holdener BC (2010) O-fucosylation of thrombospondin type 1 repeats restricts epithelial to mesenchymal transition (EMT) and maintains epiblast pluripotency during mouse gastrulation. *Dev Biol* **346:** 25-38

Ebos JM, Lee CR, Cruz-Munoz W, Bjarnason GA, Christensen JG, Kerbel RS (2009) Accelerated metastasis after short-term treatment with a potent inhibitor of tumor angiogenesis. *Cancer Cell* **15:** 232-239

Ellies LG, Tsuboi S, Petryniak B, Lowe JB, Fukuda M, Marth JD (1998) Core 2 Oligosaccharide Biosynthesis Distinguishes between Selectin Ligands Essential for Leukocyte Homing and Inflammation. *Immunity* **9:** 881-890

Fernandez-Valdivia R, Takeuchi H, Samarghandi A, Lopez M, Leonardi J, Haltiwanger RS, Jafar-Nejad H (2011) Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. *Development* **138:** 1925-1934

Firlej V, Mathieu JRR, Gilbert C, Lemonnier L, Nakhlé J, Gallou-Kabani C, Guarmit B, Morin A, Prevarskaya N, Delongchamps NB, Cabon F (2011) Thrombospondin-1 Triggers Cell Migration and Development of Advanced Prostate Tumors. *Cancer Research* **71:** 7649-7658

Freeze HH (2001) Lectin Affinity Chromatography. In *Current Protocols in Protein Science*. John Wiley & Sons, Inc.

Freeze HH (2006) Genetic defects in the human glycome. *Nature reviews* **7:** 537-551

Freeze HH (2007) Congenital Disorders of Glycosylation: CDG-I, CDG-II, and beyond. *Curr Mol Med* **7:** 389-396

Frevert U, Sinnis P, Esko JD, Nussenzweig V (1996) Cell surface glycosaminoglycans are not obligatory for Plasmodium berghei sporozoite invasion in vitro. *Mol Biochem Parasitol* **76:** 257-266

Fritz TA, Raman J, Tabak LA (2006) Dynamic Association between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase-2. *J Biol Chem* **281:** 8613-8619

Fukuda M, Kobata A (1993) Glycobiology A Practical Approach.

Fuster MM, Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature reviews Cancer* **5:** 526-542

Garshasbi M, Hadavi V, Habibi H, Kahrizi K, Kariminejad R, Behjati F, Tzschach A, Najmabadi H, Ropers HH, Kuss AW (2008) A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. *Am J Hum Genet* **82:** 1158-1164

Gastinel LNI, Cambillau C, Bourne Y (1999) Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J* **18:** 3546-3557

Gitay-Goren H, Soker S, Vlodavsky I, Neufeld G (1992) The binding of vascular endothelial growth factor to its receptors is dependent on cell surface-associated heparin-like molecules. *J Biol Chem* **267:** 6093-6098

Gonzalez de Peredo A, Klein D, Macek B, Hess D, Peter-Katalinic J, Hofsteenge J (2002) C-mannosylation and o-fucosylation of thrombospondin type 1 repeats. *Mol Cell Proteomics* **1:** 11-18

Good DJ, Polverini PJ, Rastinejad F, Le Beau MM, Lemons RS, Frazier WA, Bouck NP (1990) A tumor suppressor-dependent inhibitor of angiogenesis is immunologically and functionally indistinguishable from a fragment of thrombospondin. *Proc Nat Acad Sci USA* **87:** 6624-6628

Greenaway J, Lawler J, Moorehead R, Bornstein P, Lamarre J, Petrik J (2007) Thrombospondin-1 inhibits VEGF levels in the ovary directly by binding and internalization via the low density lipoprotein receptor-related protein-1 (LRP-1). *J Cell Physiol* **210:** 807-818

Haines N, Irvine KD (2003) Glycosylation regulates Notch signalling. *Nat Rev Mol Cell Biol* **4:** 786-797

Hakomori S (1990) Bifunctional role of glycosphingolipids. Modulators for transmembrane signaling and mediators for cellular interactions. *J Biol Chem* **265:** 18713-18716

Hallgren P, Lundblad A, Svensson S (1975) A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine. *J Biol Chem* **250:** 5312-5314

Harris RJ, Spellman MW (1993) O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3:** 219-224

Hart GW, Akimoto Y (2009) The O-GlcNAc Modification. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor NY: The Consortium of Glycobiology Editors, La Jolla, California

9. References

Hart GW, Housley MP, Slawson C (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446:** 1017-1022

Hase S, Kawabata S, Nishimura H, Takeya H, Sueyoshi T, Miyata T, Iwanaga S, Takao T, Shimonishi Y, Ikenaka T (1988) A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factors VII and IX. *J Biochem (Tokyo)* **104:** 867-868

Hase S, Nishimura H, Kawabata S, Iwanaga S, Ikenaka T (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:** 1858-1861

Haviv F, Bradley MF, Kalvin DM, Schneider AJ, Davidson DJ, Majest SM, McKay LM, Haskell CJ, Bell RL, Nguyen B, Marsh KC, Surber BW, Uchic JT, Ferrero J, Wang YC, Leal J, Record RD, Hodde J, Badylak SF, Lesniewski RR, Henkin J (2005) Thrombospondin-1 mimetic peptide inhibitors of angiogenesis and tumor growth: design, synthesis, and optimization of pharmacokinetics and biological activities. *J Med Chem* **48:** 2838-2846

Heinonen TY, Pasternack L, Lindfors K, Breton C, Gastinel LN, Maki M, Kainulainen H (2003) A novel human glycosyltransferase: primary structure and characterization of the gene and transcripts. *Biochemical and biophysical research communications* **309:** 166-174

Henkin J, Volpert OV (2011) Therapies using anti-angiogenic peptide mimetics of thrombospondin-1. *Expert Opin Ther Targets* **15:** 1369-1386

Hess D, Keusch JJ, Oberstein SA, Hennekam RC, Hofsteenge J (2008) Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective O-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* **283:** 7354-7360

Hirel PH, Schmitter MJ, Dessen P, Fayat G, Blanquet S (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Nat Acad Sci USA* **86:** 8247-8251

Hiruma-Shimizu K, Hosoguchi K, Liu Y, Fujitani N, Ohta T, Hinou H, Matsushita T, Shimizu H, Feizi T, Nishimura S (2010) Chemical synthesis, folding, and structural insights into O-fucosylated epidermal growth factor-like repeat 12 of mouse Notch-1 receptor. *Journal of the American Chemical Society* **132:** 14857-14865

Hofsteenge J, Huwiler KG, Macek B, Hess D, Lawler J, Mosher DF, Peter-Katalinic J (2001) C-Mannosylation and O-Fucosylation of the Thrombospondin Type 1 Module. *J Biol Chem* **276:** 6485-6498

Ihara Y, Manabe S, Ikezaki M, Inai Y, Matsui I-SL, Ohta Y, Muroi E, Ito Y (2010) C-Mannosylated peptides derived from the thrombospondin type 1 repeat interact with Hsc70 to modulate its signaling in RAW264.7 cells. *Glycobiology* **20:** 1298-1310

Irvine KD, Wieschaus E (1994) fringe, a Boundary-specific signaling molecule, mediates interactions between dorsal and ventral cells during Drosophila wing development. *Cell* **79:** 595-606

Jaeken J (2010) Congenital disorders of glycosylation. *Ann N Y Acad Sci* **1214:** 190-198

Jaeken J, Carchon H, Stibler H (1993) The carbohydrate-deficient glycoprotein syndromes: pre-Golgi and Golgi disorders? *Glycobiology* **3:** 423-428

Jaeken J, Hennet T, Matthijs G, Freeze HH (2009) CDG nomenclature: Time for a change! *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1792:** 825-826

Jaeken J, Landerschueren-Lodeweyckx M, Casaer P, Snoeck L, Corbeel L, Eggermont E, Eeckels R (1980) Familial psychomotor retardation with markedly fluctuating serum prolactin, FSH and GH levels, partial TBG-deficiency, increased serum arylsulphatase A and increased CSF protein: a new syndrome?: 90. *Pediatr Res* **14:** 179

Jafar-Nejad H, Leonardi J, Fernandez-Valdivia R (2010) Role of glycans and glycosyltransferases in the regulation of Notch signaling. *Glycobiology* **20:** 931-949

Jarriault S, Brou C, Logeat F, Schroeter EH, Kopan R, Israel A (1995) Signalling downstream of activated mammalian Notch. *Nature* **377:** 355-358

Jimenez B, Volpert OV, Crawford SE, Febbraio M, Silverstein RL, Bouck N (2000) Signals leading to apoptosis-dependent inhibition of neovascularization by thrombospondin-1. *Nat Med* **6:** 41-48

Ju T, Cummings RD (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Nat Acad Sci USA* **99:** 16613-16618

Julien S, Krzewinski-Recchi MA, Harduin-Lepers A, Gouyer V, Huet G, Le Bourhis X, Delannoy P (2001) Expression of sialyl-Tn antigen in breast cancer cells transfected with the human CMP-Neu5Ac: GalNAc alpha2,6-sialyltransferase (ST6GalNAc I) cDNA. *Glycoconj J* **18:** 883-893

Kazerounian S, Yee KO, Lawler J (2008) Thrombospondins in cancer. *Cellular and molecular life sciences : CMLS* **65:** 700-712

Kentzer EJ, Buko A, Menon G, Sarin VK (1990) Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase. *Biochem Biophys Res Commun* **171:** 401-406

Kjaergaard S, Skovby F, Schwartz M (1999) Carbohydrate-deficient glycoprotein syndrome type 1A: expression and characterisation of wild type and mutant PMM2 in E. coli. *European journal of human genetics : EJHG* **7:** 884-888

Komatsu M, Jepson S, Arango ME, Carothers Carraway CA, Carraway KL (2001) Muc4/sialomucin complex, an intramembrane modulator of ErbB2/HER2/Neu, potentiates primary tumor growth and suppresses apoptosis in a xenotransplanted tumor. *Oncogene* **20:** 461-470

Kopan R, Ilagan MX (2009) The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* **137:** 216-233

Kozma K, Keusch JJ, Hegemann B, Luther KB, Klein D, Hess D, Haltiwanger RS, Hofsteenge J (2006) Identification and characterization of a β1,3-glucosyltransferase that synthesizes the Glc-β1,3-Fuc disaccharide on thrombospondin type 1 repeats. *J Biol Chem* **281:** 36742-36751

Kreppel LK, Blomberg MA, Hart GW (1997) Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J Biol Chem* **272:** 9308-9315

Lacey JM, Bergen HR, Magera MJ, Naylor S, O'Brien JF (2001) Rapid determination of transferrin isoforms by immunoaffinity liquid chromatography and electrospray mass spectrometry. *Clin Chem* **47:** 513-518

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu Rev Biochem* **77:** 521-555

Larivière L, Gueguen-Chaignon V, Moréra S (2003) Crystal Structures of the T4 Phage β-Glucosyltransferase and the D100A Mutant in Complex with UDP-glucose: Glucose Binding and Identification of the Catalytic Base for a Direct Displacement Mechanism. *J Mol Biol* **330:** 1077-1086

Lariviere L, Moréra S (2002) A base-flipping mechanism for the T4 phage beta-glucosyltransferase and identification of a transition-state analog. *J Mol Biol* **324:** 483-490

Läubli H, Borsig L (2010) Selectins promote tumor metastasis. *Semin Cancer Biol* **20:** 169-177

Lawler J, Detmar M (2004) Tumor progression: the effects of thrombospondin-1 and -2. *Int J Biochem Cell Biol* **36:** 1038-1045

Lawler J, Hynes RO (1986) The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J Cell Biol* **103:** 1635-1648

Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469:** 564-567

9. References

Lesnik Oberstein SA, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, Breuning MH, Hennekam RC (2006) Peters Plus syndrome is caused by mutations in β3GALTL, a putative glycosyltransferase. *Am J Hum Genet* **79:** 562-566

Lim K, Chang HI (2009) O-GlcNAc modification of Sp1 inhibits the functional interaction between Sp1 and Oct1. *FEBS Lett* **583:** 512-520

Lin S, Kemmner W, Grigull S, Schlag PM (2002) Cell surface alpha 2,6 sialylation affects adhesion of breast carcinoma cells. *Exp Cell Res* **276:** 101-110

Lira-Navarrete E, Valero-González J, Villanueva R, Martínez-Júlvez M, Tejero T, Merino P, Panjikar S, Hurtado-Guerrero R (2011) Structural Insights into the Mechanism of Protein O-Fucosylation. *PLoS ONE* **6:** e25365

Liu J, Mushegian A (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci* **12:** 1418-1431

Lizak C, Gerber S, Numao S, Aebi M, Locher KP (2011) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474:** 350-355

Lopez-Dee Z, Pidcock K, Gutierrez LS (2011) Thrombospondin-1: multiple paths to inflammation. *Mediators Inflamm* **2011:** 296069

Lovelace LL, Cooper CL, Sodetz JM, Lebioda L (2011) Structure of Human C8 Protein Provides Mechanistic Insight into Membrane Pore Formation by Complement. *J Biol Chem* **286:** 17585-17592

Lowe JB, Marth JD (2003) A genetic approach to Mammalian glycan function. *Annu Rev Biochem* **72:** 643-691

Luo Y, Haltiwanger RS (2005) O-fucosylation of notch occurs in the endoplasmic reticulum. *J Biol Chem* **280:** 11289-11294

Luo Y, Koles K, Vorndam W, Haltiwanger RS, Panin VM (2006a) Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *J Biol Chem* **281:** 9393-9399

Luo Y, Nita-Lazar A, Haltiwanger RS (2006b) Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J Biol Chem* **281:** 9385-9392

Luque A, Carpizo DR, Iruela-Arispe ML (2003) ADAMTS1/METH1 inhibits endothelial cell proliferation by direct binding and sequestration of VEGF165. *J Biol Chem* **278:** 23656-23665

Luther KB, Haltiwanger RS (2009) Role of unusual O-glycans in intercellular signaling. *Int J Biochem Cell Biol* **41:** 1011-1024

Marcum JA, Rosenberg RD (1985) Heparinlike molecules with anticoagulant activity are synthesized by cultured endothelial cells. *Biochem Biophys Res Commun* **126:** 365-372

Markovic SN, Suman VJ, Rao RA, Ingle JN, Kaur JS, Erickson LA, Pitot HC, Croghan GA, McWilliams RR, Merchan J, Kottschade LA, Nevala WK, Uhl CB, Allred J, Creagan ET (2007) A phase II study of ABT-510 (thrombospondin-1 analog) for the treatment of metastatic melanoma. *Am J Clin Oncol* **30:** 303-309

Matsuura A, Ito M, Sakaidani Y, Kondo T, Murakami K, Furukawa K, Nadano D, Matsuda T, Okajima T (2008) O-linked N-acetylglucosamine is present on the extracellular domain of notch receptors. *J Biol Chem* **283:** 35486-35495

Matuschewski K, Nunes AC, Nussenzweig V, Menard R (2002) Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J* **21:** 1597-1606

McClain DA, Lubas WA, Cooksey RC, Hazel M, Parker GJ, Love DC, Hanover JA (2002) Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proc Nat Acad Sci USA* **99:** 10695-10699

Molinari F, Foulquier F, Tarpey PS, Morelle W, Boissel S, Teague J, Edkins S, Futreal PA, Stratton MR, Turner G, Matthijs G, Gecz J, Munnich A, Colleaux L (2008) Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. *Am J Hum Genet* **82:** 1150-1157

Moloney DJ, Lin AI, Haltiwanger RS (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:** 19046-19050

Moloney DJ, Panin VM, Johnston SH, Chen J, Shao L, Wilson R, Wang Y, Stanley P, Irvine KD, Haltiwanger RS, Vogt TF (2000a) Fringe is a glycosyltransferase that modifies Notch. *Nature* **406:** 369-375

Moloney DJ, Shair LH, Lu FM, Xia J, Locke R, Matta KL, Haltiwanger RS (2000b) Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem* **275:** 9604-9611

Moréra S, LariviËre L, Kurzeck Jr, Aschke-Sonnenborn U, Freemont PS, Janin J, Rüger W (2001) High resolution crystal structures of T4 phage β-glucosyltransferase: induced fit and effect of substrate and metal binding. *J Mol Biol* **311:** 569-577

Murray BW, Takayama S, Schultz J, Wong C-H (1996) Mechanism and Specificity of Human α-1,3-Fucosyltransferase V: Glycosidic Cleavage Occurs Prior to Nucleophilic Attack. *Biochemistry* **35:** 11183-11195

Nagy P, Vereb G, Sebestyen Z, Horvath G, Lockett SJ, Damjanovich S, Park JW, Jovin TM, Szollosi J (2002) Lipid rafts and the local density of ErbB proteins influence

the biological role of homo- and heteroassociations of ErbB2. *J Cell Sci* **115:** 4251-4262

Nakamura H, Fujii Y, Inoki I, Sugimoto K, Tanzawa K, Matsuki H, Miura R, Yamaguchi Y, Okada Y (2000) Brevican is degraded by matrix metalloproteinases and aggrecanase-1 (ADAMTS4) at different sites. *J Biol Chem* **275:** 38885-38890

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

Okajima T, Xu A, Irvine KD (2003) Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe. *J Biol Chem* **278:** 42340-42345

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

Pääkkönen K, Tossavainen H, Permi P, Rakkolainen H, Rauvala H, Raulo E, Kilpeläinen I, Güntert P (2006) Solution structures of the first and fourth TSR domains of F-spondin. *Proteins: Structure, Function, and Bioinformatics* **64:** 665-672

Paez-Ribes M, Allen E, Hudock J, Takeda T, Okuyama H, Vinals F, Inoue M, Bergers G, Hanahan D, Casanovas O (2009) Antiangiogenic therapy elicits malignant progression of tumors to increased local invasion and distant metastasis. *Cancer Cell* **15:** 220-231

Panin VM, Papayannopoulos V, Wilson R, Irvine KD (1997) Fringe modulates Notch-ligand interactions. *Nature* **387:** 908-912

Perez-Vilar J, Randell SH, Boucher RC (2004) C-Mannosylation of MUC5AC and MUC5B Cys subdomains. *Glycobiology* **14:** 325-337

Pirard M, Matthijs G, Heykants L, Schollen E, Grunewald S, Jaeken J, van Schaftingen E (1999) Effect of mutations found in carbohydrate-deficient glycoprotein syndrome type IA on the activity of phosphomannomutase 2. *FEBS letters* **452:** 319-322

Qasba PK, Ramakrishnan B, Boeggeman E (2005) Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem Sci* **30:** 53-62

Qasba PK, Ramakrishnan B, Boeggeman E (2008) Structure and function of beta - 1,4-galactosyltransferase. *Curr Drug Targets* **9:** 292-309

Ramaekers VT, Stibler H, Kint J, Jaeken J (1991) A new variant of the carbohydrate deficient glycoproteins syndrome. *J Inherit Metab Dis* **14:** 385-388

Ramakrishnan B, Balaji PV, Qasba PK (2002) Crystal structure of β1,4-galactosyltransferase complex with UDP-Gal reveals an oligosaccharide acceptor binding site. *J Mol Biol* **318:** 491-502

Ramakrishnan B, Qasba PK (2002) Structure-based design of β1,4-galactosyltransferase I (β4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. *J Biol Chem* **277:** 20833-20839

Ramasamy V, Ramakrishnan B, Boeggeman E, Qasba PK (2003) The role of tryptophan 314 in the conformational changes of beta1,4-galactosyltransferase-I. *J Mol Biol* **331:** 1065-1076

Ramasamy V, Ramakrishnan B, Boeggeman E, Ratner DM, Seeberger PH, Qasba PK (2005) Oligosaccharide preferences of beta1,4-galactosyltransferase-I: crystal structures of Met340His mutant of human beta1,4-galactosyltransferase-I with a pentasaccharide and trisaccharides of the N-glycan moiety. *J Mol Biol* **353:** 53-67

Rana NA, Haltiwanger RS (2011) Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Curr Opin Struct Biol* **21:** 583-589

Reiher FK, Volpert OV, Jimenez B, Crawford SE, Dinney CP, Henkin J, Haviv F, Bouck NP, Campbell SC (2002) Inhibition of tumor growth by systemic treatment with thrombospondin-1 peptide mimetics. *Int J Cancer* **98:** 682-689

Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, Majerus EM (2007) O-Fucosylation Is Required for ADAMTS13 Secretion. *J Biol Chem* **282:** 17014-17023

Rini J, Esko J, Varki A (2009) Chapter 5 Glycosyltransferases and Glycan-processing Enzymes. *Glycobiology*

Ruddock LW, Molinari M (2006) N-glycan processing in ER quality control. *J Cell Sci* **119:** 4373-4380

Sakaidani Y, Ichiyanagi N, Saito C, Nomura T, Ito M, Nishio Y, Nadano D, Matsuda T, Furukawa K, Okajima T (2012) O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1. *Biochemical and biophysical research communications*

Sakaidani Y, Nomura T, Matsuura A, Ito M, Suzuki E, Murakami K, Nadano D, Matsuda T, Furukawa K, Okajima T (2011) O-linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell-matrix interactions. *Nat Commun* **2:** 583

Sato T, Sato M, Kiyohara K, Sogabe M, Shikanai T, Kikuchi N, Togayachi A, Ishida H, Ito H, Kameyama A, Gotoh M, Narimatsu H (2006) Molecular cloning and characterization of a novel human β1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain. *Glycobiology* **16:** 1194-1206

Schulz BL, Stirnimann CU, Grimshaw JPA, Brozzo MS, Fritsch F, Mohorko E, Capitani G, Glockshuber R, Grutter MG, Aebi M (2009) Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *PNAS* **106:** 11061-11066

Schwarz F, Aebi M (2011) Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **21:** 576-582

Sethi MK, Buettner FF, Krylov VB, Takeuchi H, Nifantiev NE, Haltiwanger RS, Gerardy-Schahn R, Bakker H (2010) Identification of glycosyltransferase 8 family members as xylosyltransferases acting on O-glucosylated notch epidermal growth factor repeats. *J Biol Chem* **285:** 1582-1586

Sethi MK, Buettner FFR, Ashikov A, Krylov VB, Takeuchi H, Nifantiev NE, Haltiwanger RS, Gerardy-Schahn R, Bakker H (2012) Molecular Cloning of a Xylosyltransferase That Transfers the Second Xylose to O-Glucosylated Epidermal Growth Factor Repeats of Notch. *J Biol Chem* **287:** 2739-2748

Sharma CB, Lehle L, Tanner W (1981) N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. *European journal of biochemistry / FEBS* **116:** 101-108

Sherblom AP, Buck RL, Carraway KL (1980) Purification of the major sialoglycoproteins of 13762 MAT-B1 and MAT-C1 rat ascites mammary adenocarcinoma cells by density gradient centrifugation in cesium chloride and guanidine hydrochloride. *J Biol Chem* **255:** 783-790

Shi S, Stanley P (2003) Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc Nat Acad Sci USA* **100:** 5234-5239

Soper AS, Aird SD (2007) Elution of tightly bound solutes from concanavalin A Sepharose. Factors affecting the desorption of cottonmouth venom glycoproteins. *J Chromatogr A* **1154:** 308-318

Sparrow DB, Chapman G, Wouters MA, Whittock NV, Ellard S, Fatkin D, Turnpenny PD, Kusumi K, Sillence D, Dunwoodie SL (2006) Mutation of the LUNATIC FRINGE gene in humans causes spondylocostal dysostosis with a severe vertebral phenotype. *Am J Hum Genet* **78:** 28-37

St. Hill CA, Baharo-Hassan D, Farooqui M (2011) C2-O-sLeX Glycoproteins Are E-Selectin Ligands that Regulate Invasion of Human Colon and Hepatic Carcinoma Cells. *PLoS ONE* **6:** e16281

Stahl M, Uemura K, Ge C, Shi S, Tashima Y, Stanley P (2008) Roles of Pofut1 and O-fucose in mammalian Notch signaling. *J Biol Chem* **283:** 13638-13651

Stanley P (2007) Regulation of Notch signaling by glycosylation. *Curr Opin Struct Biol* **17:** 530-535

9. References

Stanley P, Schachter H, Taniguchi N (2009) N-Glycans. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor NY: The Consortium of Glycobiology Editors, La Jolla, California

Tajiri M, Yoshida S, Wada Y (2005) Differential analysis of site-specific glycans on plasma and cellular fibronectins: application of a hydrophilic affinity method for glycopeptide enrichment. *Glycobiology* **15:** 1332-1340

Takeuchi H, Fernandez-Valdivia RC, Caswell DS, Nita-Lazar A, Rana NA, Garner TP, Weldeghiorghis TK, Macnaughtan MA, Jafar-Nejad H, Haltiwanger RS (2011) Rumi functions as both a protein O-glucosyltransferase and a protein O-xylosyltransferase. *Proc Nat Acad Sci USA* **108:** 16600-16605

Tan K, Duquette M, Liu JH, Dong Y, Zhang R, Joachimiak A, Lawler J, Wang JH (2002) Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. *J Cell Biol* **159:** 373-382

Tossavainen H, Pihlajamaa T, Huttunen TK, Raulo E, Rauvala H, Permi P, Kilpeläinen I (2006) The layered fold of the TSR domain of P. falciparum TRAP contains a heparin binding site. *Protein Sci* **15:** 1760-1768

Tucker RP (2004) The thrombospondin type 1 repeat superfamily. *The international journal of biochemistry & cell biology* **36:** 969-974

Typas A, Banzhaf M, Gross CA, Vollmer W (2012) From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol* **10:** 123-136

Ünligil UM, Rini JM (2000) Glycosyltransferase structure and mechanism. *Curr Opin Struct Biol* **10:** 510-517

Varki A, Sharon N (2009) Historical Background and Overview. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds). Cold Spring Harbor NY: The Consortium of Glycobiology Editors, La Jolla, California

Vazquez F, Hastings G, Ortega MA, Lane TF, Oikemus S, Lombardo M, Iruela-Arispe ML (1999) METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity. *J Biol Chem* **274:** 23349-23357

Vogt G, Chapgier A, Yang K, Chuzhanova N, Feinberg J, Fieschi C, Boisson-Dupuis S, Alcais A, Filipe-Santos O, Bustamante J, de Beaucoudrey L, Al-Mohsen I, Al-Hajjar S, Al-Ghonaium A, Adimi P, Mirsaeidi M, Khalilzadeh S, Rosenzweig S, de la Calle Martin O, Bauer TR, Puck JM, Ochs HD, Furthner D, Engelhorn C, Belohradsky B, Mansouri D, Holland SM, Schreiber RD, Abel L, Cooper DN, Soudais C, Casanova JL (2005) Gains of glycosylation comprise an unexpectedly large group of pathogenic mutations. *Nat Genet* **37:** 692-700

Vosseller K, Sakabe K, Wells L, Hart GW (2002a) Diverse regulation of protein function by O-GlcNAc: a nuclear and cytoplasmic carbohydrate post-translational modification. *Curr Opin Chem Biol* **6:** 851-857

Vosseller K, Wells L, Lane MD, Hart GW (2002b) Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *Proc Nat Acad Sci USA* **99:** 5313-5318

Vrielink A, Ruger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J* **13:** 3413-3422

Waller JP (1963) The Nh2-Terminal Residues of the Proteins from Cell-Free Extracts of E. Coli. *J Mol Biol* **7:** 483-496

Wang LW, Dlugosz M, Somerville RPT, Raed M, Haltiwanger RS, Apte SS (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:** 17024-17031

Wang LW, Leonhard-Melief C, Haltiwanger RS, Apte SS (2009) Post-translational Modification of Thrombospondin Type-1 Repeats in ADAMTS-like 1/Punctin-1 by C-Mannosylation of Tryptophan. *J Biol Chem* **284:** 30004-30015

Wang Y, Spellman MW (1998) Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *J Biol Chem* **273:** 8112-8118

Wopereis S, Grünewald S, Huijben KMLC, Morava É, Mollicone R, van Engelen BGM, Lefeber DJ, Wevers RA (2007) Transferrin and Apolipoprotein C-III Isofocusing Are Complementary in the Diagnosis of N- and O-Glycan Biosynthesis Defects. *Clin Chem* **53:** 180-187

Xu A, Haines N, Dlugosz M, Rana NA, Takeuchi H, Haltiwanger RS, Irvine KD (2007) In vitro reconstitution of the modulation of Drosophila Notch-ligand binding by Fringe. *J Biol Chem* **282:** 35153-35162

Young GD, Murphy-Ullrich JE (2004a) Molecular interactions that confer latency to transforming growth factor-beta. *J Biol Chem* **279:** 38032-38039

Young GD, Murphy-Ullrich JE (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:** 47633-47642

10. Curriculum Vitae

Personal Information

Name	Chun-I Chen　陳駿逸
Date of birth	1st November 1978
Nationality	Taiwan
Email	Chuni.chen@fmi.ch

Education

Jul 2007-present Expected graduation date Apr 2012	PhD. In Biochemistry, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. Structure of Human POFUT2: Insights into Thrombospondin Type 1 Repeat Fold and *O*-Fucosylation
Jul 2004	M.Sc. Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan Glycomics mapping and sequencing of colon carcinomas by Mass Spectrometry: fucosylation on extended type 1 chain
Jul 2002	BSc, Department of Botany, National Taiwan University, Taipei, Taiwan

Work Experience

Mar 2006-Jun 2007	Research Assistant. Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan
Oct 2004-Jan 2006	Second Lieutenant. Army of Republic of China. Obligatory military service
Jan 2001-Jul 2004	Freelance photographer

Laboratory Skills

Cell biology   Mammalian cell culture; transfection; generating stable cell line

Mass spectrometry   Analyzing protein, peptide and glycan

Glycan analysis N-glycan and O-glycan release; purification of glycans; permethylated glycan analysis by mass spectrometry

Biochemistry   Protein purification with different chromatographies; enzyme kinetics analysis, SDS-PAGE, Western blot

Protein crystallography   Skills from screening crystallization condition to data collection and analysis

Other Skills

Language   Mandarin, Taiwanese, English

PC-skills   Apple, Microsoft Windows, Linux, MS Office, Photoshop, Illustrator, structural analysis tools

Conference Attended

Nov 2011   Annual Conference of the Society for Glycobiology, Seattle, USA
                                Selected for oral presentation
                                Structural and functional characterization of human protein
                                O-fucosyltransferase 2

Course attended

Jan 2011   CCP4 Study Weekend: Model Building, Refinement and Validation

Publication

Chun-I Chen, Jeremy J. Keusch, Dominique Klein, Daniel Hess, Jan Hofsteenge, Heinz Gut
Structure of Human POFUT2: Insights into Thrombospondin Type 1 Repeat Fold and O-Fucosylation
Paper submitted (2012)

Chi-Hung Lin, Yao-Yun Fan, Yen-Ying Chen, Shih-Hsin Wang, **Chun-I Chen**, Lung-Chih Yu and Kay-Hooi Khoo
Enhanced expression of β3-galactosyltransferase 5 activity is sufficient to induce in vivo synthesis of extended type 1 chains on lactosylceramides of selected human colonic carcinoma cell lines
Glycobiology 19 (2009), 418-427

Motoki Terada, Kay-Hooi Khoo, Risa Inoue, **Chun-I Chen**, Kanako Yamada, Hiromi Sakaguchi, Naoko Kadowaki, Bruce Yong Ma, Shogo Oka, Toshisuke Kawasaki, Nobuko Kawasaki
Characterization of oligosaccharide ligands Expressed on SW1116 cells recognized by mannan-binding protein: A HIGHLY FUCOSYLATED POLYLACTOSAMINE TYPE N-GLYCAN
J Biol Chem 280 (2005), 10897-10913