Glycosylation of Thrombospondin type 1 repeats

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Carsten Krantz
aus Grünberg, Deutschland

Basel, März 2008

Friedrich-Miescher-Institut
Maulbeerstr. 66
4058 Basel

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von
Professor Nancy Hynes
Professor Kurt Ballmer-Hofer
PD Dr. Jan Hofsteenge

Basel, den 22.04.2008

Prof. Dr. Hans-Peter Hauri
Dekan

Erklärung:

Ich erkläre, dass ich die Dissertation “Glycosylation of Thrombospondin type 1 repeats” nur mit der darin angegeben Hilfe verfasst und bei keiner anderen Universität und keiner anderen Fakultät der Universität Basel eingereicht habe.

If I have been able to see farther than others, it was because I stood on the shoulders of giants.
Isaac Newton

Table of contents

Figures

Tables

Abbreviations

(C^2-Man-)Trp: C^2-α-mannosyltrytophan
ATCC: American Type Culture Collection
BSA: Bovine serum albumin
CDG: Congenital disorder of glycosylation
CHO: Chinese Hamster Ovary
CID: Collision induced dissociation
C-Man-Trp: C-mannosylated tryptophan
DTT: Dithiothreitol
ECM: Extracellular matrix
EGF: Epidermal growth factor
ER: Endoplasmatic reticulum
Fuc: Fucose
Gal: Galactose
GalNAc: N-acetylgalactosamine
GDP: Guanidin diphosphate
Glc: Glucose
GlcNAc: N-acetylglucosamine
GNA: Snowdrop *Galanthus nivalis* lectin
GPI: Glycosylphosphatidylinositol
HRP: Horseradish peroxidase
IMAC: Immobilized metal affinity chromatography
K_{av}: Partition coefficient
LCA: *Lens culinaris* lectin
LC-MS: Mass spectrometry interfaced with RP-HPLC
LLO: Lipid linked oligosaccharide
m/z: Mass to charge ratio
Man: Mannose
MBL: Mannose binding lectin
MRM: Multiple reaction monitoring mode

MS: Mass spectrometry

MSMS: tandem mass spectrometry

MW: Molecular Weight

NeuNAC: N-acetyl neuramic acid

OST: Oligosaccharyltransferase

PAGE: Polyacrylamide gel electrophoresis

PBS: Phosphate buffered saline

PNA: Peanut agglutinin

RP-HPLC: Reversed-phase High-performance liquid chromatography

rpm: revolutions per minute

Rs: Stokes radius

TEV: Tobacco-etch virus

TFA: Trifluoro acetic acid

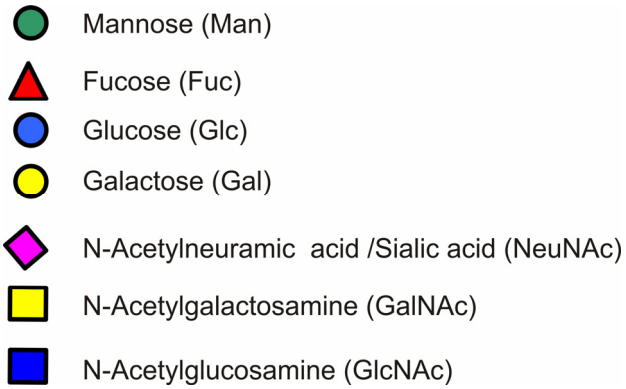
Tris: α,α,α-tris(hydroxymethyl)methylamine

TSR: Thrombospondin-Type 1 repeat

UDP: Uridine diphosphate

Xyl: Xylose

Sugar symbols



1 Introduction

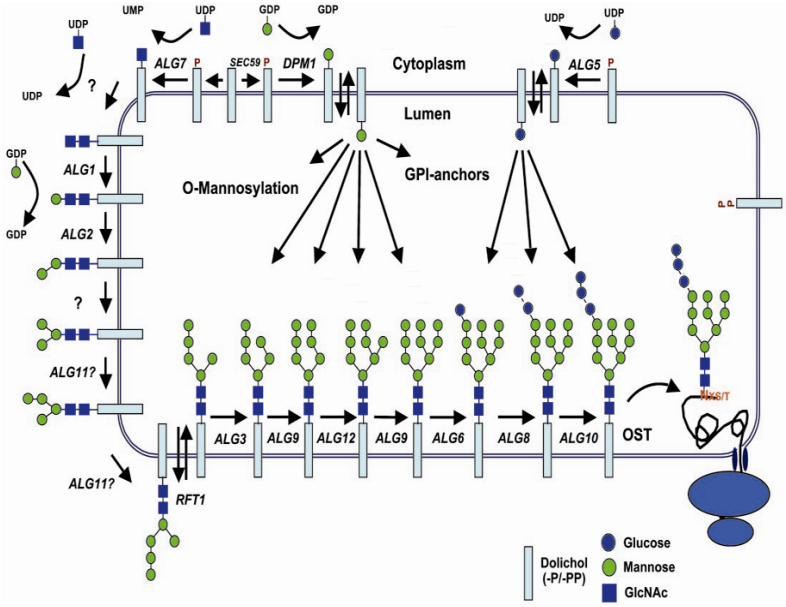
1.1 Introduction into glycosylation

Glycosylation involves the transfer of carbohydrates onto proteins, lipids or carbohydrates themselves. It is the most abundant and extensive modification in higher-ordered cells. Carbohydrates can be transferred by specific enzymes or in a non-enzymatic process termed glycation. Carbohydrates, oligosaccharides and polysaccharides are summarized in the collective term glycans. The different glycans occurring in mammals are often summarized as the mammalian glycome. This glycome is evolutionarily conserved and its importance is reflected in the fact that 1-2% of the genome encodes for substrates or cellular enzymes involved in glycosylation (*1*). Nine different sugar precursors can be coupled enzymatically onto proteins or lipids giving rise to 14 different glycans. These monosaccharides can then be extended by one out of 49 different glycosyltransferase reactions (*1*). Depending on the amino acid to which the sugars are transferred, protein glycosylation can be *N*-linked (asparagine), *O*-linked (serine or threonine) or *C*-linked (tryptophan). Additional amino acids that can be glycosylated include hydroxylysine (*2*), cysteine (*3*) hydroxyproline (*4, 5*) and tyrosine (*6*). The focus of this introduction will be carbohydrates *N*-,*O*- or *C*-linked to mammalian proteins. Large glycosaminoglycans will not be discussed (reviewed in (*7, 8*))

1.2 *N*-glycosylation

N-glycosylation occurs on the asparagine in the consensus sequence Asn-X-Ser/Thr (so-called sequon). It is found in all eukaryotic cells and in prokaryotes such as the mucosal gut pathogen *Campylobacter jejuni* (reviewed in (*9*)). *N*-glycosylation is initiated on the cytoplasmic side of the ER when the precursor dolichol-pyrophosphate receives the oligosaccharides Man_5 $GlcNAc_2$ by sequential biosynthesis steps, involving seven different enzymes using UDP-GlcNAc and GDP-Man as sugar precursors. This lipid-linked oligosaccharide (LLO) is then translocated into the luminal space of the ER. The reaction is believed to be catalyzed by a bidirectional flippase, which in yeast requires the protein Rft1 (*10, 11*). Upon translocation, four mannose and three glucose moieties

from the precursor dolichol-P-Man or dolichol-P-Glc respectively are transferred onto the LLO. The terminal α-1,2 linked Glucose moiety serves subsequently as a recognition signal for the multisubunit enzyme oligosaccharyltransferase (OST), which transfers the LLO onto the Asn of the sequon (Figure 1). The OST is closely connected to the translocation complex and is tightly associated with the chaperone BIP (GRP78) and the lectins calnexin and calreticulin. It is believed that the OST complex scans the nascent polypeptide chain, and once the polypeptide chain as has reached a length of 12-14 amino acids (30-40Å) the glycans are transferred. The *N*-glycans protrude about 30Å from the protein backbone and depending on localization and protein, they either facilitate or are even obligatory for proper folding. The lectins calnexin and calreticulin play an important role in the so-called quality control mechanism where *N*-glycosylated proteins are scanned for their native structure.



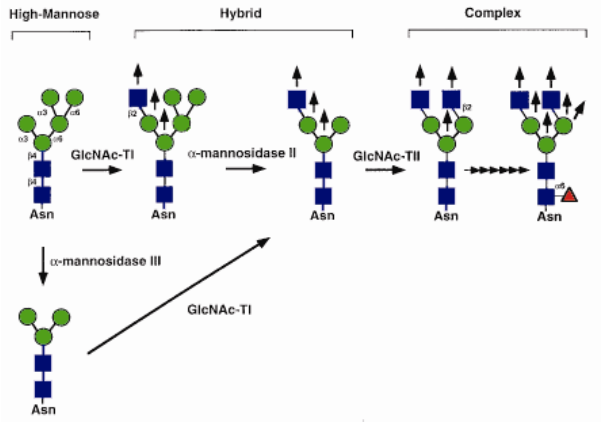
adapted from (*10*) modified

Figure 1: *N*-glycosylation in the ER

N-glycosylation starts on the cytosolic side of the ER where the lipid carrier dolichol-pyrophosphate receives Man_5 $GlcNAc_2$ through successive addition by seven different enzymes. The lipid-linked oligosaccharide flips into the ER lumen, where it is extended with mannose and glucose into the core *N*-glycan Glc_3-Man_9-$GlcNAc_2$. The core *N*-glycan is subsequently transferred by the large oligosaccharyltransferase complex onto the sequon (NXT or NXS) of nascent polypeptide chains.

The Glc_3-Man_9-$GlcNAc_2$ core *N*-glycan is subsequently trimmed by two sequential glucosidases. Glucosidase I (GI) recognizes the terminal α1,2Glc moiety and glucosidase II (GII) removes the second α1,3-linked glucose. GII has a carbohydrate recognition domain (CRD) for α1,3-linked glucose and one mannose-6-phosohate receptor subunit, which recognizes the tetra-mannose branch of a secondary *N*-glycan (*12*). The terminal α1,3-linked glucose is then recognized by the globular domain of either calnexin or calreticulin which is further associated with a phospho-disulfide isomerase (PDI) termed ERp57. Together these lectins and chaperons assist folding of the protein. Upon release from calnexin and calreticulin the α1,3-linked glucose is removed by GII. Native proteins can now leave the ER to the Golgi where their *N*-glycan is further processed. Unfolded or incompletely folded proteins are recognized by uridine diphosphate glycosyltransferase 1 (UGT1) which re-glycosylates the terminal α1,3Glc. The binding of UGT1 is believed to be dependent on non-native hydrophobic patches or unpaired cysteins (*13*). Re-glycosylation of glucose results in re-association with calnexin and calreticulin and the protein undergoes a second round of chaperone-assisted folding (i.e. calnexin-calreticulin cycle). If the protein can not adopt its native conformation and shuttles continuously between GII-deglycosylation and UGT1-mediated re-glycosylation the protein is finally translocated into the cytosol for destruction (ER-associated destruction ERAD). The molecular switch between re-glycosylation by UGT1 or destruction in the ERAD is still a matter of debate. It was shown that UGT1 can only re-glycosylate nearly-native but not completely misfolded proteins. Maturing proteins can spend quite some time in the folding cage of the calnexin-careticulin cycle (*14*). If a protein is destined for destruction, it associates with chaperones like BIP. BIP goes hand-in-hand with other chaperones and specific ER-α1,2-mannosidases of the EDEM (ER degradation enhancer, mannosidase alpha-like) family. These glycosidases remove mannose from the core mannose glycan which renders proteins unrecognizable by calnexin, calreticulin and UGT1 (*13, 15*). Finally retro translocation of terminally misfolded proteins can be carried out through a porous channel termed Sec61 (*16*) into the cytosol where proteins become polyubiquitinated and are subsequently degraded by 26S proteasome. *N*-glycans of proteins that successfully folded into their native structure are trimmed by an α1,2-mannosidase

yielding Man_8-$GlcNac_2$-Asn *N*-glycans and are subsequently translocated to the Golgi. The *N*-glycosylated protein can then either be modified by GlcNAc phosphotransferase, which upon trimming results in high mannose-type glycans containing mannose-6-phosphate. This structure is a recognition signal for the Mannose-6-phosphate receptors, which transport their cargo protein to the endosomes (*17*). Alternatively, various α1,2-mannosidases can remove mannose moieties resulting in the processed high mannose Man_5-$GlcNAc_2$-structure. This structure can then be elongated with GlcNAc, Gal, Fuc or sialic acid either into the hybrid type (i.e. substituted GlcNAc and unsubstituted Man residues) or into the complex type (both α2,3 and α2,6Man are substituted with GlcNAc). Hybrid as well as complex *N*-glycans can contain GlcNAc bearing branches commonly known as antenna. Different *N*-glycan structures can occur on different sites of the same protein, commonly known as micro heterogeneity, depending on accessibility and availability of sugar donors to glycosyltransferase (*8*) (Figure 1.2). As the *N*-glycan protein traverses through the Golgi it becomes a substrate for various glycosyltransferases that modify the different antenna with GlcNAc, Gal or sialic acids resulting in a diversified carbohydrate tree with various functions.



adopted from (*8*) modified

Figure 2: Types of *N*-glycans in the Golgi

The processed glycan Man_5-$GlcNAc_2$ that is produced in the ER serves as a substrate of *N*-glycan diversification in the Golgi. Addition of a GlcNAc to the first branch by GlcNAc-TI (encoded by Mgat1) in β1,2 position produces *N*-glycan of the hybrid type, which can be trimmed by α-mannosidase II cleaving α1,3 and α1,6-linked mannose. Alternatively, hybrid types can also be produced through mannose trimming by α-mannosidase III which is a substrate for GlcNAc-TI. Addition of a second GlcNAc by GlcNAc-TII catalyzes the conversion of hybrid to the complex type, which eventually can be core fucosylated in the α1,6 position on the first GlcNAc linked to Asn.

1.2.1 Function and disorders of *N*-glycosylation

N-glycosylation is occurring on virtually all proteins that traverse the secretory pathway. Its importance is reflected in experiments performed with tunicamycin the competitive inhibitor of the enzyme that transfers the first GlcNAc onto dolchichol (GPT: UDP-GlcNAc: dolicholphosphate *N*-acetylglucosamine-1-phosphate-transferase), which is associated with severe cellular malfunctions and death in cells. Furthermore, mice with a targeted deletion of GPT die at embryonic day 5.5 (*7*), implicating that mammals having a complete absence of *N*-glycans do not survive to term. Not only the complete absence but also the lack of diversification in *N*-glycans is embryonically lethal as demonstrated by Mgat1 knock-down animals. Mgat1 encodes for the GlcNAc transferase that converts high-mannose *N*-glycans to hybrid *N*-glycans. Knock-out mice for this key-enzyme die at embryonic day 9.5-10.5 due to defects in neural tube formation and vascularization, and severe heart abnormalities (*18*). In humans, genetic defects resulting in incomplete glycan assembly are commonly referred as congenital disorders of glycosylations (CDG). CDGs of *N*-glycosylation affect about 600 people world wide and are subdivided into Type I, affecting one of the steps in biosynthesis or addition of the LLO to the protein (CDG Ia-IL) and Type II, which affects trimming, remodeling or extensions of the glycan (CDG IIa-IIf). Type I CDGs result from incomplete or LLO precursors that are ineffectively transferred to proteins. The symptoms are often mental retardation, hypotonia, or seizures but can vary considerably amongst patients. Type II CDGs alter the processing of protein *N*-glycans. In patients, these diseases are often characterized by growth and mental retardation, dysmorphism and hypotonia. In CDG Type II patients, a new group is emerging that are defective in functions attributed to the COG (Conserved Oligomeric Golgi) genes. These genes encode an eight-subunit protein complex that is involved in intra-Golgi retrograde trafficking mediated by coat protein I vesicles (*19*). CDG patients with mutations in these genes show various alterations in glycosylation. In future, this group of CDGs will probably be further extended with patients, which could not been classified into the known CDG classes (*20*). The suspicion for the presence of a CDG, arises clinically on the observed symptoms of patients. Grouping into the known CDG subclasses can then be done based on the glycan profile of marker proteins like transferrin or α1-antitrypsin. However, the

emerging numbers of patients classified as CDG-IIx, which can not be grouped into the known classes, require novel marker proteins and analysis tools. Especially, the application of mass spectrometry provides possibilities to identify CDG-patients (*21*). Additionally, there are diseases associated with *N*-glycosylation which are not listed as CDGs. Patients suffering from Mucolipidosis I or II have defects in the GlcNAc phosphotransferase complex required for the targeting of mannose rich *N*-glycans to the lysosomes (*17*). Other Type II diseases include congenital dyserythropoietic anemia and a recently described disease mapped to the a2 subunit of the V-type H^+ ATPase (*20, 22*). A list of the currently known CDGs is given in Table 1 (taken from (*20*)).

Disorder	Gene	Enzyme	OMIM	Key Features
CDG-Ia	PMM2	Phosphomannomutase II	212065	Mental retardation, hypotonia, esotropia, lipodystrophy, cerebellar hypoplasia, stroke-like episodes, seizures
CDG-Ib	MPI	Phosphomannose isomerase	602579	Hepatic fibrosis, protein-losing enteropathy, coagulopathy, hypoglycaemia
CDG-Ic	ALG6	Glucosyltransferase I Dol-P-Glc: Man9-GlcNAc2-P-P-Dol glucosyltransferase	603147	Moderate mental retardation, hypotonia, esotropia, epilepsy
CDG-Id	ALG3	Dol-P-Man:Man5-GlcNAc2-P-P-Dol mannosyltransferase	601110	Profound psychomotor delay, optic atrophy, acquired microcephaly, iris colobomas, hypsarrhythmia
CDG-Ie	DPM1	Dol-P-Man synthase I GDP-Man: Dol-P-mannosyltransferase	603503	Severe mental retardation, epilepsy, hypotonia, mild dysmorphism, coagulopathy
CDG-If	MPDU1	Man-P-Dol utilization 1/Lec35	608799	Short stature, icthyosis, psychomotor retardation, pigmentary retinopathy
CDG-Ig	ALG12	Dol-P-Man:Man7-GlcNAc2P-P-Dol mannosyltransferase	607143	Hypotonia, facial dysmorphism, psychomotor retardation, acquired microcephaly, frequent infections
CDG-Ih	ALG8	Glucosyltransferase II Dol-P-Glc: Glc1-Man9-GlcNAc2-P-P-Dol glucosyltransferase	608104	Hepatomegaly, protein-losing enteropathy, renal failure, hypoalbuminaemia, oedema, ascites
CDG-Ii	ALG2	Mannosyltransferase II GDP-Man: Man1-GlcNAc2-P-P-Dol mannosyltransferase	607906	Normal at birth; mental retardation, hypomyelination, intractable seizures, iris colobomas, hepatomegaly, coagulopathy
CDG-Ij	DPAGT1	UDP-GlcNAc: Dol-P-GlcNAc-P transferase	608093	Severe mental retardation, hypotonia, seizures, microcephaly, exotropia
CDG-Ik	ALG1	Mannosyltransferase I GDP-Man: GlcNAc2-P-P-Dol mannosyltransferase	608540	Severe psychomotor retardation, hypotonia, acquired microcephaly, intractable seizures, fever, coagulopathy, nephrotic syndrome, early death
CDG-Il	ALG9	Mannosyltransferase Dol-P-Man: Man6- and Man8-GlcNAc2-P-P-Dol mannosyltransferase	608776	Severe microcephaly, hypotonia, seizures, hepatomegaly
CDG-IIa	MGAT2	GlcNAc transferase 2	212066	Mental retardation, dysmorphism, stereotypies, seizures
CDG-IIb	GLS1	Glucosidase I	606056	Dysmorphism, hypotonia, seizures, hepatomegaly, hepatic fibrosis; death at 2.5 months
CDG-IIc	SLC35C1/FUCT1	GDP-fucose transporter	266265	Recurrent infections, persistent neutrophilia, mental retardation, microcephaly, hypotonia; normal transferrin
CDG-IId	B4GALT1	beta1,4 galactosyltransferase	607091	Hypotonia (myopathy), spontaneous haemorrhage, Dandy–Walker malformation
CDG-IIe	COG7	Conserved oligomeric Golgi complex subunit 7	608779	Fatal in early infancy; dysmorphism, hypotonia, intractable seizures, hepatomegaly, progressive jaundice, recurrent infections, cardiac failure
CDG-IIf	SLC35A1	CMP-sialic acid transporter	605634	Thrombocytopaenia, no neurological symptoms; normal transferrin, abnormal platelet glycoproteins
CDG-II/COG1	COG1	Conserved oligomeric Golgi complex subunit 1	606973	Hypotonia, growth retardation, progressive microcephaly, hepatosplenomegaly, mild mental retardation
Mucolipidosis II and III	GNPTA	UDP-GlcNAc: lysosomal enzyme, GlcNAc-P transferase	252500	Coarsening features, organomegaly, joint stiffness, dysostosis, median neuropathy at the wrist; MLIII is less severe than MLII, which presents in infancy
Congenital dyserythropoietic anaemia (CDA II)	Unknown	Unknown	224100	Anaemia, jaundice, splenomegaly, gall bladder disease
autosomal recessive cutis laxa (ARCL) type II [a]	ATP6V0A2	a2 subunit of the V-type H+ ATPase	219200	growth and developmental delay , bone dystrophy, joint laxity and retarded development

[a]: adapted from (*22*)

Abbreviations: CDG-congenital disorder of glycosylation CMP-Cytidine monophosphate, Dol-dolichol GDP:guanidine diphosphate Glc-glucose GlcNAc-*N*-acetylglucosamine, Man-mannose, UDP-uridine diphosphate

Table 1: Diseases associated with *N*-glycosylation

1.3 *O*-glycosylation

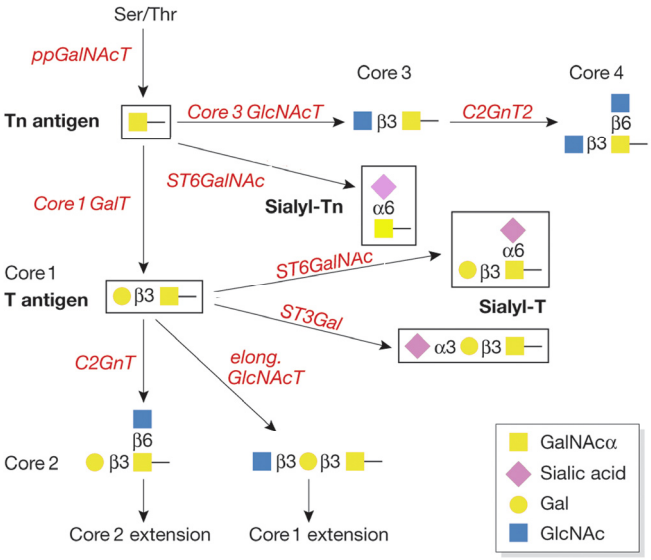
Another protein glycosylation is *O*-glycosylation. Serine or threonines can be *O*-glycosylated by *N*-acetylgalactosamine, fucose, glucose, mannose, *N*-acetylglucosamine or xylose (*1*). The most common *O*-glycosylation is initiated by a GalNAc residue through a α-glycosidic bond to Ser/Thr residues. This *O*-glycosylation has been termed mucin-type *O*-glycosylation.

1.3.1 Mucin-type *O*-glycans

Mucin-type *O*-glycosylation is initiated by a GalNAc moiety transferred by one of the UDP-GalNAc: polypeptide *N*-acetylgalactosaminyltransferases (GalNAc Ts). The modification was originally described to occur on the mucins-highly glycosylated proteins that form the mucus of the respiratory or digestive tract. Meanwhile, these glycans have also been described on a variety of different proteins like erythropoietin (*23*), podoplanin (*24*) or the transmembrane glycoprotein CD44v6 (*25*). In contrast to *N*-glycosylation in the ER, which occurs co-translationally, *O*-glycosylation of the mucin-type occurs post-translationally in the Golgi apparatus. The GalNAc-T family comprises 15 members with five or more still predicted (*26*). It has been reported that these enzymes are regulated in different tissues or expressed in different stages of development, but they can also be expressed simultaneously. Redundant expression of several GalNAc T homologues often comes along with overlapping substrate specificity. An observation that might explain the lack of phenotype observed in knock-out animals (*27*). Some GalNAc-Ts are also restricted in their substrate repertoire and their specific glycosylation pattern has been shown to be important in development (*28, 29*). *O*-glycosylation of the mucin-type often appears as highly glycosylated clusters. The density of these clusters is influenced by competing GalNAc Ts having peptide or glycopeptide activity (*30*). Since not all possible residues are equally modified, the existence of a sequence comparable to the sequon in *N*-glycans had been proposed. Unfortunately, a general valid consensus sequence could not be defined. This is largely due to a large number of peptides required to characterize the requirements of all the GalNAc T family members *in vitro*

and various problems like redundancy of transferases *in vivo (26).* Threonines rather than serines have been demonstrated to be preferred residues for GalNAc transferases (*31*). In the neighborhood of the glycosylated amino acid uncharged hydrophobic residues like Pro at position -1 and +3 (relative to the glycosylated residue) are favored, whereas charged residues at the same positions are inhibitory (*26, 30*). Many peptides can be glycosylated *in vitro* even at residues which are cryptic under native conditions in the protein, indicating that the glycosylated residue *in vivo* needs to be exposed preferably in an extended conformation (β-strand conformation) at the surface of the protein (*31*).

Once the GalNAc moiety has been added onto the substrate, the GalNAc (also known as the Tn-antigen) can be extended into various glycan structures termed core structures. The core structures 1-4 (see Figure 3) are the most common. Core structures 5-7 have been described to occur only in specific tissues or at specific developmental stages (*8*). Under pathological conditions e.g. in cancer, the Tn-antigen can be sialylated by α6-sialyltransferase (ST6GalNAc) into sialyl-Tn. Extension of the Tn-antigen with Gal in β1,3-linkage by the Core 1 β1,3-galactosyltransferase (T-synthase) results in the Core 1 structure, which is also called T-antigen. The Core structure can also be sialylated by α,3-sialyltransferase (ST3Gal) or ST6GalNAc into sialyl-T-antigen. Tn, sialyl-Tn, T-antigen and its sialylated variants are commonly referred to Thomsen-Friedenreich related antigens. Using the Core 1 structure as building block for further extension, three different Core 2 GlcNAc transferases (C2GnT) can transfer a GlcNAc onto the GalNAc in β1,6-linkage. Alternatively, to the formation of the Core 1 glycan, the Tn antigen can also be extended by Core 3 β1,3-GlcNAc transferase (Core 3 GlcNAc-T) with GlcNAc in β1,3-linkage to the GalNAc. Core 3 can now itself form a building block for Core 4 which is performed by C2GnT2 adding a GlcNAc in β1,6-linkage. All core structures can be extended with multiple Galβ1,4GlcNAc (lactosamine) structures in a mono-or biantennary fashion. These glycans can additionally be sialylated, fucosylated, GalNAcylated or sulfated at the end.



adapted from (*32*) modified

Figure 3: Core structures of Mucin type *O*-Glycans

Mucin-type *O*-glycosylation is initiated by transfer of a αGalNAc moiety by a GalNAc T. This structure is also called Tn antigen, which can be sialylated by (ST6GalNAc) to yield sialyl-Tn. Extension with Gal, results in the Core 1 structure also called T-antigen. T-antigen can be sialylated by ST6GalNAc or ST3Gal resulting in sialyl-T-antigen. The Core 1 structure can be alternatively, extended to the Core 2 structure by C2GnT. Glycans of Core 3 structures are obtained by modification of the Tn antigen with Core GlcNAc-T. The product is a substrate for C2GnT2 to yield Core 4. All the Core structures can be further extended as mono-or biantennary glycans.

1.3.1.1 Functions and diseases of O-glycosylation

Similar to *N*-glycosylation, *O*-glycosylation of the mucin-type is required for various functions, ranging from a barrier function to leukocyte homing (reviewed under (*33*)). These glycans are often tightly regulated and their expression is altered under physiological as well pathological conditions. A well-known function of *O*-glycans is the formation of the ABO blood group antigens. The blood group system is based on antibodies recognizing specific carbohydrates located on *O*-glycans and glycolipids. The precursor glycan based on multiple lactosamine (Galβ1,4-GlcNAc) moieties is fucosylated by α1,2-fucosyltransferase adding a fucose in α1,2-linkage onto the terminal galactose. The unmodified glycan is called the H-antigen and defines blood group 0. Alternatively, it can be extended by α1,3 GalNAc T (blood group A) or α1,3

galactosyltransferase (blood group B). These glycans are displayed on the surface of erythrocytes. They trigger the formation of IgM antibodies directed against glycans of the ABO system which are absent in the individual. The “matching” of the correct blood group system in transfusion medicine is important to prevent blood transfusions reactions due to antibody-and complement-dependent red cell lysis (*8*).

The importance of mucin-type glycosylation in development is illustrated with the finding that mutations in the GalNAc T3 are associated with a CDG termed familial tumoral calcinosis (FTC). FTC is an autosomal recessive disorder characterized by phosphatemia and calcium deposits in the skin and subcutaneous tissues. It was found that this disease can result from nonsense mutations in the gene encoding for GalNAc T3 (GALNT3) (*29*) or from mutations in the regulator for calcium homeostasis FGF23 (*34*). FGF23 contains a Thr^{179} residue adjacent to a subtilisin-like pro-protein convertase cleavage site. It was demonstrated that this residue is *O*-glycosylated by GalNAc T3. Further processing of the *O*-glycan into the sialylated Tn or the sialyl-T-antigen with a α2,6-sialylated GalNAc abrogated the proteolytic cleavage. It was suggested that a competition between *O*-glycosylation and proteolytic cleavage tightly regulates FGF23 and that deregulation might be a possible reason for the disease (*35*). Another disease where the Tn antigen plays a major role is the so-called Tn-syndrome. The Tn-syndrome is an autoimmune disease where subpopulations of blood cells are *O*-glycosylated only with the Tn antigen resulting in anemia (i.e. decreased number of red blood cells), leukopenia (i.e. decreased number of leucocytes) or thrombocytopenia (i.e. decreased number of platelets). The molecular defect for this disease was attributed to the protein Cosmc. Cosmc is a chaperone, which is required by the Core 1 β1,3-galactosyltransferase for its function and expression (*36*). In Tn syndrome, Cosmc is mutated in subpopulations of hematopoetic stem cells leading to expression of the autoimmune Tn antigen on blood cells from different lineages.

In cancer, the mucin-type *O*-glycans are often altered. In colon cancer, the Tn, sialyl-Tn and T-antigens have been described to be significantly increased. Normal colonic epithelium expressed *O*-glycans of Core 3. In cancer, a shift of glycosylation results in the increased formation of shorter glycans of the Thomsen-Friedenreich related antigens. This was attributed to a decrease of Core 3 GlcNAc-T in favor of expression of

increased Core 1 glycans. Additionally, increased expression of sialyltransferases and *O*-acetyltransferases modifying sialic acids simplify *O*-glycan biosynthesis in colon cancer (*32, 37*). Similarly, in breast cancers *O*-glycans are also simplified. These cells often show increased expression of sialyltransferases in particular ST6GalNac-I throughout the Golgi. This enzyme competes with Core 1 transferase resulting in increased expression of sialyl-Tn structures, which is associated with poor prognosis and formation of metastases.

1.3.2 *O*-fucosylation

Another functionally relevant *O*-glycosylation is *O*-fucosylation. This modification involves an L-fucose α1-linked to the hydroxyl group of Ser or Thr residues. This modification has been found on epidermal growth factor like repeats (EGF) on thrombospondin type 1 repeats (TSRs) and on the protease inhibitor from *Locusta migratoria* (*38-40*). *O*-fucosylation of EGF repeats is performed by protein *O*-fucosyl transferase 1 (*41*) (POFUT1) and TSRs are modified by POFUT2 (*42*). Both ER resident enzymes have been shown to be unique for their substrate and cannot glycosylate the other (*43*).

1.3.2.1 O-fucosylation of Notch

POFUT1 transfers a fucose moiety onto serine or threonine in the recognition motif $C^2X_{4-5}S/TC^3$ of properly folded EGF repeats (*41*). A SwissProt database search against the consensus sequence revealed 102 proteins, potentially modified with *O*-fucose. Within these proteins, the transmembrane receptor Notch was present. Notch is a single-pass transmembrane receptor forming a heterodimer upon processing by a furin-like convertase in the Golgi (S1 cleavage). In mammals four Notch receptors are known (Notch 1-4) containing 36, (Notch 1,2), 34 (Notch 3) or 29 (Notch 4) EGF repeats and 3 NOTCH/LIN repeats on the extracellular part. The intracellular part encompasses 6 ankyrin repeats, a transactivation domain (TAD) and a PEST domain (*44*). The importance of Notch in particular Notch 1 and 2 is reflected in embryonic lethality at day 11.5 in homozygous knock-out mice (*45, 46*). Notch receptor signaling occurs by

interaction with transmembrane ligands on adjacent cells. These ligands are called DSL (Delta, Serrate or Lag). In mammals, Serrate is called Jagged and two Jagged (Jagged 1 and 2) and four Delta-like ligands (Delta –like 1-4) have been described. Upon interaction with one of the ligands, the Notch receptor changes its conformation. Subsequent proteolytic cleavage inside and outside the membrane (S2 and S3 cleavage) releases the intracellular domain. The intracellular domain interacts with transcription factors and initiates the transcription of target genes (*44*). Interestingly a POFUT1 knock-out in mice displayed similar phenotypes as observed in knock-out from Notch downstream targets, implicating the importance of *O*-fucosylation in Notch signaling (*47*). 21 out of 36 EGF repeats in Notch 1 contain the recognition motif for POFUT1 and can undergo *O*-fucosylation. EGF repeats on Notch can either be glycosylated with a single *O*-fucose or extended to the tetrasaccharide NeuNAcα2,3/6-Gal-β1,4-GlcNAc-β1,3-Fuc-α1-*O*-Ser/Thr (*48*). The GlcNAc residue is added onto fucose by the fucose-specific β1,3-GlcNAc transferase fringe. In *Drosophila,* fringe was shown to be important for boundary formation in the wings (*49*). In mammals, three fringes are known which are called Manic, Lunatic, and Radical fringe. Lunatic fringe knock-down mice have been shown to display multiple phenotypes in somitogenesis. Radical fringe seems to be important for chick limb growth (reviewed in (*44, 48*)). Different fringes display different preferences for EGF repeats, which are necessary to modulate Notch signaling (see below). Addition of Galcatose onto the disaccharide GlcNAc-β1,3Fuc-α-Ser is performed by β4Gal-T1 which also modulates Notch activity in a Chinese hamster ovary (CHO) expression model (*50*) but was shown to be dispensable in an *in vitro* system with purified components of the *Drosophila* Notch-signaling pathway (*51*). The α2,3/6-linked sialic acid is transferred by unknown sialyltransferases. This sugar was demonstrated to be unimportant for Notch signaling (*44*).

1.3.2.1.1 Functions and diseases

As mentioned above POFUT1 knock-out in mice phenocopies a Notch knock-out and shows a more severe phenotype than mice or flies lacking fringe (*44*). These results suggest that the protein plays an important role in development. If this function is located in the fucose moiety or in POFUT1 is currently a matter of discussion. In *Drosophila* a chaperone function of the fly POFUT1 orthologue (in *Drosophila* OFUT1) was demonstrated independent of its glycosyltransferase activity. It was discovered that Notch accumulated in the ER when OFUT1 was depleted in cells in the *Drosophila* imaginal wing disc. This phenotype could be rescued by a mutant murine POFUT1 where the GFD-fucose binding site had been mutated. It was shown that the mutant had lost its glycosyltransferase activity but retained a chaperone activity, necessary exclusively for intracellular trafficking of Notch (*52*). Furthermore, a recent study demonstrated, that Drosophila embryos expressing an OFUT1 mutant lacking its enzyme activity but retaining its chaperone activity, completed embryogenesis and hatched without grossly evident abnormalities. The authors speculated that the glycosyltransferase function of OFUT1 is not necessary for all Notch signaling (*53*). By contrast, in mammalian cells lacking POFUT1 Notch receptors are equivalently expressed on the surface (*54*), requiring further experimental proof of the chaperone hypothesis of POFUT1 in Notch signaling.

An important role of the *O*-fucose moiety was proposed in an *in vitro* co-culturing assay. Co-culturing with L-cells expressing Jagged1 resulted in a four-fold induction of Notch signaling in wild type CHO cells. This induction was due to *O*-fucosylation as demonstrated by the CHOlec13 cell line that transfers only minimal levels of fucose to glycoconjugates (i.e. deficiency in GDP-D-mannose-4,6dehydratase). Co-culturing of CHOlec13 together with L-cells decreased significantly Notch signaling and could partially be restored by adding fucose into the medium (*49*). These experiments argue that either the fucose moiety alone or the lack of fringe mediate the observed effects. In the CHO co-culturing assay fringe, inhibited Jagged1 induced Notch signaling and in *Drosophila* ectopically expressed fringe eliminated the normal fringe expression border in the wing and induced a novel expression border in ventral cells (*49*). It was shown that the three fringes in mammals modulate Notch signaling in a different manner. While

lunatic fringe enhanced signaling mediated by Delta and inhibited signaling by Jagged 1, manic fringe inhibited only the Jagged 1 signaling. Radical fringe even enhances Delta and Jagged 1 signaling. It was proposed that the different specificities of the fringes on the EGF repeats can increase or attenuate binding of the ligands and therefore mediate a specific Notch signal (*55*). These results indicate that regulation of Notch signaling strongly depends on the glycosylation by protein *O*-fucosyltransferases or the fringes, which can influence directly the interaction with Delta or Jagged (Serrate in *Drosophila*). How the individual proteins and carbohydrates mediate exactly the observed effects is still a matter of debate.

Due to the significance of Notch signaling in development various diseases are known where correct Notch signaling is disturbed. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is a human disease characterized by repeating ischemic attacks resulting in dementia and decline. It was discovered that mutation of the Notch 3 gene accounts for this disease. Mutations were predominantly present in the first 5 EGF domains and often resulted in the replacement of a conserved residue with a cysteine or the mutation of a cysteine. These findings suggested a phenotype due to misfolding or defective receptor trafficking. Investigation of the glycosylation revealed normal *O*-fucosylation but reduced elongation by Lunatic fringe and increased tendency for aggregation as hetero-or homodimers (*56*). In cancer, Notch overexpression was observed in T cell acute lymphoblastic leukemia (T-ALL) where often a translocation leads to a constitutive active Notch 1 protein. The glycosylation of Notch 1 in T-ALL has not been studied in detail. However, the observation that Notch1 and Lunatic fringe regulate T cell progenitors and suppress B cell potential (*57*) implies the involvement of fringe and probably POFUT1 in cancer (*44*).

Besides Notch, an effect on signaling of *O*-fucosylation has also been observed in the urokinase-plasminogen activator (uPA). uPA binds a specific receptor and this triggers cell proliferation in Sa-OS-2 osteosarcoma cells. It was shown that de-fucosylated uPA retains its receptor binding activity but loses its growth promoting activity. A phenomenon attributed to *O*-fucosylation of Thr^{18} (*58*). Initially it was also thought that also *O*-fucosylation of Cripto is important for signaling. Cripto is a GPI-anchored co-receptor for the Nodal receptor, which is essential for mesoderm formation and left-right

asymmetry. It was shown earlier in signaling assays, that Cripto-nodal interaction required an *O*-fucoslyated Thr (*59*) However, later mutagenesis work revealed, that only the Thr but not the *O*-fucosyl moiety mediated the interaction required for signaling (*60*).

1.3.2.2 O-fucosylation of Thrombospondin Type 1 repeats (TSRs)

L-Fucose can also be added on serines or threonines in the consensus sequence $C^1X_{2-3}S/TC^2X_2G$. This type of *O*-fucosylation has been described to occur on Thrombospondin type 1 repeats (TSRs) (view chapter 4.1.1 for a detailed description of TSRs). Meanwhile, the modification has been demonstrated on TSRs of thrombospondin-1 (*39*), the axonal guidance protein F-spondin (*61*), the positive regulator of complement properdin (*62*), the metalloprotease ADAMTS13 (*63*) and the metalloprotease-like protein puctuin-1 (*64*). The fucose moiety is transferred by a POFUT1 homologue designated POFUT2. This enzyme has been shown to be localized in the ER, although it lacks ER-retention signals like POFUT1 (*65*). Similarly, to POFUT1, which modifies only properly folded EGF repeats, POFUT2 was shown to require a properly folded TSR (*42, 43*). The fucose moiety on TSRs can be extended with glucose to the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr (*43, 61, 66*). Recently, the β1,3-Glucosyltransferase (β3Glc-T) responsible for adding the glucose was identified and characterized (*65, 67*). Similarly, to POFUT1 also β3Glc-T required a properly folded TSR module. Interestingly it was found that β3Glc-T contained a REEL-sequence at its C-terminus, which is similar to the KDEL sequence an ER-retention signal. As previously mentioned POFUT2 lacks an ER-retention signal but was found to localize in the ER. Since the activity of POFUT 2 and β3Glc-T strongly overlaps it was speculated that both enzymes could interact (*65*).

1.3.2.2.1 Function and diseases

POFUT2 is conserved from protozoa to mammals. The *C. elegans* POFUT2 orthologue is called PAD-2. This enzyme has been shown to be the O-fucosyltransferase in the nematode that modifies TSRs. Targeted deletion of the PAD-2 in *C. elegans* resulted in loss of enzymatic activity. Mutant animals were homozygous viable and fertile but displayed a gonad phenotype. The bilobed gonad is the reproductive organ of the nematode; its shape is determined by migration of the two distal tip cells (DTCs). During development in L2 larvae stage, the DTCs start on the ventral side of the animal and migrate horizontally. Subsequently the DTCs turn towards the dorsal side and after a right angle turn they migrate centripetally along the dorsal site of the dorsal body wall muscles. The migration is depended on attractive and repulsive cues of the netrin and TGFβ signaling pathways. The pad-2 mutant displayed an early dorsal migration phenotype of the anterior arm of the gonad. Genetic analysis revealed that pad-2 acts upstream or parallel to the netrin pathway. This is the first report that the *O*-fucosyltransferase PAD-2 is important role in cell migration in a multicellular organism (*68*).

In another study using the TSR-containing metalloproteinase ADAMTS13, the *O*-fucosylated serines of TSR1-8 were mutated to alanines. This resulted in decreased secretion efficiency of the full-length ADAMTS13 protein compared to wild type. Mutagenesis of two particular *O*-fucosylation sites on the TSRs or RNAi of POFUT2 strongly diminished secretion. Overexpressed POFUT 2 on the other hand, was able to rescue the secretion of two mutated TSRs (*63*). A similar result was obtained with the ADAMTS metalloproteinase-like protein punctuin-1. Here, mutations in TSR2-4 and culturing in the GDP-D-mannose-4,6-dehydratase deficient cell line CHOlec13 resulted in significant decreased secretion levels (*64*). Based on these result, the authors suggested that POFUT2 like POFUT1 may be involved in quality control recognizing correctly folded modules. However, unlike in POFUT1, in POFUT2 the glycosyltransferase activity was still required (*63*).

Another function for *O*-fucosylation on TSRs was recently discovered by the identification of human disease termed Peter-Plus syndrome. It was reported that truncating mutations in a gene designated B3GALTL are responsible for the clinical

manifestation. The disease is characterized by anterior eye chamber defects, short statue, developmental delay and characteristic craniofacial features. DNA sequencing revealed that, all examined Peters-Plus patients had acquired a c.1020+1G→A mutation. This mutation most likely results in skipping of exon 8 and an out-of frame mRNA (*69*). Interestingly the described B3GALTL gene was the same as the one found by Kozma et al or Sato et al (*65, 67*) being the β1,3-Glucosyltransferase (β3Glc-T) that modifies *O*-fucosylated TSRs. In order to study the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr in these patients properdin was purified and its glycan composition analyzed. Properdin is the positive regulator of complement, it contains six TSRs. In four TSRs the Glcβ1,3Fuc-*O*-Ser/Thr disaccharide is present and its position is known (*61, 70*). The results demonstrated that in all patients examined the Glcβ1,3Fuc-*O*-Ser/Thr disaccharide was missing and the TSRs were demonstrated to contain solely *O*-Fuc (*71*). These results validated Peters-Plus syndrome as a novel CDG and emphasize the important role of the Glcβ1,3Fuc-*O*-Ser/Thr disaccharide for development.

1.3.3. *O*-GlcNAcylation

O-GlcNAcylation involves a GlcNAc moiety β-linked to serines or threonines. This modification is one of the most common posttranslational modifications and is conserved in all metazoan studies so far. In contrast to *N*-glcyans or other *O*-glycans, *O*-GlcNAcylated proteins are not further extended and can be found in cytosolic, as well as in nuclear proteins. *O*-GlcNAc is transferred by *O*-GlcNAc transferase (OGT) using UDP-GlcNAc as a sugar donor. Targeted deletion of this X-linked gene results in loss of embryonic stem cell viability (*72*). The *O*-GlcNAc can be removed by an enzyme called *O*-GlcNAcase, which exhibits a similar phenotype in *C. elegans* than deletion of OGT. These animals are viable but show defects in dauer formation and macronutrient storage in a daf-2 knock-out background (i.e. insulin-like receptor in *C*. elegans) (*73*). *O*-GlcNAc proteins are present in almost every cellular compartment. So far, no consensus sequence could be defined. However, a preference was observed for PVS (Pro-Val-Ser) or for PEST (Pro-Glu-Ser-Thr) sequences. *O*-GlcNAcylation is observed on residues, which are also found to be phosphorylated. This adds another factor of complexity and

diversity to signaling via phosphorylation. Proteins can not just be phosphorylated by a kinase and de-phosphorylated by a phosphatase but also GlcNAcylated or de-GlcNAcylated. The tight balance between phosphorylation on one hand and *O*-GlcNAcylation on the other hand renders the system susceptible for misregulations which manifest in diseases (*74*). Insulin resistance is the hallmark of type 2 diabetes, which is accompanied by increased levels of *O*-GlcNAcylation. It was shown that even moderate overexpression of OGT in muscle and fat tissues lead to insulin resistance, suggesting that increased transfer of *O*-GlcNAc onto proteins impairs directly glucose metabolism. While *O*-GlcNAc levels are elevated in diabetes, in Alzheimer's disease these levels are drastically reduced. This neurodegenerative disease is characterized by neurofibrillary tangles that result from hyperphosphorylation of the protein tau. The microtubule-associated protein tau is also *O*-GlcNAcylated in healthy individuals and a negative regulation of various phosphate sites has been observed. *O*-GlcNAcylation was suggested to compete with phosphates for the same positions. In Alzheimer's disease the glucose uptake into neurons is impaired, resulting in decreased levels of *O*-GlcNAc which could result in hyperphosphorylation of tau (*75*).

1.4 *C*-glycosylation

C-glycosylation involves α-mannose *C*-linked to the C^2 atom of the indol ring of tryptophan (*76, 77*) (see Figure 4). *C*-mannosylation was first reported in human RNase2 (*76*) but in the meantime was demonstrated to be present in a variety of different proteins (see Table 2). Similarly, to *O*-hexoses, *C*-hexoses add 162 Da to the mass of a peptide. They can be compared in tandem MS experiments, where *C*-glycosides exhibit a characteristic loss of 120 Da (formally four times CH_2=O) which is often accompanied by multiple water losses (*78, 79*). *O*-glycans on the other hand show a characteristic neutral loss of 162 Da. The recognition sequence of this modification was determined to be WXXW in a chimeric RNAse designated RNase2.4 (*80*). Later it was found in TSR proteins of the human complement system (C6-C9) containing the motif WXXWXXW, that one, two or three tryptophans can be *C*-mannosylated. In these proteins, *C*-mannosylation of tryptophan was even reported when no tryptophan or

aromatic amino acid at position +3 (relative to the glycosylated tryptophan) was present (*81*). Therefore, it was suggested that a secondary *C*-mannosylation signal may exist which is directed by the primary or tertiary structure (*81, 82*). The modification was demonstrated to be an enzyme-catalyzed process in the ER, using Dol-P-Man as a sugar precursor (*83*). The observation that the WXXW motif is not strictly conserved in all proteins, suggests the existence of several *C*-mannosyltransferases. These protein *C*-mannosyltransferases (PCMT) have not been cloned and only the activity or the discoveries of C-mannosylated proteins indicate their existence. Enzymatic activity of *C*-mannosylation was found from nematodes to humans suggesting a high conservation of this modification. In mammalian tissues a *C*-mannosyltransferase activity was detected in most of the organs (*83*) indicating an important function. Studies with native RNase produced in *E.coli* revealed that only proteins that have not acquired their three-dimensional structure can be efficiently *C*-mannosylated (*80, 83*). Similar to *N*-glycosylation, *C*-mannosylation must consequently take place as an early event on the nascent polypeptide chain before it has adopted its final tertiary structure (*84*).

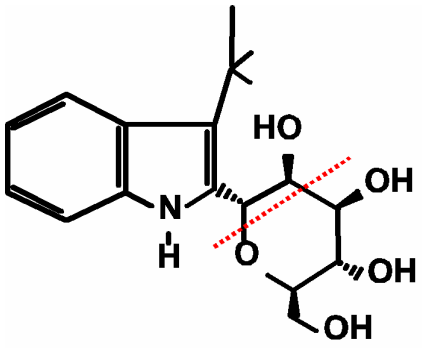


Figure 4: Structure of C-mannosyl tryptophan

The Mannosylpyranose is linked via its anomeric carbon to the C2 of the indol ring of tryptophan. The dashed line indicates the loss of 120 Da, which is characteristic for *C*-glycosides

Protein	Reference	No. of C^2-Man -Trp	Remarks
RNAse 2	(*76*)	1	
Hypertrehalosemic hormone	(*85*)	1	First insect hormone described to contain *C*-Man-Trp; no consensus sequence
Interleukin-12	(*83*)	1	
C6	(*81*)	6	*C*-mannosylation detected on more than one tryptophan. *C*-mannosylation detected in absence of consensus sequence
C7		4	
C8α		4	
C8β		4	
C9		2	
Properdin	(*70*)	15	
thrombospondin 1	(*39, 86*)	4	Ihara et al, used a *C*-Man-Trp specific antibody
F-spondin	(*61*)	8	
soluble Erythropoietin receptor	(*87*)	1	
MUC5B/Cys3	(*84*)	1	*C*-mannose detected with mannose binding lectins (GNA and LCA) [a]
MUC5A/Cys5		1	
bovine lense MP20	(*88*)	1	Neutral losses of 162 and 120 Da; no consensus sequence [b]
Ebola virus soluble glycoprotein sGP	(*89*)	1	

[a] The mannose binding lectins ConA and MBL do not recognize *C*-Man-Trp (*90*).

[b] In the insect Aedes aegypt N-mannosyl-tryptophan was described that exhibited a clear 162 Da loss (*91*). Since a 162 Da loss has never been observed for C2-linked α-mannopyranose, the linkage of this protein needs to be carefully evaluated.

Table 2: Proteins described to be *C*-mannosylated

C-mannosylated proteins reported in the literature. Proteins, where the carbohydrate has not been detected by mass spectrometry or novel observations/conclusions, are indicated

1.4.1 Proposed functions

Cytokine receptors contain in their extracellular domain a so-called WSXWS motif, which was shown to be critical for surface expression of the erythropoietin receptor (EPOR) (*92*). This motif is also the recognition motif for *C*-mannosylation. Therefore, the modification was studied in a soluble variant of the receptor. The soluble EPOR was found to be *C*-mannosylated on the first tryptophan to a degree of about 50%. Similar levels of modification where observed in a mutant (A234E) which promotes secretion

and in a mutant (S233A) which totally blocks secretion. Both mutants had a degree of modification inside and outside the cell comparable to wild type. Based on these results, it was concluded that *C*-mannosylation is not necessary for the secretion of the soluble EPOR but rather fulfills a stabilizing function (*87*). The results are in contrast to a study using the cystein-rich subdomains Cys3 and Cys5 of the mucin proteins MUC5B and MUC5A. These domains contain a WXXW motif close to their N-terminus. *C*-mannosylation was proven indirectly by binding to the mannose-specific lectins (GNA and LCA) and by competing binding with mannose. Mutagenesis of the WXXW motif resulted in decreased expression of the cysteins subdomains in COS 7 cells. This secretion deficiency was attributed to the missing *C*-mannosyl moieties. The authors confirmed this hypothesis by expressing the cysteins subunits in CHOlec35.1 cells. CHOlec35.1 are deficient in Dol-P-Man utilization and the gene mutated in these cells is necessary for all modifications requiring Dol-P-Man (*N*-glycans, GPI anchored proteins, *O*-mannosylation, *C*-mannosylation) (*93*). It was shown that C-mannosylation is significantly decreased in these cells (*93*). Expression of mucin cysteins subdomains in CHOlec35.1 cells resulted in proteins, which are retained in the ER. These effects were attributed to the missing *C*-mannosylation suggested to play a critical role for the secretion efficiency of mucin cystein-rich subdomains (*84*).

A connection between diabetes and *C*-mannosylation was suggested in macrophage RAW264.7 cells and in the aortic vessels of diabetic rats. Using a specific *C*-Man-Trp antibody, the authors could show that *C*-mannosylation was increased in secreted proteins in RAW264.7 cells cultured under hyperglycemic conditions. In a rat diabetes model, the level of *C*-Man-containing proteins was increased in the aortic vessels, other organs were not affected. In these tissues, an elevated *C*-mannosylated thrombospondin 1 was reported. It was suggested that *C*-mannosylation could play a role in the etiology of diabetes (*86*).

Using again RAW264.7 macrophage cells the same group demonstrated that chemically synthesized peptides containing *C*-mannosyl tryptophan potentiated the damage induced by the bacterial outer membrane component lipopolysaccharide (LPS). This effect was mediated through activation of *C*-Jun activated kinase (JNK) and resulted in increasing production of the inflammatory cytokine TNF-α. At the moment, it is not clear,

why these peptides mediated this effect and if degradation products of specific *C*-mannosylated proteins also elicit such an inflammatory effect.

In viruses, *C*-mannosylation was found in the glycoprotein sGP from the *Zaire Ebola virus*. The modification was observed by the characteristic loss of 120 Da in tandem MS on a peptide containing the WXXW repeat. Mutation of the WXXW motif to WXXA abrogated *C*-mannosylation. In contrast to the inflammatory response triggered by LPS and *C*-mannosylated peptides no change in function, biosynthesis or structure of the mutated sGP was observed. Therefore, it was suggested that the modification could be a consequence from an anomaly resulting from unedited transcripts. The WXXW motif and *C*-mannosylation are widespread in viruses. Database searches against the WXXW motif considering passage through the secretory pathway of the host revealed 373 proteins. These numbers and the fact that some viruses even contain multiple conserved *C*-mannosylation consensus sequences indicate that this modification might play a significant role in some processes of viral infections (*89*).

1.5 Rare glycosylations

Carbohydrates attached to serines or threonines encompass a huge field in protein glycosylation. In addition to the glycosylation described above, the last decade has brought major insights in the involvements of glycans at multiple levels in the cell. Especially with the help of transgenes and modern analysis tools like mass spectrometry the role of glycans in development or in diseases could be elucidated. These results position the glycobiology more and more in the field of system biology, where glycans are involved in various functions in-and outside the cell. In this regard, also less-well understood “rare” glycosylations become the focus of investigation. Some of them will be discussed in this chapter.

In addition to *O*-fucosylation EGF repeats also contain *O*-glucose in the recognition sequence C^1XSXPC^2 (the modified Ser is underlined). This modification has been found in Notch and in EGF repeats of the blood coagulation factors VII and IX. The glycan can be extended to the structure Xyl-α1,3-Xyl-α1,3-Glcβ1-*O*-Ser. Activities of the UDP:glucose protein *O*-glucosyltransferase (Poglut) and of the enzymes adding both

xyloses have been observed in CHO cells. Recently the Drosophila protein Rumi was described as the enzyme that adds *O*-glucose to EGF repeats. Interestingly, it was observed that Rumi mutants phenocopy Notch in a temperature dependent fashion ,providing an additional mode of regulation of Notch signaling (*94*).

A very important glycosylation is mammalian *O*-mannosylation. In yeast, *O*-mannosylation is essential for viability. In mammals, *O*-mannose has been described to occur in brain, eye and skeletal muscle. The best studied protein is α-dystroglycan having the tetrasaccharide Sia-α2,3-Gal-β1,4-GlcNAc-β1,2-Man-*O*-Ser/Thr. Similarly, to *C*-mannosylation, O-mannosylation requires Dol-P-Man as a sugar precursor. The mannose moiety is transferred by a hetero-complex of protein O-mannosyltransferase 1 and 2 (POMT1, POMT2). Mutations in POMT1 have been associated with Walter-Warburg syndrome (WWS). This disease is characterized by severe brain malfunctions and eye abnormalities and patients usually die within the first year. The mannose can be extended by protein *O*-mannose β1,2-*N*-acetylglucosamyltransferase 1 (POMGnT1) transferring a GlcNAc in β1,2-linkage to the mannose. Loss-of-function mutations of POMGnT1 in humans result in muscle eye brain disease (MEB), which is characterized by similar but less severe phenotypes than WWS. Both diseases and probably other muscular dystrophies like Fukuyama muscular dystrophy or congenital muscular dystrophy, are associated with under-glycosylation of α-dystroglycan. These under-glycosylations inhibit engagement with extracellular matrix components like laminin, agrin or neurexin and account for the various features associated with disturbed muscle development (*7, 95*).

2 Aim of the Thesis

The study of glycosylation is becoming increasingly important for understanding processes within and outside the cell. For example, glycans play pivotal roles in signaling as powerfully illustrated by the modulation of Notch signaling by glycosylation through the GlcNAc transferase fringe. Furthermore, many mutations in enzymes that carry out protein glycosylation cause severe, often systemic disorders (e.g. Peters-Plus-syndrome). Today, the major tasks in glycobiology are (i) to define the exact molecular role of glycans, (ii) to characterize the factors how a particular protein becomes modified and which glycans are involved and (iii) to obtain a complete list of glycan structures. In this thesis, I am addressing aspects of the first two questions.

In the first part, I am extending preliminary observations that a change in structure of a well-studied protein module, the thrombospondin type 1 repeat (TSR), causes an alteration of its pattern of glycosylation. This task was performed using the following approaches:
(i) Characterization of the acquired glycosylation by mass spectrometry
(ii) Biochemical characterization of the mutant
(iii) Identification of glycosyltransferases involved in the glycosylation.

In the second part, I aim to produce the tools to address, in molecular detail, the role of normal glycosylation of TSRs, i.e. C-mannosylation of Trp and modification with Glcβ1,3Fuc-*O*-Ser/Thr. Therefore, the following methods were applied:

i) Expression of TSRs in high amounts in organisms performing *C*-mannosylation and *O*-fucosylation
ii) Expression of TSRs in organisms which do not glycosylate TSRs
iii) Analysis of the expressed proteins and discussion of their suitability as model proteins in functional assays.

3 Material and Methods

In this chapter only those materials and methods are described which have not been included in the manuscript in Chapter 4.2.

3.1 Materials

Vectors: The vectors IRES-GFP-pRS5a (a kind gift from S. Geisse Novartis) pSecTagB–TSR4, pSecTagB–TSR1-4, pSecTagB-F-spondin and pPICZα (Invitrogen ™ Switzerland) were applied. Buffers: Acetonitrile and HPLC-grade water were from JT Baker (Griesheim, Germany), Trifluoro acetic acid (TFA) was from Pierce (Rockford IL, USA) and formic acid, methanol, ethanol were of ultrahigh purity and were purchased from Fluka (Buchs, Switzerland). Oligonucleotides: Primers used for PCR experiments were purchased from Microsynth (Balgach, Switzerland). Cell lines: HEK-EBNA (ATCC CRL-10852), HEK 293T (ATCC CRL-11268), COS 7 (ATCC CRL-1651), CHO-K1 (ATCC CCl-61). Yeast: *Pichia pastoris* (ATCC 28485).

3.2 Transformation of plasmids

The vectors were propagated in DH5α (Invitrogen ™) by adding 100 ng to 50 µl competent DH5α. The bacteria were incubated 20 min on ice, 50 s at 42°C and 2 min on ice. 950 µl 2xYP medium was added and incubated at 37°C for 2h. The bacterial solution was centrifuged at 1000xg for 5 min, 900 µl medium were discarded and the bacterial pellet was resuspended in the remaining 100 µl. The transformed bacteria were grown on LB (5 g/l yeast extract, 10 g/l tryptone, 10 g/l NaCl pH 7.5, 15 g/l agarose) plates and incubated over night at 37°C. One colony was picked, incubated for 8h in 4ml LB +50 mg/ml Ampicillin and 500 µl were used to incoluclate 250 ml LB +50mg/ml Ampicillin. On the next day, plasmids were purified according to the Macherey Nagel ™ Maxi Prep protocol.

3.3. Cloning of TSRs into mammalian and yeast expression vectors

TSR4 and TSR1-4 were amplified by PCR from pSecTag B –TSR4 and pSecTag B – TSR1-4 using the primers ckTSRIgK and ckTSR(4)3rev. This PCR created IgK-TSR4-TEV and IgK-TSR1-4-TEV having a SalI restriction site at the 5'-end. Subsequently the Fc tag was amplified from the pRS5a-IgG using the primers ckTSR(4)4 and ckTSR(1-4)6rev resulting in TEV-Fc-His with a BglII site at the 3'-end. The PCR products were separated by 1% agarose gel electrophoresis in 1x TAE buffer (4.84 g Tris base, 1.142 ml acetic acid 2 ml 0.5M EDTA) and purified according to the manufacturer's instructions(Gel extraction kit Qiagen, Hilden, Germany). A third PCR amplification was performed using both fragments and the primers ckTSRIgK and ckTSR(1-4)6rev.
This resulted in amplicons encoding IgK-TSR4-TEV- Fc-His (1117 bp) and IgK-TSR1-4-TEV- Fc-His (1551 bp) which will be called TSR4fchis and TSR1-4fchis. For expression in yeast TSR4fchis and TSR1-4fchis were amplified using the primers ckTSRpischia3 and ckTSRpischia4rev resulting in an amplicon of 1145 and 1487 bp respectively. The PCR products and plasmids were separated on a 1% agarose gel, extracted and digested over night at 37°C using the enzymes SalI and BglII (Roche™) for the mammalian vector and XhoI and NotI for the yeast vector. In parallel 2 µg of the mammalian target vector IRES-GFP-pRS5a and the yeast *Pichia pastoris* expression vector pPICZα were digested with the same enzymes. IRES-GFPpRS5a is a bicistronic vector containing a GFP, which is under the same promoter control than the target protein and can therefore be used to monitor expression. After a second gel purification to remove restriction enzymes, the vectors were de-phosphorylated using alkaline phosphatase according to the manufacturer's instructions (Roche™). The PCR products were ligated using T4 ligase kit as described (Roche™) and 5 µl were pipetted to 50 µl of chemical competent DH5α (Novagen™) bacteria. Transformation of competent bacteria was performed as described before. 10-30 colonies were inoculated in 4 ml LB+Amp/colony. Plasmids were extracted according to Qiagen MiniPrep kit and 5 µl eluate was digested over night with the corresponding restriction enzymes. After separation on an 1% agarose gel, clones were screened for the correct insert and positive ones were subjected to DNA sequencing.

5' primer	**3' primer**	
Name	*Name*	**PCR conditions**
Sequence	Sequence	
ckTSRIgK	*ckTSR(4)3rev*	- 94°C 2min
ACGCGTCGACAGCCAC	AGATTTGCCCTGAAAAT	- (94°C 1min, 50°C 1min, 72°C 1.5min) x 5 cycles
CATGGAGACAGACACAC	ACAAATTCTCGGGGCAC	- (94°C 1min, 55°C 1min, 72°C 1.5min) x 25 cycles
TCCTGC	TCTGGCAGCATACA	- 72°C 5min
		- 4°C
ckTSRIgK	*ckTSR(1-4)6rev*	- 94°C 2min
ACGCGTCGACAGCCAC	GAAGATCTTCTCAGTGG	- (94°C 1min, 50°C 1min, 72°C 1.5min) x 5 cycles
CATGGAGACAGACACAC	TGGTGGTGGTGGTGTTT	- (94°C 1min, 55°C 1min, 72°C 1.5min) x 25 cycles
TCCTGC	ACCCGGAGACAGGGA	- 72°C 5min
		- 4°C
ckTSRpischia3	*ckTSRpischia4rev*	- 94°C 2min
CCGCTCGAGAAAAGAGA	TTTTCCTTTTGCGGCCG	- (94°C 1min, 55°C 1min, 72°C 1.5min) x 5 cycles
GGCTGAAGCTGCGGCC	CTCAGTGGTGGTGGTG	- (94°C 1min, 65°C 1min, 72°C 1.5min) x 25 cycles
CAGCCGGCCAGG	GTGGTGTTTACC	- 72°C 5min
		- 4°C

Table 3: Primer and PCR conditions

3.4 Expression systems

3.4.1 The mammalian expression system HEK-EBNA

HEK-EBNA cells are derived from the human embryonic kidney cell line HEK 293 containing a stably integrated Epstein-Barr Nuclear Antigen (EBNA). EBNA encodes for a protein that drives episomal replication of plasmids carrying the oriP derived from the Epstein - Barr virus (*96*). These cells were successfully used for large-scale protein expression (*97*). It was demonstrated that HEK-EBNA cells are positive for *C*-mannosylation (*82*). The presence of POFUT2 and β1,3GlcT in HEK-EBNA to glycosylate the disaccharide Glc-Fuc-*O*-Ser/Thr on TSRs has not been demonstrated

directly in HEK-EBNA cells but in the related human embryonic kidney cell variant HEK 293T (*65*).

3.4.2. The methylotrophic expression system *Pichia pastoris*

The methylotrophic yeast *P. pastoris* is suitable for high extracellular expression of recombinant proteins. It is able to form correct disulfide bridges even on difficult proteins. *N*-linked as well as O-mannosylated *O*-linked glycans have been described (*98*). Neither C-mannosylation nor O-fucosylation has been reported.

3.5 Cell culture

HEK-EBNA, HEK 293T and CHO-K1 cells were routinely cultivated in a humidified incubator at 5% CO_2. HEK EBNA and HEK 293T were cultured in DMEM + 10%FCS Penicillin, Streptomycin and 2.5 mM glutamine. CHO-K1 cells were cultured in αMEM 10%FCS Penicillin, Streptomycin and 2.5 mM glutamine

3.6 Transfection

HEK-EBNA, COS7 and CHO-K1 cells were transfected with Lipofectamine (Invitrogen ™) according to the manufacturer's instructions.

3.7 Cell sorting of stable HEK-EBNA cell

Two days after transfection 100 µg/ml Zeocin (Invitrogen™) was added and incubated for 2 weeks. Culture medium was replaced every 3 days. Subsequently two 10 cm dishes containing stable transfected cells expressing TSR4fchis and TSR1-4fchis were subjected to cell sorting based on the GFP fluorescence using a MoFlo (DakoCytomation™). Subclones were subsequently expanded and screened for high expression of the target protein.

3.8 Transformation and expression of recombinant proteins in *Pichia Pastoris*

The vector pPICZα containing TSR4fchis and TSR1-4fchis was linearized with SacI over night. *Pichia pastoris* was made competent using a modified Lithium acetate method (A. Hein, Novartis). A preculture of 5 ml YPD (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose pH 6.5) was inoculated with one colony *P. pastoris* and grown over night at 37°C. On the next day 100 ml YPD were inoculated with 20 µl o/n culture and grown to a OD_{600}=1-2. The cells were pelleted by centrifugation at 500 xg for 5min) and $1x10^{10}$ cells were incubated in 96 ml LiAc-DTT buffer (100 mM lithium acetate, 10 mM DTT, 0.6 M sorbitol, 10 mM Tris pH 7.5) for 30 min. Subsequently cells were pelleted down for 10 min at 1500xg and resuspended in 25ml ice cold 1M sorbitol. This step was repeated twice, then the cells were pelleted and resuspended in 960 µl 1 M sorbitol. Linearized DNA was mixed with 8 x 10^8 cells and pipetted into a cold electroporation cuvette, incubated for 5 min and then pulsed with 1500V, 25 µF, 200 Ohm in a BioRad Micropulser. The transformed cells were then diluted in 1ml ice cold 1 M sorbitol, incubated for 1 h at 30°C.Then 1 ml YPD medium was added, incubated for 1h at 30°C at 225 rpm and spread onto YPSD-Zeocin plates (2% peptone, 1% yeast extract, 2% dextrose, 2% agar, and 1 M sorbitol)

3.9 Expression of recombinant TSRs in *P. pastoris*

A 5 ml preculture of BMGY (1% yeast extract, 2% peptone, 100 mM potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 0.4 mg/L biotin, 1% glycerol) +100 µg/ml Zeocin was inoculated with a Zeocin resistant *P. pastoris* clone and grown over night at 30°C and 225 rpm. Then 50 ml BMGY + Zeocin was inoculated with the overnight culture and grown at the same conditions. The next day the cells were pelleted and resupended in 50 ml BMMY (same as BMGY but with 0.5% methanol instead of glycerol) in a 250 ml baffled Erlenmeyer flask at a OD_{600}=10 and grown for 48 h. 0.5% Methanol was added twice a day to a concentration of 0.5% (v/v).

3.10 Purification of recombinant proteins using IMAC

F-spondin or TSRs containing a His_6-tag were transfected into HEK 293T using Lipofectamine. 24 h after transfection cells were cultured with Optimen (Gibco ™) without serum or antibiotics fro 48 h. The supernatant was spun and cells or cell debris removed. Subsequently the supernatant was dialyzed against a 100 x excess of 50 mM Tris pH 8 100 mM NaCl. The buffer was replaced twice. The next day the dialysate was incubated with $^1/_{100}$ of the total volume of NiNTA beads over night at 4°C. Beads were collected by centrifugation and washed with at least 20 column volumes of washing buffer (50 mM Tris, pH 8, 500 mM NaCl, 20 mM imidazole). Elution was performed with two bead volumes of elution buffer (50 mM Tris, pH 8, 150 mM NaCl, 500 mM imidazole) for two-times five minutes. The protein was reduced, carboxyamidomethylated and digested as described in 3.14.

3.11 Purification of recombinant TSR-Fc fusion proteins

Supernatant from stable clones expressing TSR4fchis and TSR1-4fchis was collected and 500 ml to 1 L were loaded at a flow of 0.5 ml/min on a commercial 1ml NiNTA column (GE Healthcare™). Proteins were washed with buffer A (20 mM NaPi pH 7.4, 500 mM NaCl, 20 mM Imidazol) at 1 ml/min for 10 min. Elution was performed by switching to 100% buffer B (20 mM NaPi pH 7.4, 500 mM NaCl, 500 mM Imidazol) for 5 min. Fractions of 0.5 ml were collected and protein concentration was measured manually at OD_{280}. Fractions containing high protein amounts were pooled, dialyzed over night at 4°C against PBS and incubated over night at 4°C with protein G beads (GE Healthcare™) with $^1/_{100}$ of the total fraction volume. Proteins were washed in PBS containing 0.5 M NaCl and eluted using a pH shift with buffer containing 0.2 M Glycine pH 3, 150 mM NaCl. In order to prevent precipitation eluted proteins were immediately neutralized using a $^1/_{33}$ of the total volume of 1 M Tris pH 9. Protein purity was examined by SDS-PAGE.

Supernatant of TSRs expressed in yeast was cleared by centrifugation at 1000 xg for 10 min and incubated with $^1/_{100}$ of the total volume of protein G beads. Washing and elution was performed as described above.

TSRs (5 µg / protein) expressed in HEK-EBNA cells were additionally analyzed by HPLC (Agilent 1100) on a 1 mm C4 column (Vydac™). The system was operated in the microflow mode at a flow rate of 50µl/min. The HPLC was connected to a Triple Quadrupole LC/MS/MS Mass Spectrometer (API 300 10x, Perkin Elmer Sciex-Instruments, Toronto Canada). A gradient from 5-80% buffer B in 60 min (Buffer A: 2% ACN, 0.05% TFA, Buffer B: 80% ACN, 0.045% TFA) was used.

3.12 Sandwich ELISA of proteins having a Myc-His-tag

Quantification of His-tagged proteins was performed with commercially available Ni-NTA HisSorb Plates (Qiagen, Hilden, Germany). As a standard purified TSR4 was used at a concentration of 0.25-0.005 ng/µl. TSR4 was prepared as previously described (*65*). Protein concentration was measured by absorption at 280 nm and the protein diluted in 0.2% BSA in PBS + 0.05% Tween 20. Subsequently the samples were diluted 1:1000 to 1:2000 in the same buffer and pipetted into the wells. Samples and standards were always applied in triplicate at a volume of 200 µl and incubated for two hours at room temperature or at 4°C over night. The wells were washed four times with PBS + 0.05% Tween and subsequently anti-Myc antibody (Sigma) at a dilution of 1:5000 in 0.2% BSA in PBS +0.05% Tween was added into the wells. After two hours incubation at room temperature, wells were washed four times with 0.2% BSA in PBS + 0.05% Tween and the secondary anti-mouse antibody was added at a dilution of 1:2500 in 0.2% BSA in PBS +0.05% Tween. The plate was incubated for one hour at room temperature and washed four times. SIGMA*FAST*™ OPD (*o*-Phenylenediamine dihydrochloride) (Sigma) was used for detection according to the manufacturer's recommendation. The reaction was stopped after 30 min incubation in the dark with 50 µl M H_2SO_4 and absorbance was read at 490 nm on a Spectra MAX Plus spectrophotometer (Bucher biotech, Basel, Switzerland). Figure 5 shows a representative diagram of the standards measured in the ELISA that were used for quantification.

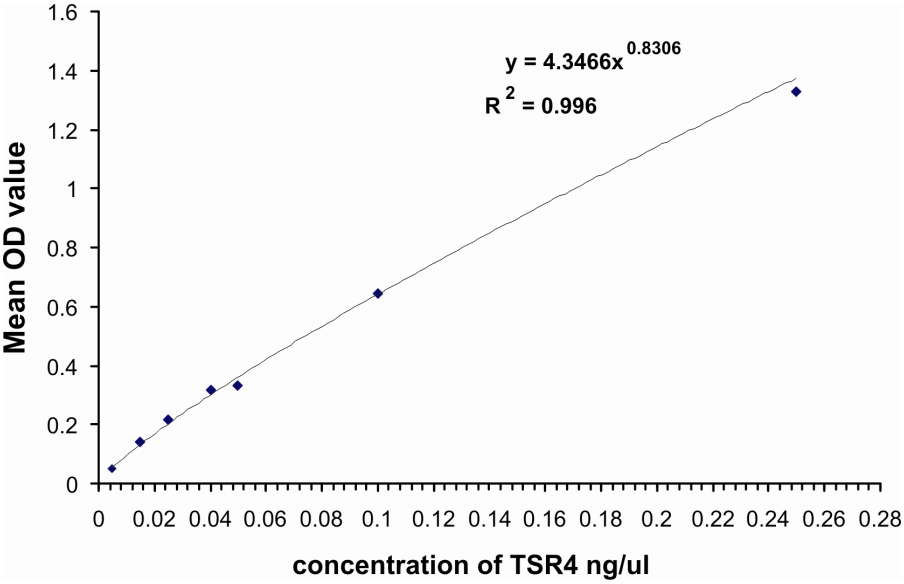


Figure 5: ELISA of TSR4 standards

3.13 Analytical gel filtration

Gel filtration is used to determine molecular weights of unknown samples and to estimate the Stokes Radius (Rs) of an unknown molecule. The Stokes Radius is the radius of a hard sphere that diffuses at the same rate as the protein to be measured. It defines a value that reflects the compactness and molecular size of globular proteins, assuming that a long extended molecule has a greater Stokes radius than a compact molecule of the same molecular mass. In practice, a gel filtration column is calibrated with known molecular weight standards. For calibration, the following parameters are necessary to define a column: V_t represents the total volume of the column. Large proteins, which are not retained by the column, elute in the void volume V_0. The elution volume of the sample is given by the parameter V_e. The partition coefficient K_{av} is defined as:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad \text{(Equation 1)}$$

Using a plot K_{av} versus log molecular weight (MW) of standards, the molecular weight of an unknown sample can be determined. The retention coefficient R is given by the division of V_0/V_e. A plot $\sqrt{}$–log K_{av} versus known Stokes radii of the standard proteins can

be used to estimate the Stokes radius of the sample (*99*). For gel filtration experiments, a Superdex 75 100/300 GL connected to an Aekta Prime (GE Healthcare) chromatography system was used. The column had a total volume of 24 ml and an experimentally determined void volume of 7.65 ml. Bovine serum albumin (BSA) dimer (134 kDa), BSA monomer (67 kDa, Rs=35.5 Å), Ovalbumin (43 kDa, Rs=27.3 Å), RNAse A (13.7 kDa, Rs=16.4 Å) and Aprotinin (6.5 kDa, Rs=10.8 Å) were used to calibrate the system. Stokes radii were taken from the literature (*99, 100*). For each standard 50 µg of protein was loaded. The running buffer was 20 mM NaPI pH 7.4, 150 mM NaCl. The flow rate was set to 0.2 ml/min. Chromatography was performed until the salt peak, indicated by an increase in the conductivity was detected. Figure 6A shows the plot K_{av} versus log molecular weight and figure 6B $\sqrt{-\log K_{av}}$ versus the Stokes radius (for this plot only BSA monomer was considered).

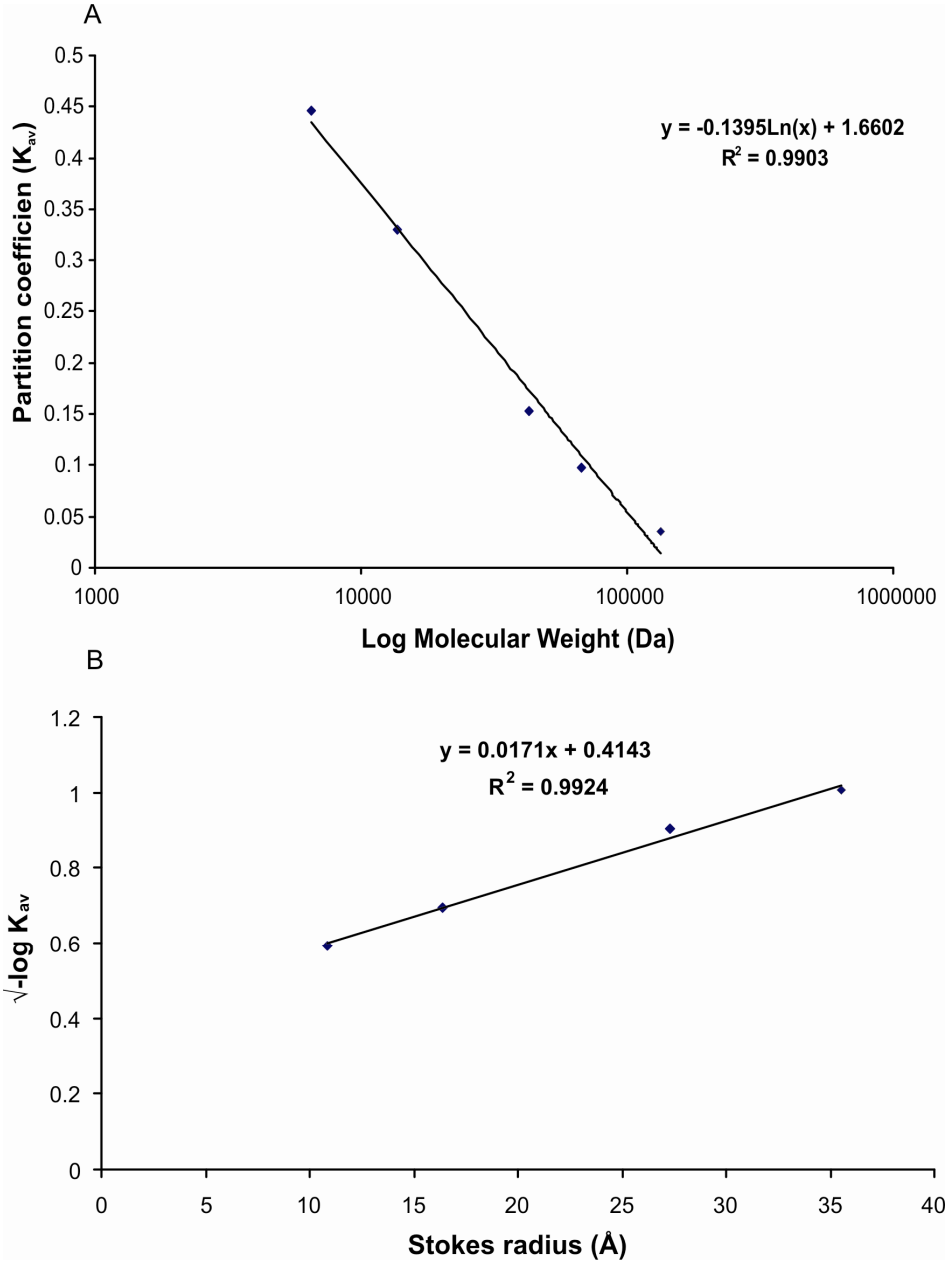


Figure 6: Calibration of Superdex 75 100/300 GL gel filtration column:

(A) The partition coefficient K_{av} was blotted against the log molecular weight of K_{av} of known standards. (B) Plot of √–log K_{av} against known Stokes radii of the standard proteins.

Samples of native TSR4wt and ΔDS2,3-TSR4 were chromatographed under the same conditions as standards. Denatured TSRs were generated by treatment with TCEP at a concentration of 5 mM for 1 h at 60°C, reduced cysteines were carboxyamidomethylated with 85 mM iodacetamide for 30 min in the dark. Fractions of 0.2 ml were collected and analyzed by TSR4-specific ELISA as described.

3.14. Reduction, carboxyamidomethylation and digestion

Protein samples (3-10 µg) were dialyzed over night at 4°C against H_2O or 50 mM NH_4HCO_3 and subsequently dried in the speed vac. Proteins were dissolved in 6 µl argon-purged RCM (500 mM Tris pH 8.6, 6 M guanidium-HCl) + DTT and incubated at 37°C for 1-3 h. Cysteines were carboxyamidomethylated using 1 µl iodacetamide in the dark at a final concentration of 115 µg/µl for 30 min at room temperature. The reaction mix was diluted with 35 µl digestion buffer (500 mM NaPi pH 8, 700 mM guanidium-HCl) and digested over night at 37°C with 400 ng endoproteinase Lys-C.

3.15 LC-MS

Peptides were separated on a HPLC-MS (Agilent 1100 HPLC system connected to a triple quadrupole mass spectrometer IONICS, EP10+, a modified PE SCIEX API 360, Concord Canada) at a flow rate of 50 µl/min. In order to separate glycosylated TSRs a flat gradient of 0-50% B in buffer system 1 (A=2% ACN 0.05% TFA B=80% ACN, 0.045% TFA) in 120 min was used. The UV_{280} signal of C-mannosylated glycoforms was divided by the factor 1.3 for peptide with a single (C^2-Man)Trp and by 1.6 for those containing two (C^2-Man)Trp residues. For analysis of glycopeptides at high resolution, buffer system 2 (A=2% ACN 0.1%FA, B=80% ACN 0.1% FA) was used. Relevant Peptides were collected manually in Low-Bind tubes (Eppendorf ™), neutralized with an equal volume of 50 mM NH_4HCO_3 and dried in a Speed Vac.

3.16 SDS-PAGE and Western Blotting

SDS-PAGE was performed according to Laemmli (*101*) using 72 x 1.5 x 102mm minigels in a Protean II chamber (BioRAD). Gels were stained by Coomassie using the NuPAGE™ staining kit (Invitrogen). Alternatively, proteins were blotted onto nitrocellulose (BioRad) using 10 mM CAPS pH 11, 10%(v/v) methanol. Probing of the membrane with different antibodies was performed according to the manufacturer's instructions. HRP-labeled secondary antibodies were detected using the ECL detection kit, according to the manufacturer's instructions (GE Healthcare).

4 Mucin-type glycosylation of ΔDS2,3-TSR4

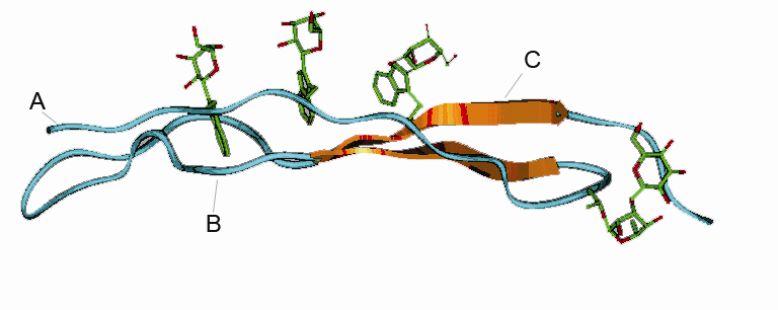
4.1 Introduction

The fourth thrombospondin type 1 repeat (TSR4) of the axonal guidance protein F-spondin has been demonstrated to contain C-mannosylation on the WXXW motif and the disaccharide Glcβ1,3-Fuc-O- Thr in the consensus sequence C^1X_{2-3}(S/T)C^2X_2G (*61*). In this chapter it will be shown, that a structural mutant of TSR4 lacking two out of three disulfide bridges (termed ΔDS2,3-TSR4), exhibits a third type of glycosylation. The glycan was determined to be NeuNAcα2,3Galβ1,3[NeuNAc]GalNAc-*O* linked to the same residue modified with the disaccharide Glcβ1,3-Fuc-O- in wild type TSR4. It will be shown that ΔDS2,3-TSR4 behaves as an independently folding protein and that mucin-type O-glycosylation on this module can be initiated *in vitro* and *in vivo* by GalNAc T3. In order to place these findings in a suitable background a short structural and functional overview on TSRs will be provided.

4.1.1 Thrombospondin Type 1 repeat as a model protein

Thrombospondin Type 1 repeats (TSR) are highly conserved molecules of about 60 amino acids found from protozoa to humans. BLASTp searches against non-redundant databases revealed that 187 TSRs in 41 different proteins occur in the human genome (*102, 103*). The crystal structure of TSR2 and TSR3 of thrombospondin 1 (TSP-1) revealed an anti-parallel three-stranded fold (strands A-C). Strand A has a rippled conformation whereas strand B and C form regular β-sheets. The conserved tryptophans in the WXXW motif interact in a cation-π –interaction with conserved arginines in the B strand. This structure is capped by disulfide bridges at the top and at the bottom forming a CWR-layered structure. A cysteine in the loop between strand A and B forms the third disulfide bridge with a cysteine in the C strand located at the C-terminus. Based on their disulfide connectivity TSRs have been grouped into two groups. Group 1 containing proteins like thrombospondin, properdin, ADAMTS-1 and 4 or TSRs from the complement factors C6, C7 C8α and C9β, have a connectivity of 1-5,

2-6 and 3-4. Group 2 has a connectivity of 1-4, 2-5 and 3-6 (cysteines have been numbered according to the appearance in the molecule as proposed by Huwiler et al (*104*)) and is found in TSRs of F-spondin, Mindin or the plasmodium protein TRAP (*105*). TSRs have been shown two contain two glycosylations, *C*-mannosylation of the Trp in the WXXW motif (*61, 66, 81*) and the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr occurring in the consensus sequence C^1X_{2-3}(S/T)C^2X_2G (*61*). It has been suggested that the *C*-mannosyl tryptophans protrude to the outside of the molecule on strand A. Glcβ1,3-Fuc-*O*-Ser/Thr is located in the loop between strand A and B (*105*) (Figure 7).



adapted from (*105*), modified

Figure 7: Crystal structure of TSR of TSP-1 with modeled glycosylation

The TSR contains an anti-parallel three-stranded fold consisting of by strand A with a rippled conformation and strands B and C with a regular β-sheet structure. The conserved tryptophans interact with conserved arginines in a cation-π –interaction capped by disulfide bridges. TSRs of thrombospondin 1 have three conserved Trp residues in a WXXWXXW motif, which can be modified with *C*-mannose. These mannose moieties have been suggested to protrude to the outside of the molecule. The second glycosylation, the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr, is located in the loop between strand A and B as illustrated here by modeling.

The function of TSRs is diverse and will be presented only exemplarily. The function of the TSRs in F-spondin will be mentioned in chapter 5.1.1. TSRs of TSP-1 and TSP-2 have been shown to bind the scavenger receptor CD36 through their CSVTCG-motif (*106*). This interaction was shown to be responsible for the anti-angiogenic response of thrombospondins (*107, 108*). TSRs in the transmembrane protein TRAP from the protozoa Plasmodium have been shown to be necessary for invasion of the parasite into the cells of the mosquito salivary glands (*109*). Finally, peptides of a TSR located in the metalloproteases ADAMTS-4 have been implicated to bind to the glycosoaminoglycans

of aggrecan, which was shown to be necessary for the aggrecanase activity of ADAMTS-4 (*110*).

4.2 Manuscript: Alternative glycosylation of Thrombospondin Type 1 repeats

Alternative Glycosylation of Thrombospondin Type 1 repeats*

Carsten Krantz‡,' Dominique Klein‡, Ragna Sack‡, Kentaro Kato§, Daniel Hess‡, Henrik Clausen§ and Jan Hofsteenge‡1

‡Friedrich-Miescher Institute for Biomedical Research, Basel Switzerland §Department of Medical Biochemistry and Genetics, Faculty of Health Sciences, University of Copenhagen

Address correspondence to: Jan Hofsteenge, Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058, Switzerland. Telephone: +41 61 6974722; FAX: +41 61 6973976; E-mail: jan.hofsteenge@fmi.ch

Running title: *Gain of O-glycosylation in a mutant of TSR4 of F-spondin*

SUMMARY

Thrombospondin type 1 repeats have been demonstrated to contain the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr that is synthesized by protein-*O*-fucosyltransferase 2 and β1,3-glucosyltransferase. Here we demonstrate that a mutant of the fourth Thrombospondin type 1 repeat (TSR) of f-spondin, lacking two disulfide bridges undergoes an alternative *O*-glycosylation. This mutant lacks *O*-fucosylation, but carries the disialylated core 1 modification (NeuNAcα2,3Galβ1,3[NeuNAcα2,6]GalNAc-*O*-) on the same residue that is otherwise *O*-fucosylated. In a peptide screen we identified ppGalNAc transferase 3 to be able to initiate this modification *in vitro*. This observation was confirmed by co-expressing ppGalNAc transferase 3 and the mutant TSR in CHO-K1 cells and analyzing the modification by relative quantification. To the best of our knowledge, this is the first example of a switch in glycosylation on one-and-the-same residue. The results suggest that mutations may exist, that cause a gain of mucin type *O*-glycosylation.

INTRODUCTION

Glycosylation is the most abundant form of post-translational modification in eukaryotes. Besides *N*-glycosylation, *O*-glycosylation is the second most frequent type (*111*). *O*-glycosylation of the mucin-type is initiated by the attachment of a α-GalNAc moiety to Ser or Thr, which is catalyzed by a member of the polypeptide GalNAc-transferase family. At the moment this family consists of 15 characterized proteins and at least five predicted ones (*26, 35*). Unsubstituted *O*-linked GalNAc is commonly known as the Tn antigen, which can be sialyated to yield sialyl-Tn (STn). Alternatively, extension of the Tn antigen by core I β1,3Gal-transferase results in the formation of the T-antigen (Galβ1,3GalNAc-*O*-Ser/Thr)[2]. Addition of sialic acid in α2,3-linkage results in the sialyl-T antigen (ST), which can be further modified to yield the disialyl-T-antigen (dST) (Siaα2,3Galβ1,3[Siaα2,6]GalNAc-*O*-Ser/Thr). Tn and T-antigens, as well as their sialylated counterparts have been associated with a variety of different cancers including breast (*112*), colon (*113*) and gastric cancer (*114*). A different type of *O*-glycosylation is *O*-fucosylation, the attachment of L-fucose to Ser or Thr. It has been found in a protease inhibitor from *Locusta migratoria* (*40*), in epidermal growth factor like repeats (EGF)[3] (*38*) and in thrombospondin type I repeats (TSRs) (*39, 61*). *O*-fucosylation of TSRs is performed by POFUT2, which similar to *O*-fucosylation of EGF repeats by POFUT1, requires a properly folded protein as the acceptor substrate (*41*). Importantly, POFUT1 and -2 are specific for their respective acceptor substrates (*43, 65*). In *C. elegans* *O*-fucosylation by POFUT2 is necessary for normal gonad development, since a knock-out of the nematode orthologue, PAD-2, results in abnormal distal tip cell migration (*115*). The fucose moiety on TSRs can be further

extended by an ER- resident β1,3-glucosyltransferase to form the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr (*65*). Mutations in the gene encoding this transferase were recently described to cause Peters-Plus-Syndrome (*69*) and to result in the lack of glucose in the TSRs of the reporter protein, properdin[4]. This autosomal recessive disorder is characterized by anterior eye-chamber defects, short stature, developmental delay and cleft lip and/or palate. Together, these observations emphasize the importance of this form of glycosylation of TSRs for normal development.

The three-dimensional structure of TSRs from thrombospondin and the axonal guidance protein f-spondin consists of a unique three-stranded antiparallel fold that is stabilized by disulfide bridges (*102, 116*). Recently, Haltiwanger and coworkers demonstrated that reduced and carbamidomethylated TSR is no longer a substrate for POFUT2 *in vitro* (*43*). We wanted to address the question whether correct disulfide bridge formation in a TSR is important for recognition by POFUT2 *in vivo*. Towards that aim, we expressed a mutant of the fourth TSR of rat f-spondin, which lacks its second and third disulfide bridge, in HEK293T cells. In contrast to the wild type TSR, this mutant did not undergo *O*-fucosylation on Thr-601. Instead, it carried a disialylated T-antigen on this amino acid residue. To the best of our knowledge, this is the first example of a complete switch of the type of *O*-glycosylation on one-and-the same amino acid position. In addition, we obtained *in vitro* and *in vivo* evidence that GalNAc-T3 can initiate the observed *O*-glycosylation of this TSR mutant.

EXPERIMENTAL PROCEDURES

Materials - Neuraminidase from *Clostridium perfrigens* (which cleaves α2-3-, α2-6-, and α2-8-linked N-acetylneuraminic acid, *O*-Glycosidase and endopeptidase Asp-N were purchased from Roche (Basel, Switzerland). α2-3-Neuraminidase from *Salmonella typhimurium* was obtained from Calbiochem (San Diego, CA., USA). Endoproteinase Lys-C was from Wako (Dallas, TX, USA). Mouse anti-Myc monoclonal antibodies and Extravidin-labelled peroxidase were purchased from Sigma (Buchs, Switzerland). Biotinylated peanut agglutinin (PNA) and rabbit anti-mouse-HRP were from Vector Laboratories (Burlingame, USA) and GE Healthcare (Wädenswil, Switzerland), respectively. Acetonitrile and HPLC-grade water were from JT Baker (Griesheim, Germany), TFA was from Pierce (Rockford IL, USA) and formic acid, methanol, and ethanol (ultrahigh purity) were purchased from Fluka (Buchs, Switzerland). Oligonucleotides used for PCR experiments were purchased from Microsynth (Balgach, Switzerland). *Plasmid for the expression of recombinant TSR4* - The fourth thrombospondin type I repeat of rat f-spondin (residues 615-666; Swiss-Prot P35446) was amplified from cDNA (a kind gift from Dr. A. Klar, Hebrew University, Jerusalem) using the following primers (forward: AAGCTTTGCTTGCTGTCTCCTTGGTCCG and reverse: TCTAGAGCCCTGAAAATACAAATTCTCG GGGCACTCTGGCAGCATA), introducing a HindIII- and XbaI site at the 5`- and 3` end, respectively. The reverse primer also introduced a tobacco-etch-virus cleavage site (TEV). This construct was cloned in-frame into pSecTagB (Invitrogen). This strategy resulted in 14 additional amino acids at the N-terminus and a TEV cleavage site, a Myc-epitope and a His_6 tag at the C-terminus (all in italics), resulting in the following amino acid sequence: *DAAQPARRARRTKL*CLLSPWSEWSDCSVTC GKGMRTRQRMLKSLAELGDCNEDLEQAE KCMLPECP*ENLYFQGSRGPEQKLISEEDLNS AVDHHHHHH*. This protein is indicated as wild type TSR4. The mutant TSR4 lacking the second and third disulfide bridges (designated ΔDS2,3-TSR4) was obtained by inverse PCR using *pfu* DNA polymerase (Promega, Wallisellen, Switzerland) with overlapping primers containing the desired Cys to Ala mutations (underlined in the sequence above) (*117*). The coding regions of all expression plasmids were sequenced to verify the presence of the desired mutation and the absence of unwanted ones. *Cell culture* - HEK293T cells were maintained in DMEM, supplemented with 10% FCS,

penicillin, streptomycin and 2.5mM glutamine in a humidified incubator at 37°C with 5% CO_2. CHO K1 cells were grown in MEMα (Gibco, Switzerland) supplemented with 10% FCS, 2.5mM glutamine, penicillin and streptomycin.

Expression and purification of recombinant TSR4 - Plasmids encoding wild type or mutant TSR4 were transiently transfected into HEK293T in Optimem (Gibco, Basel, Switzerland), using Lipofectamine (Invitrogen), according to the manufacturer's instructions. Conditioned medium was collected after 48h, and dialyzed overnight against 20 mM Tris-HCl pH 8.0, containing 150 mM NaCl. TSR4 from 50 ml of conditioned medium was extracted with 50 µl NiNTA resin (Qiagen) for 18h at 4°C. After washing the resin with 20 volumes of 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, TSR4 was eluted with 500 mM imidazole in the same buffer. The purity of the protein was verified by SDS-PAGE. The purified proteins were dialyzed over night against 50 mM NH_4HCO_3 or H_2O prior to further analysis.

Analysis of glycosylation status of TSR4 by peptide mapping – Samples of TSR4 was dried, carboxamidomethylated (*80*) and digested with Lys-C (8 % w/w) at 37°C overnight. The digests were fractionated by LC-MS , using an Agilent 1100 HPLC system that was interfaced with an upgraded (IONICS, EP10+) API 300. MS experiments were performed in the positive ion mode using an atmospheric pressure ionization source and a declustering potential of 40V. For initial separation of glycopeptides a reversed-phase C_{18} column (1 x 250 mm; Grace Vydac, Deerfield, USA) was used with a 120-min linear gradient of 0-50% B at a flow rate of 50 µl/min (solvent A: 0.05% TFA, 2% CH_3CN; solvent B: 0.045% TFA 80% CH_3CN). Relevant glycopeptides were collected manually, neutralized with an equal volume of 50 mM NH_4HCO_3, dried and digested overnight at 37°C with endopeptidase Asp-N in 50 mM NH_4HCO_3. Purification of the polar glycopeptides was performed on a Hypercarb column (1 x 150 mm, Thermo Electron Waltham, MA, USA) using a 35-min linear gradient of 0-70% B (solvent A: 0.1% formic acid, 0.01% TFA, 2% CH_3CN; solvent B: 0.1% formic acid, 0.01% TFA, 80% CH_3CN) at a flow rate of 50 µl/min. This procedure yielded peptide K2*D2.

Determination of the structure of the tetrasaccharide - Peptide K2*D2 was analyzed by CID tandem MS in the static nano-electrospray mode on a quadrupole linear 4000 Q TRAP instrument (Applied Biosystems, Foster City, CA, USA) as described (*118*). The collision energy was chosen so that sequence information on the *O*-glycan was generated. Further structural information was obtained by treating the glycopeptide with endoglycosidases. Digestion with neuraminidase from *Clostridium perfringens* (0.05 mU) was performed overnight in 25 µl of 50 mM ammoniumacetate, pH 4.5, at 37°C, in the presence of 1x complete protease inhibitor cocktail (Roche). Part of the product was dried, dissolved in 25 µl 100 mM Na-phosphate, pH 6.0 and treated with *O*-glycanase (0.5 mU) at 37°C overnight. Another aliquot of the peptide was treated overnight at 37°C with α2-3 neuraminidase from *Salmonella typhimurium* (100 units) in 25 µl 50 mM sodium citrate, 100 mM NaCl, pH 6.0. Reactions were terminated by dilution into 1% formic acid, containing 30% methanol, and the mass of the reaction products was determined by LC-MS using the Hypercarb system described above. In addition, the product of the digestion with the neuraminidase from *Clostridium perfringens* was analyzed by Western blotting using biotinylated PNA and antibodies against the Myc tag.

Determination of the modified amino acid residue – Glycosylated peptide K2*-D2 was examined by CID tandem MS using static nanospray in the enhanced product ion mode with variable collision energies. Ions representing a putatively glycosylated fragments were further fragmented in the MS^3 mode to examine the presence of the glycan.

In vitro assay of polypeptide GalNac-transferase activity – The ΔDS2,3-TSR4-derived peptide, α-biotinyl-GAGAGWSDASVTAGK-NH_2 and the MUC2-derived peptide, α-biotinyl-APTTTPITTTTTVTPTPTPTGTQTK-NH_2) were obtained from Bachem (Bubendorf, Switzerland). They were glycosylated by recombinant human GalNAc-T1, -T2, -T3 and –T4 from insect cells as previously described

(*119*). The reactions were performed in 25 µl 25 mM cacodylic acid, pH7.4, containing 10 mM $MnCl_2$, 0.25% Triton X-100, 1.5 mM UDP-GalNAc, 10 µg of ΔDS2,3-TSR4 and 0.4 µg of recombinant human GalNAc-transferases. Reaction conditions for the MUC2-derived peptide were same as for TSR4-derived one, except that the reaction mixtures contained 4% (v/v) DMSO. Reactions were terminated by diluting 0.5 µl incubation mixture with 3.5 µl of 0.1% TFA/H_2O and 1µl of each sample was applied on a tip and mixed with 25 mg/ml of 2,5-DHB dissolved in H_2O/CH_3CN (2:1) solution. MALDI-TOF mass spectra were obtained on a Voyager-DE™ Pro instrument (Applied Biosystems) operating at an accelerating voltage of 20 kV (grid voltage 94%, guide wire voltage 0.1%) in the linear mode with the delayed extraction setting. Recorded data were processed using Data Explorer ver.4.

Co-expression of GalNAc-T3 and ΔDS2,3-TSR4 in CHO-K1 cells – The plasmids encoding human polypeptide-N-acetylgalactosaminyltransferase 3 (a kind gift from Dr. E. Bennett) and ΔDS2,3-TSR4 were co-transfected into CHO-K1 cells using Lipofectamine as described above. As a control the plasmid pEGFP-N1 (Clontech) was substituted for the one encoding GalNAc-T3.

Analysis of ΔDS2,3-TSR4 from CHO-K1 by tandem LC-MS in the MRM mode - ΔDS2,3-TSR4 was purified and digested with endoproteinase Lys-C as described above. Peptides were separated by reversed-phase chromatography on a Magic C_{18} column (75 µm x 100 mm, Michrom BioResources, Auburn USA), using an Agilent 1100 Nanospray LC System (Agilent Technologies, Santa Clara, USA) interfaced to a quadrupole linear 4000 Q TRAP mass spectrometer. Digests were loaded onto a Peptide Captrap (Michrom BioResources, Auburn, USA) in 0.05% TFA and 2% acetonitrile at a flow rate of 10µl/min. Peptide were eluted with a 60-min linear gradient of 15-50% B (Solvent A: 0.1% formic acid, containing 2% acetonitrile; solvent B: 0.1% formic acid, 80% acetonitrile) at a flow rate of 300 nl/min. For determination of the glycosylation status of ΔDS2,3-TSR4 from CHO-K1 cells, the mass spectrometer was operated in an information-dependent acquisition mode. This involved monitoring specific transitions in the MRM mode to measure the amount of a particular peptide, triggering CID tandem MS to confirm its structure. Peptide-specific transitions for the K2* peptides of ΔDS2,3-TSR4 were first determined by tandem MSMS experiments using peptides from the protein obtained from HEK 293T cells. This experiment provided the required information on abundance, charge state and collision energy, necessary to select suitable transitions. A total of 28 transitions were selected (supplemental Table S1) and the dwell time for each of them was optimized. As a control for the amount of protein analyzed the non-glycosylated peptide K4* (SLAELGDCNEDLEQAEK) was also monitored. Its signal was used to normalize the data. The threshold for tandem MSMS was set to 500 cps and target ions were excluded for further measurements at 2 occurrences for 40 s.

Comment [h1]: This needs clarification!

RESULTS

Production of recombinant TSR4 from rat f-spondin. Initially, with the aim to determine whether proper protein folding is required for *O*-fucosylation of TSRs *in vivo*, we expressed the fourth TSR from rat f-spondin in HEK293T cells. The wild type protein (TSR4) and a mutant lacking the second and third disulfide bridge (ΔDS2,3-TSR4) migrated as two bands on an SDS-PAA gel, with apparent molecular weights of 18 and 20 kDa (data not shown). Protein sequencing by Edman degradation showed that the two bands differed in their N-terminal amino acid sequence, starting with DAA and TKL, respectively. The former start point represents the full-length recombinant TSR4, whereas the latter one appears to results from a furin-like cleavage (see experimental procedures for the complete amino acid sequence)

Analysis of glycosylation status. In full-length f-spondin, the fourth TSR carries Glc-β1,3-Fuc-*O*- on Thr-601 and is *C*-mannosylated on Trp-592 and Trp-595 (*61*). To examine the glycosylation pattern of the isolated modules, purified TSR4 and ΔDS2,3-TSR4 were reduced, carboxamidomethylated and digested with endoproteinase Lys-C. The resulting peptides

were analyzed by reverse-phase LC-MS (Fig. 1). Peptide K2, which contains all three known glycosylation sites, eluted as a series of glycoforms that differed in mass by 162 Da (Fig. 1A and Table 1). Low energy CID tandem MS demonstrated that this series resulted from the presence of 2 (MM), 1 (M0), or no *C*-mannosyl residues on the tryptophan residues[6]. Furthermore, their mass (Table 1) and the facile loss of 308 Da by in-source fragmentation showed the presence of the Glc-Fuc-*O*- (FG) disaccharide. Thus, three major glycopeptides were obtained from TSR4: K2+MMFG, K+2M0FG, and K2+00FG; Fig. 1A)[4]. In addition, we observed, in most cases, minor accompanying peaks that contained peptide forms with masses that were 162 and 308 Da lower (Table 1). These data agreed with the glycosylation pattern observed in full length f-spondin, i.e. nearly complete *C*-mannosylation of Trp-592, very partial *C*-mannosylation of Trp-595, and nearly complete modification of Thr-601 with the disaccharide Glc-Fuc-*O*- (*61*).

Also the peptide map of ΔDS2,3-TSR4 contained a series of peptides that differed in mass by 162 Da (K2*+MM00, K2*+M000 and K2*, Fig. 1B), that resulted from partial *C*-mannosylation of the two tryptophan residues. Importantly, the masses of these peptides indicated that the Glc-Fuc-*O*- disaccharide was absent (Table 1). However, each of these peptides was preceded by a peak containing a peptide that was 948 Da heavier (K2*+MM+ 948 Da, K2*+M0 + 948 Da and K2* + 948 Da; Fig.1B). An increase in mass of 948 Da would exactly correspond to a tetrasaccharide consisting of 1 hexosyl-, 1 N-acetylhexosaminyl- and 2 N-acetylneuraminic acid residues. Indeed, for each of these peptides we observed in-source fragmentation that produced a series of fragment ions that differed in mass by values corresponding to single sugar units (162, 203 and 291 Da). Thus ΔDS2,3-TSR4 appeared to have undergone some other form of glycosylation, in addition to *C*-mannosylation.

Peptide K2* contains four hydroxyl amino acids that could potentially be *O*-glycosylated. In order to simplify subsequent structural analysis by MS, the relevant peptides from the Lys-C digest of ΔDS2,3-TSR4 were pooled and further digested with endopeptidase Asp-N. This should yield the *N*-terminal fragments K2*-D1 (LCLLSPWSEWS, carrying variable number of *C*-mannosyl residues) and the *C*-terminal peptide K2*-D2 (DASVTAGK). In initial fractionation experiments, using reversed-phase C_{18} chromatography, only the differentially *C*-mannosylated peptides K2*-D1+MM, K2*-D1+M0 and K2*-D1 were recovered (data not shown). This suggested that the putative tetrasaccharide was linked to peptide K2*-D2, which was too polar to bind to the C_{18} reversed phase material. Therefore, we used a column containing porous graphitized carbon (Hypercarb) to isolate this fragment (Fig. 2). Static nanospray MS of the purified peptide revealed a doubly charged ion at *m/z* 848.5, corresponding to a mass of 1695 Da (Fig. 2, inset). The amino acid sequence of this peptide was confirmed by CID tandem MS (data not shown). Together, these results indicated that peptide K2*-D2 carried the putative tetrasaccharide (expected mass: 1695.66 Da).

Determination of the structure of the putative tetrasaccharide – To examine whether peptide K2*-D2 is glycosylated, we performed nanospray CID tandem MS experiments on the singly charged precursor (m/z 1696), using relatively low collision energies (10-30 V). This should yield structural information on the putative glycan, but cause minimal fragmentation of the polypeptide chain. The highest observed fragment mass (*m/z* 1404.7) is 291 mass units lower than the precursor and can be explained by the neutral loss of a terminal NeuNAc residue (Fig. 3). This fragment undergoes both a 162 and 291 Da loss, which can be explained by the presence of a linear or branched glycan (Fig. 3). The resulting fragment ions at *m/z* 1242.7 and 1113.7 can be explained as K2*D2 – NeuNAc – Hex and K2*D2 – 2NeuNAc, respectively. Further 291 and 162 Da losses yielded K2*D2 – 2 NeuNAc - Hex (*m/z* 951.6). Finally, the neutral loss of a HexNAc residue explains the formation of the fragment ion at *m/z* 748.4 (K2*D2 – 2 NeuNAc – Hex - HexNAc). A series of doubly charged fragment ions could be explained in the same way. These data showed that peptide K2*-D2 is indeed glycosylated, and the tandem MS spectrum is fully consistent with the presence of a

tetrasaccharide with the structure: NeuNAc-Hex-[NeuNAc]-HexNAc-*O*-.

To determine the structure of the tetrasaccharide in detail, we treated glycosylated peptide K2*D2 with glycosidases and analyzed the reaction products by LC-MS using the Hypercarb column. Treatment with the neuraminidase from *Clostridium perfrigens*, which cleaves α2-3-, α2-6- and α2-8-linked NeuNAc residues, resulted in a peptide with a mass of 1112 Da (Fig. 4A), showing the cleavage of two NeuNAc residues. Subsequent digestion of the desialylated glycopeptides with *O*-glycanase, which is specific for the core 1 structure (*120*), removed all carbohydrate as indicated by the mass of the product (747 Da; Fig. 4B). This result was confirmed by Western analysis of ΔDS2,3-TSR4, using the core 1-specific lectin PNA. We only observed binding of the lectin to ΔDS2,3-TSR4 that had been treated with the neuraminidase from *Clostridium perfringens* (Fig. 4C). Thus, the desialylated glycopeptide carries the structure Galβ1,3GalNAc-*O*-

Incubation of glycosylated K2*D2 with an α2-3-specific neuraminidase from *Salmonella typhimurium* yielded a peptide with a mass of 1403.7 Da (Fig. 4D), indicating the loss of a single α2-3-linked NeuNAc residue. We propose that the product of this reaction is K2*D2 with the trisaccharide Galβ1,3[NeuNAc]GalNAc-*O*- for the following reasons: 1. Since the 3-OH position of the GalNAc residue is occupied by the Gal, the NeuNAc residue removed in this experiment was attached to the 3-OH of the Gal residue. 2. The tandem MS spectrum of the completely glycosylated peptide showed the loss of a NeuNAc residue from the quasi-molecular ion, but not of a hexose (Fig. 3). Such a loss was only observed after loss of one NeuNAc residue. This indicates that the Gal moiety is sub-terminal to the NeuNAc residue. Furthermore, the remaining NeuNAc residue must be attached to GalNAc.

We have not determined the position on the GalNAc to which the NeuNAc is attached. It most likely occupies position 6, since NeuNAc attachment to the 4-OH has not been reported. Taken together, the data in Figs. 3 and 4 demonstrate that the tetrasaccharide has the structure NeuNAcα2,3Galβ1,3[NeuNAc]GalNAc-*O*-, an known *O*-linked glycan called disialyl-T-antigen (dST) (*121*)

Determination of the glycosylated amino acid residue. In glycosylated peptide K2*D2 Ser-599 and Thr-601 are possible attachment sites for the tetrasaccharide. Tandem MS experiments using collision energies that are sufficiently high to achieve backbone fragmentation always resulted in the loss of the two NeuNAc residues. We performed MS^3 experiments on diagnostic fragment ions that could potentially carry the core 1 structure or the GalNAc residue. This approach resulted in two fragment ions, which allowed us to deduce the modified amino acid (Figure 5). The MS^3 experiment on the fragment ion at *m/z* 741.5 (putatively y4: T(+365)AGK) exhibited sequential neutral losses of 162 and 203 Da (Fig. 5). Importantly, the b3 or b4 ions (DAS and DASV) were only observed as such, and no neutral sugar losses from these fragments were observed. From these data we conclude that the tetrasaccharide is attached to Thr-601 in ΔDS2,3-TSR4.

GalNActransferase 3 can glycosylate a ΔDS2,3-TSR4-derived peptide in vitro - Formation of the disialyl-T-antigen is initiated by the transfer of a GalNAc moiety, a reaction that probably is catalyzed by a member of the family of polypeptide GalNAc transferases (GalNAc-T). In order to identify potential candidate enzymes, a ΔDS2,3-TSR4-derived peptide spanning the putative Thr-601 acceptor amino acid and a MUC2 peptide, serving as a control, were synthesized. The peptides were incubated with purified recombinant human GalNAc-T1, -T2, -T3 and -T4 and the incorporation of a GalNAc moiety (203 Da) was examined by MALDI-TOF-MS. All examined GalNAc-T homologues modified the MUC2a peptide (Fig. 6B). However, only GalNAc-T1 and-T3 were found to incorporate a GalNAc moiety into the ΔDS2,3-TSR4-derived peptide. Contrary to transferase 1, which added two GalNAc moieties to the TSR-derived peptide, GalNAc-T3 attached a single GalNAc residue in a time- dependent manner. In order to identify the modified amino acids we purified the products of the reaction catalyzed by GalNAc-T1 an-T3 by HPLC. In both cases we found a

putative y4 ion at m/z 578.2 in CID tandem MS experiments, which would indicate glycosylation of Thr-601 (data not shown). Using the same MS^3 approach as described above, it was found that these ions exhibited a loss of 203 Da (shown in Fig. 7A for the product of GalNac-T3). The same result was obtained for the ppGalNAc-T1 product. However, in that case an additional glycosylation was indicated by the presence of a fragment ion at m/z 1202.5 (data not shown). This would correspond to the glycosylated b9 ion, GAGAGWS (+203 Da) suggesting modification of Ser-596. This was confirmed in an MS^3 experiment, which showed the neutral loss of the GalNAc residue (Fig. 7B). These data confirmed that of the tested GalNAc transferases, only GalNAc-T3 was able to modify a single residue in the ΔDS2,3-TSR4-derived peptide, i.e. Thr-601.

GalNAc-T3 modifies ΔDS2,3-TSR4 in cultured cells - In an initial screen of different cell lines, we noticed a very low level of modification of ΔDS2,3-TSR4 with dST in CHO-K1 cells. This allowed us to examine the effect of overexpression of GalNAc-T3 on the modification of ΔDS2,3-TSR4 with GalNAc *in vivo*. For this, the degree of glycosylation of ΔDS2,3-TSR4 purified from CHO-K1 cells that were co-transfected with GalNAc-T3 or GFP were compared by tandem LC-MS in the multiple reaction monitoring mode (MRM). In MRM experiments peptide-specific precursor ions are selected in the first quadrupole, fragmented in the collision cell and characteristic fragment ions are selected in the third quadrupole (the precursor-/specific fragment ion pair is called a 'transition'). The high selectivity and the short scan times necessary for measuring multiple transitions, allows sensitive and quantitative analysis (*122*). Glycan- as well as peptide backbone-specific transitions (Supplemental data; table S1) were used to monitor the disappearance of unmodified peptide K2*, as well as the appearance of the five different glycoforms, i.e. K2*+MMdST, K2*+MM0, K2*+M0dST, K2*+M00 and K2*+00dST [5]. An overlay of representative MRM chromatograms is shown in Fig. 8. Upon overexpression of GalNAc-T3 we observed a strong decrease in the species lacking *O*-glycosylation, i.e. K2*+MM0 and K2*+M00 (Fig. 8A and B). Concomitantly, a 5.3 ± 0.9 and 9.1 ± 2-fold increase in peptides K2*+MMdST and K2*+M0dST occurred, respectively. These changes were not due to differences in the amount of protein injected for analysis, as can be seen from the equal intensity of the non-glycosylated control peptide K4*. In all cases the identity of the dST was confirmed, and thus the mass of the polypeptide backbone, by a full tandem MS experiment that was triggered by the relevant transition (Supplemental data Fig.S1). Thus, GalNAc-T3 can not only modify the ΔDS2,3-TSR4-derived peptide *in vitro*, but also strongly enhances the formation of disialylated T-antigen in CHO-K1 cells.

DISCUSSION

1. Description of the results

TSRs undergo two types of glycosylation: attachment of the disaccharide Glcβ1,3-Fuc-*O*- to a Thr or Ser residue, and *C*-mannosylation of Trp residues (*61*). Here we report a mutant of f-spondin, lacking the second and the third disulfide bridge (ΔDS2,3-TSR4), which shows *C*-mannosylation on both Trps but harbors a disialylated-T-antigen (NeuNAcα2,3Galβ1,3-[NeuNAcα2,6]GalNAc-O-) on Thr-601. In a peptide screen we identified GalNAc transferase T3 as the putative enzyme initiating the *O*-glycan. This enzyme modifies exclusively Thr-601 and its co-expression together with the mutated TSR significantly increases the expression of the disialylated-T-antigen in CHO-K1 cells

2. Lack of Glc-Fuc-O

It has been shown recently that the *O*-fucosyltransferase POFUT2 and the glucosyltransferase β3Glc-T, that glycosylate TSRs, require a properly folded molecule to attach Glcβ1,3-Fuc-*O*- (*42, 65*). Deletion of two out of three disulfide bridges most likely alters the three-dimensional structure of the TSR in way that these enzymes loose their capability to recognize it as a substrate. Work from our own laboratory indicates that less severe changes affecting the 3D fold of the TSR molecule (i.e. replacement of only one disulfide bridge), still lead to a substantial modification with the

disaccharide Glcβ1,3-Fuc-*O*. This suggests that the native module has a compact structure which can only be distorted by severe structural changes[7]. The fact that the mutated TSR is not degraded in the ER but receives the Disialyl-T-antigen in the Golgi implies that the module still retains sufficient secondary and tertiary structure to traverse the secretory pathway. These features of ΔDS2,3-TSR4 seem to be valid not only in human embryo kidney cells (HEK 293T) but also in Chinese hamster ovary cells (CHO-K1).

3. Gain of Disialyl-Tantigen initiated by GalNAc T3

Alteration of the loop between strand A and strand B in the TSR module can lead to abrogation of the Glc-Fuc-*O* disaccharide without formation of the Disialyl-T-antigen[7]. Therefore, formation of the Disialyl-T-antigen was not merely an effect of lack of *O*-fucosylation. The enzyme initiating the observed core 1 modification is a UDP-N-acetyl-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase (ppGalNAc-T). Currently, this family consists of up to 20 predicted members, 16 of which have been cloned (*123*). Although a consensus sequence has not been defined, it is believed that the primary amino acid sequence is critical for enzyme substrate recognition (*124*) and that the *in vitro* specificity for distinct residues reflects the *in vivo* condition (*12*). In order to confirm this hypothesis, we screened for potential candidate enzymes using a peptide of ΔDS2,3-TSR4 spanning the Thr-601 residue and a MUC2a peptide as a control. This screen discovered UDP-GalNAc T1, T3 of the tested enzymes as the ones, capable of modifying the ΔDS2,3-TSR4-peptide. GalNAc T1 was shown to glycosylate as well Ser-596 and Thr-601, whereas GalNAc T3 solely glycosylated Thr-601 (Figure 7). Interestingly, we could confirm the potential of GalNAc T3 to glycosylate ΔDS2,3-TSR4 in CHO-K1 cells using a semi-quantitative MS approach. This method allowed us to monitor multiple glycoforms at the same time. CHO-K1 cells exhibited already basal activity to modify ΔDS2,3-TSR4 with a disialylated-T-antigen, but this activity could be enhanced significantly upon over expression of GalNAc-T3. The observation that GalNAc T3 can glycosylate ΔDS2,3-TSR4 *in vitro* and *in vivo* extends the list of known substrates for this glycosyltransferase. GalNAc T3 has been described to be overexpressed in various cancers (*125, 126*). It was suggested that in cancer overexpression of tightly regulated GalNAc transferases could result in initiation of *O*-glycosylation at normally unoccupied sites, which may contribute to cancer (*127*) In this context the usage of specific peptides of known substrates of GalNAc T3 like ΔDS2,3-TSR4, could help to further define the specificity of this glycosyltransferase.

4. ΔDS2,3-TSR4 is an experimental proof of a gain of O-glycosylation

Mutations resulting in a loss of glycosylation have been reported in various different genes. It was estimated that at least 170 mutations result in loss of *N*-glycosylation and 230 mutations influence directly *O*-glycosylation (*128*). Mutations that affect synthesis of glycans can clinically manifest in diseases known as congenital disorders of glycosylation (CDGs). Interestingly, mutations resulting in a gain of glycosylation are less well studied. The best characterized example of a gain of *N*-glycosylation is the T168N mutation in the IFNγR2, which is sufficient and necessary for an inherited disease called Mendelian susceptibility for mycobacterial disease (MSMD) (28). This paper provides an experimental proof of a gain of *O*-glycosylation. The same threonine residue that is modified with the disaccharide Glcβ1,3-Fuc-*O*- in wild type TSR4 receives the tetrasaccharide NeuNAcα2,3Galβ1,3 [NeuNAcα2,6]GalNAc-O in ΔDS2,3-TSR4. Whereas the disaccharide Glcβ1,3-Fuc-O is synthesized in the ER, all enzymes required for the synthesis of mucin-type *O*-glycans are known to localize in the Golgi. Thus the discovered glycosylation represents not only a switch in specificity but also encompasses different cellular organelles. This shows the intriguing flexibility of glycosylation and provides experimental evidence that gain of *O*-glycosylation may exist.

REFERENCES

1. Oriol, R., Martinez-Duncker, I., Chantret, I., Mollicone, R., and Codogno, P. (2002) *Mol Biol Evol* **19**, 1451-1463

2. Kato, K., Jeanneau, C., Tarp, M. A., Benet-Pages, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., and Clausen, H. (2006) *J Biol Chem* **281**, 18370-18377
3. Gerken, T. A., Raman, J., Fritz, T. A., and Jamison, O. (2006) *J Biol Chem* **281**, 32403-32416
4. Baldus, S. E., Wienand, J. R., Werner, J. P., Landsberg, S., Drebber, U., Hanisch, F. G., and Dienes, H. P. (2005) *Int J Oncol* **27**, 1289-1297
5. Campbell, B. J., Yu, L. G., and Rhodes, J. M. (2001) *Glycoconj J* **18**, 851-858
6. Santos-Silva, F., Fonseca, A., Caffrey, T., Carvalho, F., Mesquita, P., Reis, C., Almeida, R., David, L., and Hollingsworth, M. A. (2005) *Glycobiology* **15**, 511-517
7. Nakakura, N., Hietter, H., Van Dorsselaer, A., and Luu, B. (1992) *Eur J Biochem* **204**, 147-153
8. Harris, R. J., and Spellman, M. W. (1993) *Glycobiology* **3**, 219-224
9. Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F., and Peter-Katalinic, J. (2001) *J Biol Chem* **276**, 6485-6498
10. Gonzalez de Peredo, A., Klein, D., Macek, B., Hess, D., Peter-Katalinic, J., and Hofsteenge, J. (2002) *Mol Cell Proteomics* **1**, 11-18
11. Wang, Y., and Spellman, M. W. (1998) *J Biol Chem* **273**, 8112-8118
12. Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006) *J Biol Chem* **281**, 9385-9392
13. Kozma, K., Keusch, J. J., Hegemann, B., Luther, K. B., Klein, D., Hess, D., Haltiwanger, R. S., and Hofsteenge, J. (2006) *J Biol Chem* **281**, 36742-36751
14. Canevascini, S., Kozma, K., Grob, M., Althaus, J., Klein, D., Chiquet-, and Ehrismann, R., and Hofsteenge, J. . (2006). in *European Worm Meeting, April 29-May 3, 2006*, Hersonissos, Crete, Greece
15. Lesnik Oberstein, S. A., Kriek, M., White, S. J., Kalf, M. E., Szuhai, K., den Dunnen, J. T., Breuning, M. H., and Hennekam, R. C. (2006) *Am J Hum Genet* **79**, 562-566
16. Tan, K., Duquette, M., Liu, J. H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J., and Wang, J. H. (2002) *J Cell Biol* **159**, 373-382
17. Paakkonen, K., Tossavainen, H., Permi, P., Rakkolainen, H., Rauvala, H., Raulo, E., Kilpelainen, I., and Guntert, P. (2006) *Proteins* **64**, 665-672
18. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* **77**, 51-59
19. Krieg, J., Hartmann, S., Vicentini, A., Glasner, W., Hess, D., and Hofsteenge, J. (1998) *Mol Biol Cell* **9**, 301-309
20. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) *Nature* **379**, 466-469
21. Sorensen, A. L., Reis, C. A., Tarp, M. A., Mandel, U., Ramachandran, K., Sankaranarayanan, V., Schwientek, T., Graham, R., Taylor-Papadimitriou, J., Hollingsworth, M. A., Burchell, J., and Clausen, H. (2006) *Glycobiology* **16**, 96-107
22. Dwek, R. A., Edge, C. J., Harvey, D. J., Wormald, M. R., and Parekh, R. B. (1993) *Annu Rev Biochem* **62**, 65-100
23. Dall'Olio, F., and Chiricolo, M. (2001) *Glycoconj J* **18**, 841-850
24. Hulsmeier, A. J., Paesold-Burda, P., and Hennet, T. (2007) *Mol Cell Proteomics* **6**, 2132-2138
25. Luo, Y., Koles, K., Vorndam, W., Haltiwanger, R. S., and Panin, V. M. (2006) *J Biol Chem* **281**, 9393-9399
26. Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., and Clausen, H. (2007) *Glycobiology* **17**, 374-387
27. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) *J Biol Chem* **275**, 38197-38205

28. DeFrees, S., Wang, Z. G., Xing, R., Scott, A. E., Wang, J., Zopf, D., Gouty, D. L., Sjoberg, E. R., Panneerselvam, K., Brinkman-Van der Linden, E. C., Bayer, R. J., Tarp, M. A., and Clausen, H. (2006) *Glycobiology* **16**, 833-843
29. Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., and Mandel, U. (2007) *Acta Histochem* **109**, 273-284
30. Sutherlin, M. E., Nishimori, I., Caffrey, T., Bennett, E. P., Hassan, H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M. A. (1997) *Cancer Res* **57**, 4744-4748
31. Vogt, G., Vogt, B., Chuzhanova, N., Julenius, K., Cooper, D. N., and Casanova, J. L. (2007) *Curr Opin Genet Dev* **17**, 245-251

ACKNOWLEDGMENTS
Special thanks to Ms. C. Heinrich (FMI) for critical reading of the manuscript, to Mrs. S. Schleeger (Novartis AG) and Mr. Chen (FMI) for discussion and to Ms. C. Keller (FMI) for technical assistance.

FOOTNOTES

***** The work at the Friedrich Miescher Institute was supported by the Novartis Research Foundation

[1] To whom correspondence should be addressed
Telephone: +41 61 6974722
FAX: +41 61 6973976
Email: jan.hofsteenge@fmi.ch

[2].The Tn and T-antigen as well as sialylated substitutions are commonly referred as Thomsen-Friedenreich-related antigens

[3] The abbreviations used are: CDG, congenital disorders of glycosylation; CID, collision induced dissociation; IDA, information dependent acquisition; LC, liquid chromatography; MRM, multiple reaction monitoring; MS, mass spectrometry; PAA, Polyacrylamide; POFUT, protein *O*-fucosyltransferase; SDS, sodium dodecylsulphate; S/N: Signal to noise ratio; TFA: trifluoro acetic acid; TSR, thrombospondin type 1 repeat

[5] The nomenclature used here is: M:C-mannose, F:Fucose, G:Glucose, 0:No modification.

[6] The glycoforms of K2* are indicated in the order of their occurrence in the peptide. The first M being *C*-mannosylation of Trp-592, the second M being *C*-mannosylation of Trp-595 and dST being Disialyl-T-antigen on Thr-601. '0' indicates no glycosylation of the corresponding residue.

[7] Klein, D., Macek, B. and Hofsteenge, J., manuscript in preparation.

FIGURE LEGENDS

Figure 1. **Peptide mapping of wild type TSR4 and the mutant ΔDS2,3-TSR4 by LC-MS.** *A,* Partial UV chromatogram of the peptide map generated by cleavage of wild type TSR4 with endoproteinase Lys-C. Only the portion with the different glycoforms of peptide K2 (inset) is shown. These were identified by their mass and CID tandem MS, and are indicated with M representing a *C*-mannosylated tryptophan (underlined in the amino acid sequence), FG a threonine (indicated in bold in the amino acid sequence) carrying the Glc-Fuc-*O*- disaccharide and '0' the absence of modification. In addition to glycopeptides carrying the Glc-Fuc-*O*- disaccharide, also peptides lacking Glc (162 Da) and/or Fuc (146 Da) were observed in chromatographically separated peaks. *B*, Partial UV chromatogram of the Lys-C peptide map of ΔDS2,3-TSR4, the mutant lacking the second and third disulfide bridge. Glycoforms of peptide K2* (see inset, with Cys-598 and -602 replaced by Ala indicated in italics) were identified as in '*A*'. Peptides lacking the Glc-Fuc-*O*- disaccharide were readily identified: K2* + MM00, K2* + M000 and K2*. Each of these peptides was preceded by a peptide that has a 948 Da higher mass.

Figure 2. **Isolation and LC-MS analysis of peptide K2*-D2, carrying a putative tetrasaccharide.** Peptides K2*+MM+947Da, K2*+M0+948 Da and K2*+948 Da (see Fig. 1B) were pooled and cleaved with endopeptidase Asp-N. The C-terminal fragment was purified on a porous graphite carbon column (Hypercarb) and its mass was determined by enhanced resolution MS to be 1695 Da (inset).

Figure 3. **CID MS characterization of the putative tetrasaccharide.** The doubly-charged precursor of peptide K2*-D2 was analyzed by CID tandem MS in the static nanospray mode at low collision energy (10-30 V). Neutral losses corresponding to the mass of monosaccharide constituents are indicated. The structures assigned to the attached glycan are based on this spectrum together with the results from the experiment depicted in Fig. 4, since the exact nature of the sugars can not be deduced from mass spectrometry. The fragment ion at m/z 1404.7 results from the loss of a NeuNAc residue from the precursor ion. The subsequent concomitant loss of a hexose and NeuNac residue indicates that the glycan has a bi-antennary structure. The spectrum also shows a series of doubly-charged fragment ions, which confirm these assignments. They are labeled with their m/z value only for sake of clarity.

Figure 4. **Determination of the tetrasaccharide structure.**
A, Digestion of glycosylated peptide K2*-D2 with neuraminidase from *Clostridium perfringens* resulted in a glycopeptide with 582 Da lower mass, corresponding to the loss of two NeuNAc residues. *B*, The product of the reaction in '*A*' was further treated with *O*-Glycanase, which yielded peptide K2*-D2 that lacks all sugars as indicated by the singly charged ion at *m/z* 748.4. *C*, ΔDS2,3-TSR4 was treated with neuraminidase from *Clostridium perfringens* and compared to the untreated protein by Western blot using peanut agglutinin (PNA; upper panel). To verify the loading of equal amounts of protein, the blot was re-probed with an antibody against the Myc tag (lower panel). *D*, Digestion of glycosylated peptide K2*-D2 with an α2-3-specific neuraminidase from *Salmonella typhimurium* resulted in a peptide with a 291 Da lower mass, indicating the presence of a single α2-3-linked NeuNAc residue in the original glycopeptide.

Figure 5. **Determination of the amino acid residue carrying the tetrasaccharide.** Doubly charged glycosylated peptide K2*-D2 was fragmented by CID in the static nanospray mode, using variable collision energies, to produce peptide sequence ions. Under these conditions, the glycopeptides underwent neutral loss of two NeuNAc residues. The fragment ion at m/z 741.4 (putatively y4 T(+ 365 Da)AGK) was further fragmented in an MS^3 experiment. The sequential loss of a hexose- (162 Da) and a HexNAc residue (203 Da) is indicated.

Figure 6. ***In vitro*** **peptide modification by GalNAc transferases.** *A*, A peptide derived from the sequence of ΔDS2,3-TSR4 (underlined in the inset) was incubated for various length of time with GalNAc-T1, -T2, -T3 and -T4. Modification was monitored by MALDI-TOF MS, monitoring the loss of

starting material (*m/z* 1560.7) and the formation of GalNAcylated products. The observed *m/z* values are indicated with in parenthesis the number of incorporated GalNAc residues. *B*, For comparison the mucine 2-derived peptide (inset) was treated and analyzed in the same way.

Figure 7. **Characterization of the glycosylated products from the *in vitro* reactions catalyzed by GalNAc-T1 and –T3.** *A*, Glycosylated ΔDS2,3-TSR4-derived peptide obtained from the reaction catalyzed by GalNAc-T3, was analyzed by CID tandem MS in the static nanospray mode (data not shown), and the fragment ion at m/z 578.2 (putatively y4 T(+ 203 Da)AGK) was further fragmented in an MS^3 experiment. The loss of 203 Da indicates the presence of the GalNAc residue on Thr-601. *B*, Analysis of the product obtained form the GalNAc-T1-catalyzed reaction by the same approach, showed this enzyme also modifies Ser-596.

Figure 8. **Modification of ΔDS2,3-TSR4 by GalNAc-T3 in CHO-K1 cells.** CHO-K1 cells were co-transfected with plasmids encoding ΔDS2,3-TSR4 and GalNAc-T3 (—) or as a control ΔDS2,3-TSR4 and GFP (…..) Purified ΔDS2,3-TSR4 was digested with endoproteinase Lys-C and the state of glycosylation of peptide K2* from the two experiments was compared by relative quantification using tandem LC-MS in the multiple reaction monitoring mode. *A*, peptide K2* containing two C-mannosyltryptophan residue was monitored using the indicated transitions (precursor/product) indicated in italic (see also supplemental data, Table S1). The non-glycosylated control peptide K4* was monitored to assess the amount of injected digest. *B*, The same analysis was performed for peptide K2* with one C-mannosyltryptophan residue. Note that the order of elution of glycopeptides with or without dST differs from that in Fig.1, as a result of the use of formic acid instead of TFA in the eluent.

Table 1: Identification of the glycoforms of peptide K2 from TSR4 and K2DS2,3-TSR4.[a]*

Protein	Average observed Mass [Da]	Assignment	Average theoretical Mass [Da][a]
TSR4	2917.2	K2+MMFG	2918.1
	2754.8	K2+MMF0	2756.0
	2609.2	K2+MM00	2609. 9
	2754.8	K2+M0FG	2756.0
	2592.4	K2+M0F0	2593. 9
	2446.8	K2+M000	2447.8
	2592.4	K2+00FG	2593. 9
	2430.8	K2+00F0	2431.8
	2284.4	K2	2285.6
ΔD2,3-TSR4	3378.8	K2*+MM+948 Da	3379.5
	2429.2	K2*+MM	2431.7
	3216.4	K2*+M0+ 948 Da	3217.4
	2268.4	K2*+M0	2269.5
	3054	K2*+ 948 Da	3055.2
	2106.8	K2*	2107.4

[a] The average theoretical mass for the given assignment is shown. The assignments are supported by CID tandem MS analyses.

Figure 1

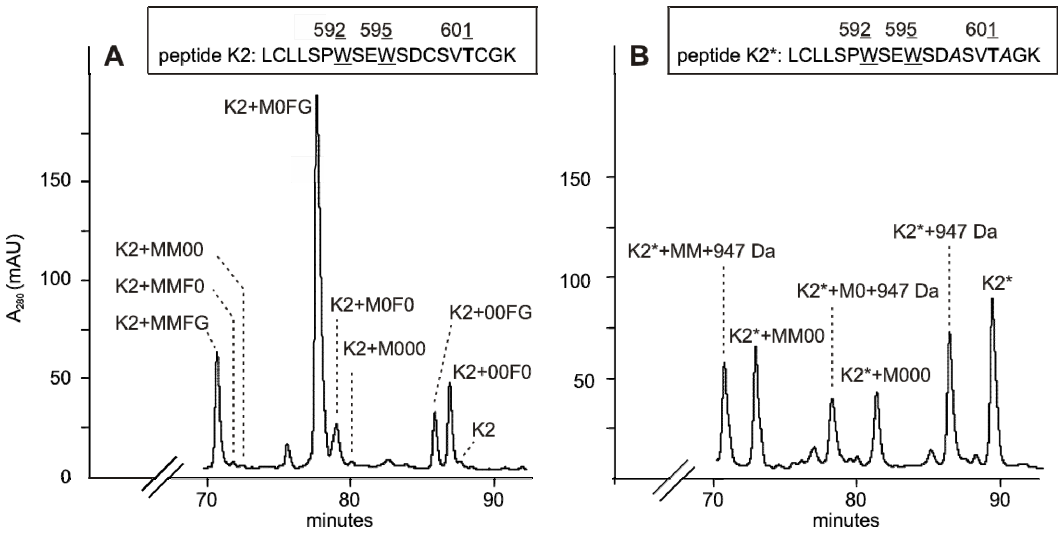

A
592 595 601
peptide K2: LCLLSPWSEWSDCSVTCGK
K2+M0FG
K2+MM00
K2+MMF0
K2+MMFG
K2+M0F0
K2+M000
K2+00FG
K2+00F0
K2
A280 (mAU)
150
100
50
0
70
80
90
minutes
B
592 595 601
peptide K2*: LCLLSPWSEWSDASVTAGK
K2*+MM+947 Da
K2*+947 Da
K2*+M0+947 Da
K2*
K2*+MM00
K2*+M000
150
100
50
70
80
90
minutes

Figure 2

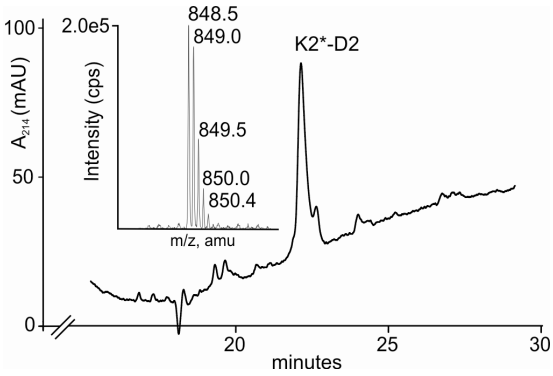

100
50
0
A_{214} (mAU)
2.0e5
Intensity (cps)
848.5
849.0
849.5
850.0
850.4
m/z, amu
K2*-D2
20
25
30
minutes

Figure 3

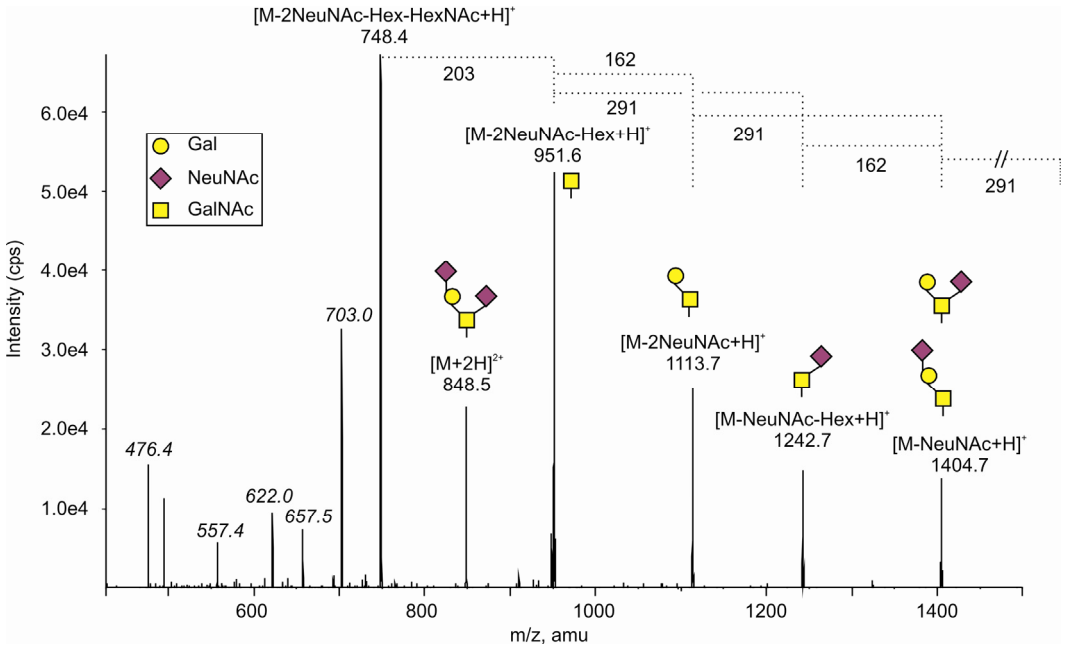

[M-2NeuNAc-Hex-HexNAc+H]+
748.4
203
162
291
[M-2NeuNAc-Hex+H]+
951.6
291
162
//
291
Gal
NeuNAc
GalNAc
Intensity (cps)
6.0e4
5.0e4
4.0e4
3.0e4
2.0e4
1.0e4
703.0
[M+2H]2+
848.5
[M-2NeuNAc+H]+
1113.7
[M-NeuNAc-Hex+H]+
1242.7
[M-NeuNAc+H]+
1404.7
476.4
622.0
657.5
557.4
600
800
1000
1200
1400
m/z, amu

Figure 4

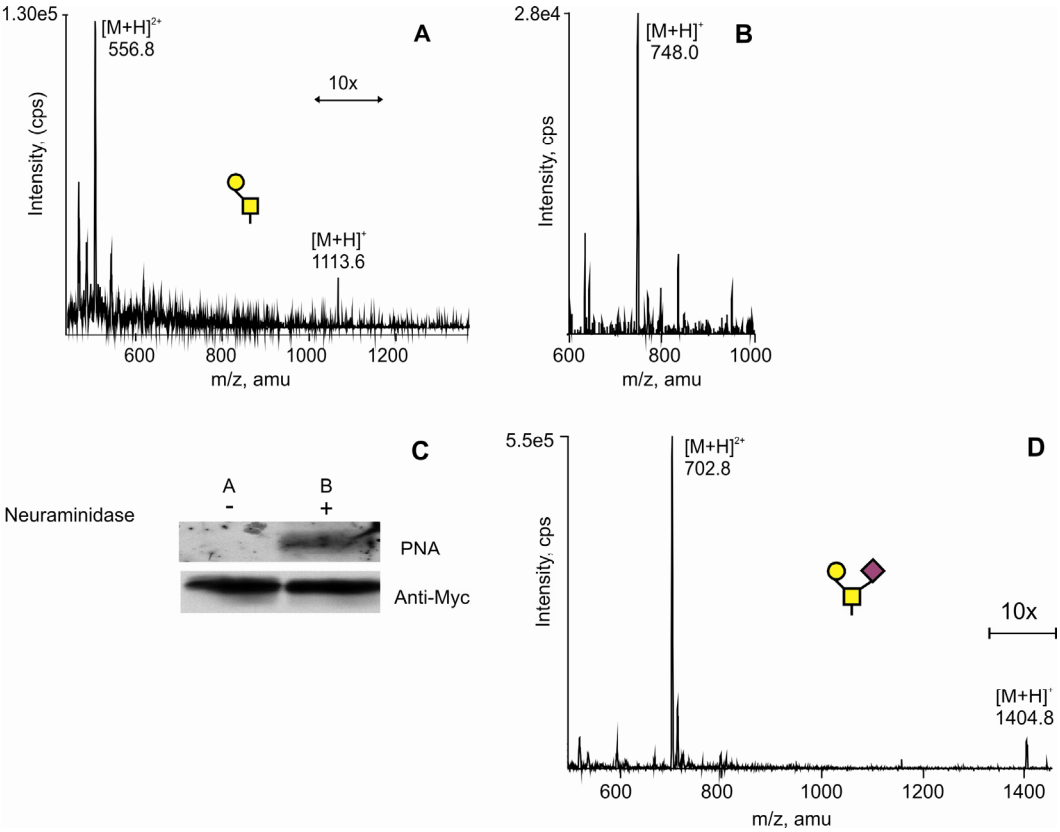

1.30e5
[M+H]2+
556.8
A
10x
Intensity, (cps)
[M+H]+
1113.6
600
800
1000
1200
m/z, amu
2.8e4
[M+H]+
748.0
B
Intensity, cps
600
800
1000
m/z, amu
C
A
-
B
+
Neuraminidase
PNA
Anti-Myc
5.5e5
[M+H]2+
702.8
D
Intensity, cps
10x
[M+H]+
1404.8
600
800
1000
1200
1400
m/z, amu

Figure 5

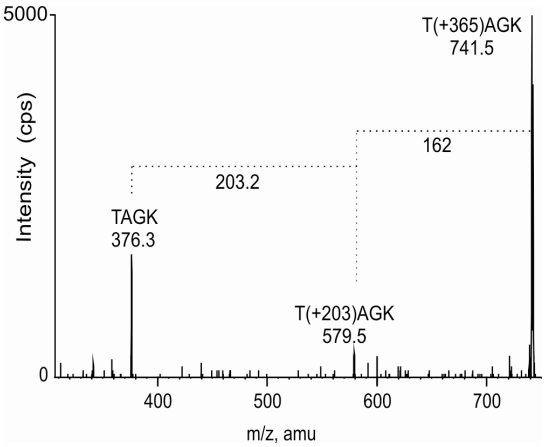

5000
0
Intensity (cps)
T(+365)AGK
741.5
162
203.2
TAGK
376.3
T(+203)AGK
579.5
400
500
600
700
m/z, amu

Figure 6

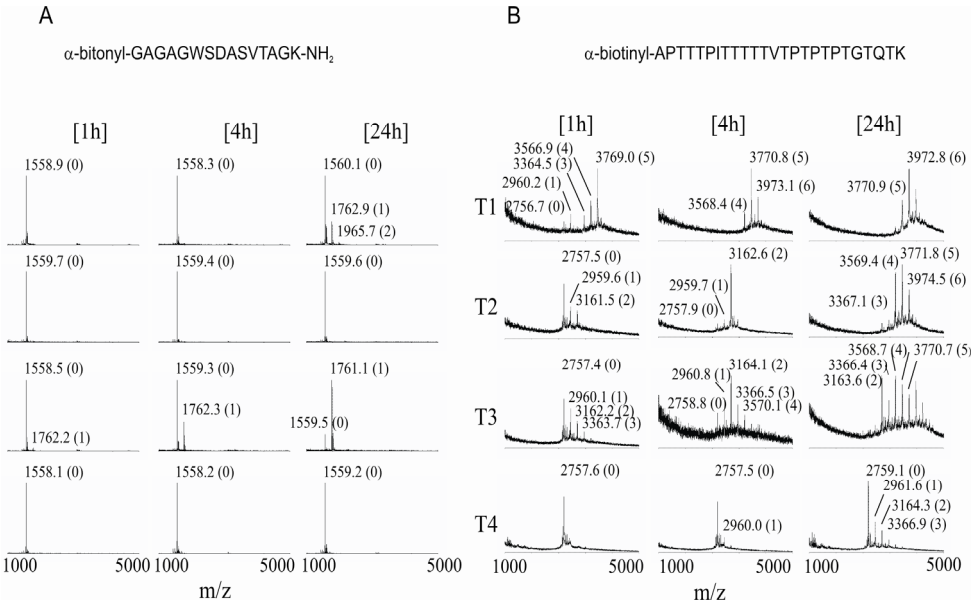

A
α-bitonyl-GAGAGWSDASVTAGK-NH2
[1h]
[4h]
[24h]
1558.9 (0)
1558.3 (0)
1560.1 (0)
1762.9 (1)
1965.7 (2)
1559.7 (0)
1559.4 (0)
1559.6 (0)
1558.5 (0)
1559.3 (0)
1761.1 (1)
1762.3 (1)
1762.2 (1)
1559.5 (0)
1558.1 (0)
1558.2 (0)
1559.2 (0)
1000
5000
m/z
B
α-biotinyl-APTTTPITTTTTVTPTPTPTGTQTK
[1h]
[4h]
[24h]
T1
T2
T3
T4
3566.9 (4)
3364.5 (3)
3769.0 (5)
2960.2 (1)
2756.7 (0)
3770.8 (5)
3973.1 (6)
3568.4 (4)
3972.8 (6)
3770.9 (5)
2757.5 (0)
2959.6 (1)
3161.5 (2)
3162.6 (2)
2959.7 (1)
2757.9 (0)
3569.4 (4)
3771.8 (5)
3974.5 (6)
3367.1 (3)
2757.4 (0)
2960.1 (1)
3162.2 (2)
3363.7 (3)
2960.8 (1)
3164.1 (2)
3366.5 (3)
2758.8 (0)
3570.1 (4)
3568.7 (4)
3770.7 (5)
3366.4 (3)
3163.6 (2)
2757.6 (0)
2757.5 (0)
2960.0 (1)
2759.1 (0)
2961.6 (1)
3164.3 (2)
3366.9 (3)
1000
5000
m/z

Figure 7

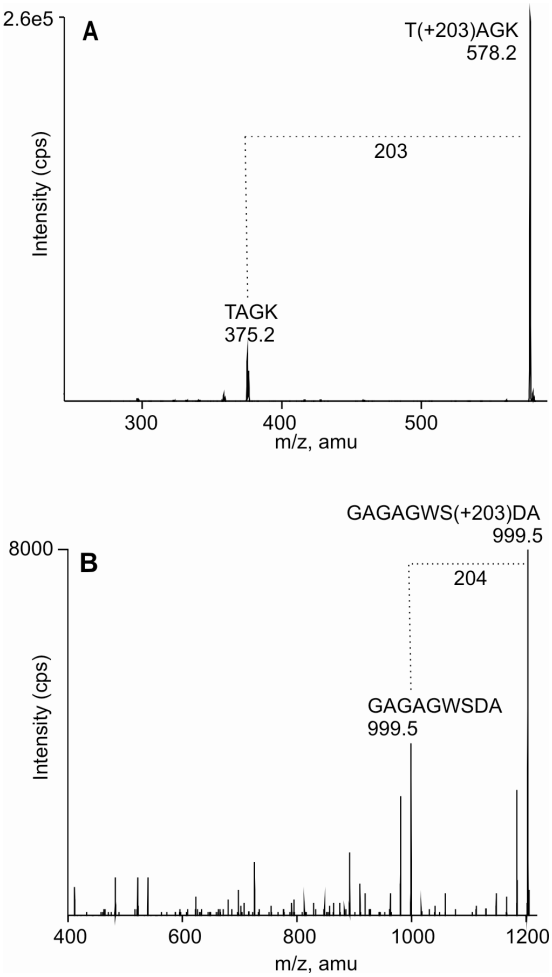

2.6e5
A
T(+203)AGK
578.2
203
TAGK
375.2
Intensity (cps)
300
400
500
m/z, amu
GAGAGWS(+203)DA
999.5
8000
B
204
GAGAGWSDA
999.5
Intensity (cps)
400
600
800
1000
1200
m/z, amu

Figure 8

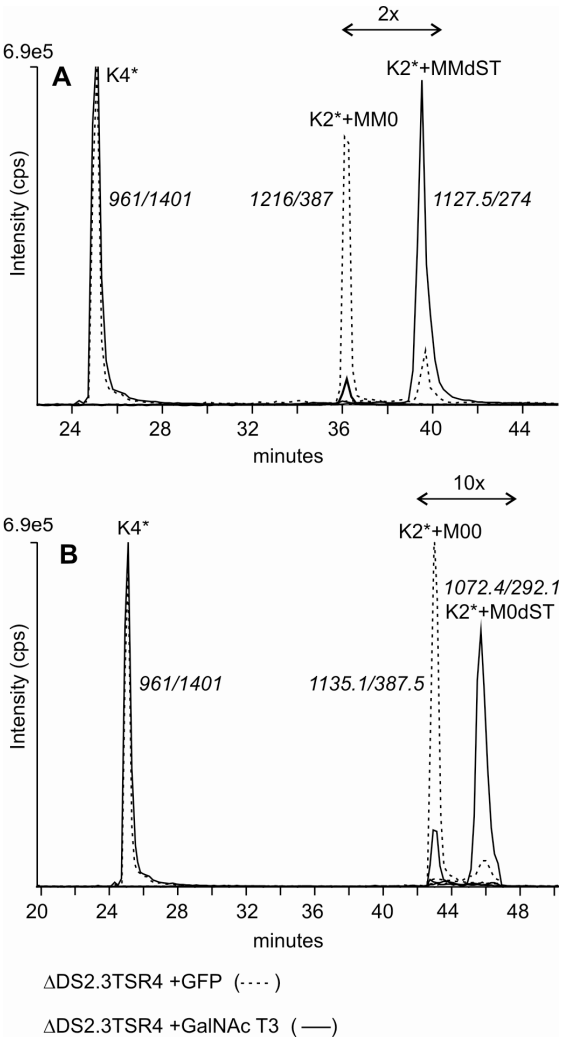
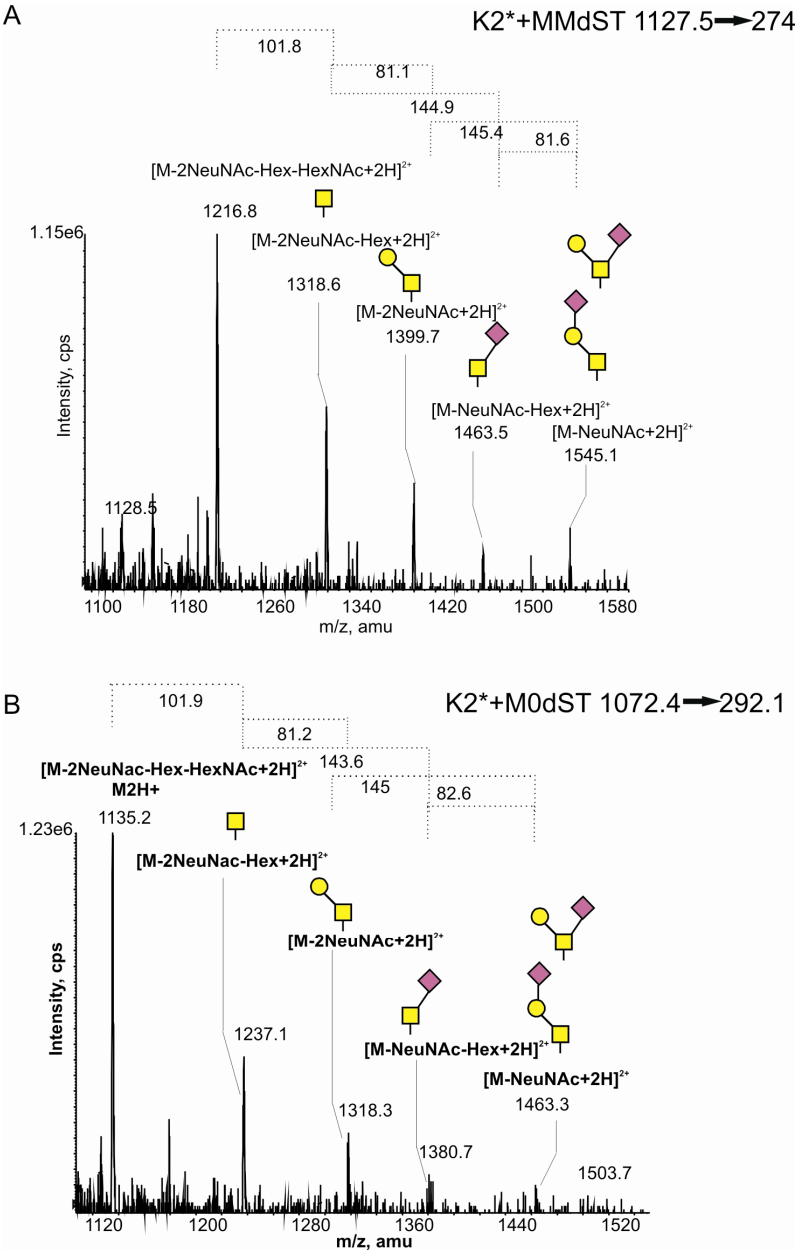

2x
6.9e5
A
K4*
K2*+MMdST
K2*+MM0
961/1401
1216/387
1127.5/274
Intensity (cps)
24
28
32
36
40
44
minutes
10x
6.9e5
B
K4*
K2*+M00
1072.4/292.1
K2*+M0dST
961/1401
1135.1/387.5
Intensity (cps)
20
24
28
32
36
40
44
48
minutes
ΔDS2.3TSR4 +GFP (- - - -)
ΔDS2.3TSR4 +GalNAc T3 (——)

Table S1. Transitions used in LC-MRM-driven IDA experiments

Precursor			Product			Collision energy (V)	Dwell time (ms)
Identity	**m/z**	**Charge**	**Identity**	**m/z**	**Charge**		
K4*	961	2	b3	272	1	53	60
K4*	961	2	b4	401	1	53	60
K2*+MMdST[a]	1127.5	3	b2	274	1	30	200
K2*+MMdST	1127.5	3	NeuNAc	292	1	20	200
K2*+MMdST	1127.5	3	y4	376.24	1	40	200
K2*+MMdST	1127.5	3	K2*+MM0	1216	2	30	200
K2*+MMdST	1127.5	3	K2*+MMmST[b]	1544.8	2	20	200
K2*+MM	1216	2	b2	274.2	1	65.8	200
K2*+MM	1216	2	LLS-H_20[c]	296.2	1	65.8	200
K2*+MM	1216	2	b3	387.5	1	65.8	200
K2*+MM	1216	2	y7	633.2	1	65.8	200
K2*+M0dST	1072.4	3	b2	274.2	1	30	200
K2*+M0dST	1072.4	3	NeuNAc	292.1	1	20	200
K2*+M0dST	1072.4	3	y4	376.2	1	40	200
K2*+M0dST	1072.4	3	K2*+M00	1135.2	2	30	200
K2*+M0dST	1072.4	3	K2*+MMTn[d]	1236.6	2	30	200
K2*+M00	1135.1	2	b2	274.1	1	61.7	200
K2*+M00	1135.1	2	LLS-H_20	296.3	1	61.7	200
K2*+M00	1135.1	2	b3	387.5	1	61.7	200
K2*+M00	1135.1	2	y7	633.2	1	61.7	200
K2*+00dST	1019.5	3	b2	274.1	1	30	200
K2*+00dST	1019.5	3	NeuNAc	292.1	1	20	200
K2*+00dST	1019.5	3	y4	376.2	1	40	200
K2*+00dST	1019.5	3	K2*	1054	2	30	200
K2*	1054	2	b2	274.1	1	61.7	200
K2*	1054	2	LLS-H_20	296.3	1	61.7	200
K2*	1054	2	b3	387.5	1	61.7	200
K2*	1054	2	y7	633.2	1	61.7	200

[a] dST: Disialyl-T-antigen
[b] mST: Monosialyl-T-antigen
[c] LLS – H_2O: Internal fragment ion, Leu-Leu-Ser -H_2O
[d] Tn: Tn antigen

Figure S1


A
K2*+MMdST 1127.5→274
101.8
81.1
144.9
145.4
81.6
[M-2NeuNAc-Hex-HexNAc+2H]2+
1216.8
1.15e6
[M-2NeuNAc-Hex+2H]2+
1318.6
[M-2NeuNAc+2H]2+
1399.7
Intensity, cps
[M-NeuNAc-Hex+2H]2+
1463.5
[M-NeuNAc+2H]2+
1545.1
1128.5
1100
1180
1260
1340
1420
1500
1580
m/z, amu
B
K2*+M0dST 1072.4→292.1
101.9
81.2
[M-2NeuNac-Hex-HexNAc+2H]2+
M2H+
143.6
145
82.6
1135.2
1.23e6
[M-2NeuNac-Hex+2H]2+
[M-2NeuNAc+2H]2+
Intensity, cps
1237.1
[M-NeuNAc-Hex+2H]2+
[M-NeuNAc+2H]2+
1463.3
1318.3
1380.7
1503.7
1120
1200
1280
1360
1440
1520
m/z, amu

4.3 Supplementary data

4.3.1 Expression and structural studies

Results

In ΔDS2,3-TSR4 two disulfide bridges were deleted. These changes result most likely in alterations of the three dimensional structure. Severe distortions of the three dimensional fold often lead to misfolded proteins that accumulate inside the cells. Since accumulated proteins are toxic, the cell has developed various ways to solve this problem. In order to enhance folding, chaperons are upregulated. Terminal misfolded proteins are removed from the ER and degraded.

An indication that the introduced mutations in ΔDS2,3-TSR4 result in a severely misfolded protein, would be decreased secretion levels of ΔDS2,3-TSR4 compared to wild type TSR4. Therefore, wild type TSR4 and ΔDS2,3-TSR4 were transiently expressed in HEK 293T cells. The proteins were purified from the conditioned medium as described in chapter 3.10 and secretion levels were quantified. For quantification, a sensitive TSR specific ELISA was developed that allowed analysis of expressed TSRs under native conditions (Figure 8).

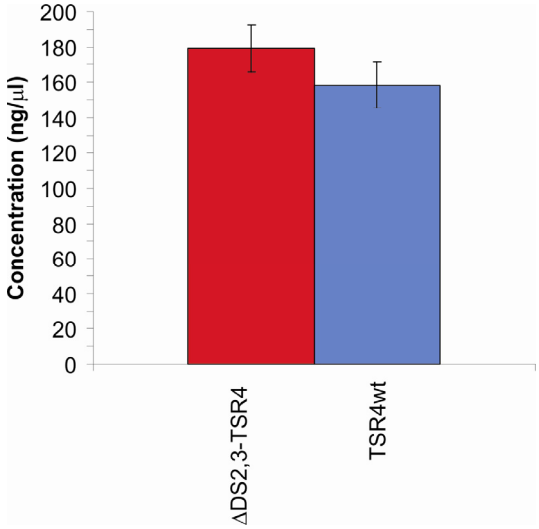


Figure 8: ELISA of secreted TSR4wt and ΔDS2,3-TSR4.

HEK 293T cells were transiently transfected with TSR4wt and ΔDS2,3-TSR4. The secreted proteins were purified based on their His_6-tag and the total amount was determined by ELISA.

Quantification of ΔDS2,3-TSR4 compared to wild type TSR4 revealed no significant differences in the secretion levels. This observation was surprising, since it indicated that ΔDS2,3-TSR4 was not misfolded but rather contained some structure allowing the molecule to traverse ER and Golgi.

Western blot analysis of wild type TSR4 and ΔDS2,3-TSR4 revealed that both proteins displayed significant heterogeneity. It was observed, that wild type TSR4 migrated as three bands between 16 band 20 kDa. The identities of the bands at about 20 kDa and at 18 kDa were determined by N-terminal sequencing (not shown). The band at 20 kDa demonstrated to be wild type TSR4 starting with the sequence DAA, whereas the major band at 18 kDa turned out to be wild type TSR4 starting with TKL. The blot of ΔDS2,3-TSR4 revealed three bands migrating at a slightly higher molecular weight than the bands of wild type TSR4. Additional bands were also detected at about 25 and 50 kDa for ΔDS2,3-TSR4 which indicated higher oligomers (see Figure 9).

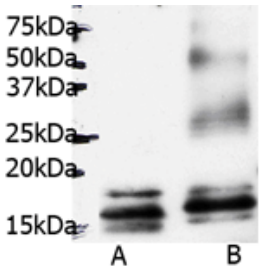


Figure 9: Western Blot of TSR4wt and ΔDS2,3-TSR4

Wild type TSR4 (lane A) and ΔDS2,3-TSR4 (lane B) were transiently expressed in HEK 293T cells, purified by IMAC and blotted using a mouse anti-Myc antibody as primary and anti-mouse-HRP as secondary antibody.

In order to gain information about the structure of ΔDS2,3-TSR4 under physiological conditions, ΔDS2,3-TSR4 and wild type TSR4 were analyzed by analytical gel filtration where proteins elute in the order un-, partially and fully-folded proteins. It was possible that ΔDS2,3-TSR4 represents a partially folded intermediate that possess native-like secondary structure, but due to the lacking of disulfide bridges can not collapse into its compact three-dimensional conformation. These folding intermediates are called molten globules (*129*) and occur under physiological conditions in the von Hippel-Lindau tumor suppressor (*99*), the apolipoproteins E3 and E4 (*130*) and in the α(S1) and κ-casein milk

proteins (*131*). An experimental value that gives indication about the folding state is the Stokes radius that decreases with the degree of folding. The Stokes radius of an unknown analyte can be estimated based on the elution volumes of globular standard proteins (see Chapter 3.13). Wild type TSR4 and ΔDS2,3-TSR4 were chromatographed on a Superdex 75 100 300/GL gel filtration column that had been calibrated with known standards (see Figure 6A). In order to mimic denatured proteins, TSRs were reduced, cysteins carboxyamidomethylated and the proteins were chromatographed under the same conditions than the un-treated "native" proteins (Figure 10).

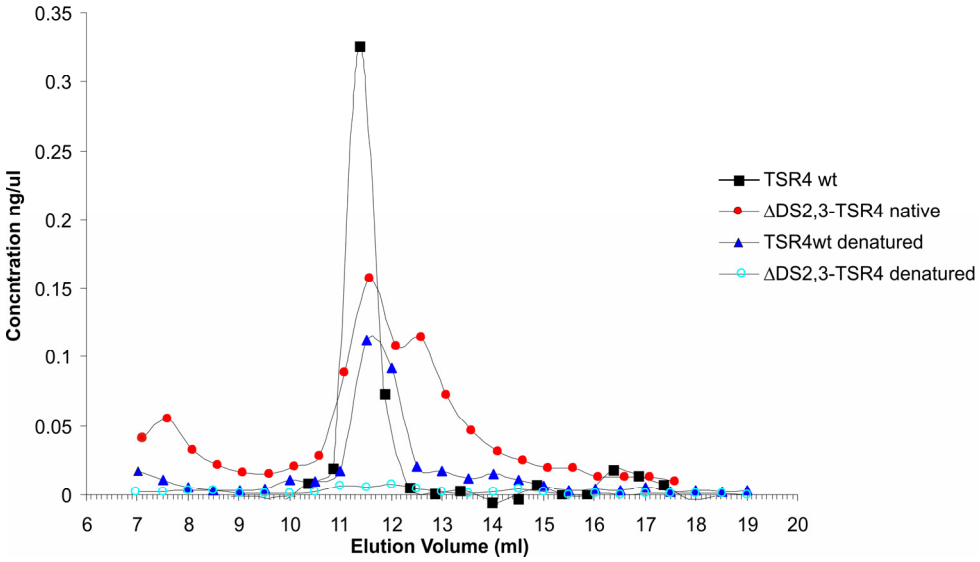


Figure 10: Analytical gel filtration of wild type TSR4 and ΔDS2,3-TSR4.

Gel filtration experiments of native and denatured TSR4 and ΔDS2,3-TSR4. 10 µg of TSR were loaded onto a Superdex 75 100 300/GL column and fractions from 7 to 17.5 ml elution were analyzed by ELISA.

As shown in Fig. 10, the majority of ΔDS2,3-TSR4 eluted as two peaks with maxima at 11.6 ml and 12.5 ml. Additionally, a minor third peak eluted at the void volume (7.59 ml) indicating the presence of a large multimer. No significant peak could be measured for denatured ΔDS2,3-TSR4. Either the protein was lost during the reduction, carboxyamidomethylation, or during the gel filtration experiment in the FPLC system. In contrast, wild type TSR4 eluted as a sharp peak having a maximum at 11.4 ml. TSR4 with denatured and modified cysteins eluted at 11.5 ml. Unlike with ΔDS2,3-TSR4,

denaturation of the protein reduced the amount of recovered wild type TSR4 only to about 3.5-fold. Based on the equations obtained for the plot of K_{av} versus molecular weights of standards (Figure 6A) and √–log K_{av} versus Stokes radii of the standards (Figure 6B), the molecular weight and Stokes radii of the major TSR4 species were calculated (Table 4).

Sample Name	Ve (ml)	Ve/V0	Kav	Calculated Stokes radius (Å)	Calculated Molecular Weight (kDa)
TSR4 native	11.4	1.49	0.23	22.54	28.44
TSR4 denatured	11.5	1.5	0.24	22.12	27.22
ΔDS2,3-TSR4 peak 1	11.6	1.52	0.24	21.7	26.05
ΔDS2,3-TSR4 peak 2	12.5	1.63	0.3	18.26	17.56

Table 4: Determined and calculated parameters of wild type TSR4 and ΔDS2,3-TSR4 in gel filtration experiments

The molecular weight of the various TSR4 forms was calculated using the equation y=0.1395Ln(x)+1.6602 determined from the plot K_{av} versus log molecular weight. The Stokes radius was determined by the equation y=0.171x + 0.4143 from the plot √–log K_{av} versus Stokes radii of standard proteins (see Chapter 3.13)

The calculated molecular weights for wild type TSR4 correspond rather to a dimer than to a monomer. Similarly the calculated molecular weight for peak of ΔDS2,3-TSR4 also suggested the presence of a dimer in solution (see discussion part). The Stokes radius of ΔDS2,3-TSR4 was calculated to be similar or even smaller than the Stokes radius of wild type TSR4, indicating that ΔDS2,3-TSR4 does not fold as an molten globule.

Discussion:

The ΔDS2,3-TSR4 protein was found to express a tetrasaccharide termed disialyl-T-antigen on amino acid Thr-601. In contrast to wild type TSR4, ΔDS2,3-TSR4 contains only one disulfide bridge. It could be possible that such a deletion of cysteines results in a misfolded protein which is predominately degraded and only secreted at minor amounts. Neither in ELISA of the native protein (see Figure 8) nor in SDS polyacrylamide gels (view figure 9), a decrease in secretion could be demonstrated. The ΔDS2,3-TSR4 protein has adopted a structure, which is not recognized by POFUT2 as a potential substrate for *O*-fucosylation. POFUT2 was shown to *O*-fucosylate only properly folded TSRs (*43*). Proteins, which contain a mutation of the O-fucosylated residue, show

decreased secretion (*63, 64*). It was proposed that POFUT2, similar to POFUT1, acts as a chaperone. Unlike POFUT1, however, the chaperone activity of POFUT2 is mediated by its glycosyltransferase function (*63*). Addition of *O*-fucose on TSRs was proposed to provide a signal for proper folding. Absence of this signal was suggested to result in poor secretion. ΔDS2,3-TSR4 seems to escape this quality control mechanism. Apparently, it is not recognized as a misfolded protein and absence of *O*-fucosylation does not lead to decreased secretion. In this context, it would be interesting to determine the structural requirements for recognition by POFUT2. For *O*-fucosylation the recognition motif was described to be $C^1X_{2-3}S/TC^2X_2G$. Studies from our laboratory indicated that replacement of either of the cysteines to alanine (but not both as in ΔDS2,3-TSR4) still leads to substantial *O*-fucosylation (D. Klein unpublished). Currently, it is not known whether these replacements result in significant distortion of the three-dimensional fold. However, the data suggests that POFUT2 can tolerate some changes in TSRs and still recognize it as a properly folded molecule that can be *O*-fucosylated.

The fact that POFUT2 does not *O*-fucosylate ΔDS2,3-TSR4 indicated that it has not adopted the native TSR conformation. One possibility was that ΔDS2,3-TSR4 might be a molten globule. One way to detect the molten globule state is the comparison of the hydrodynamic dimensions and the compactness of ΔDS2,3-TSR4 and wild type TSR4. Both proteins were examined by analytical gel filtration (Figure 10). The molecular weights calculated for wild type TSR4 and for ΔDS2.3-TSR4 corresponded rather to a dimer than to a monomer. Calibration of the gel filtration column was performed with globular standard proteins. It has to be kept in mind that a TSR4 is not a globular but an elongated protein (see Figure 7). Thus, the molecular weight of TSR4 could be over-estimated. The significance of the Stokes radius has to be taken with a similar caution. Based on the solution structure of TSR4 from F-spondin determined by Paakonen and coworkers (*116*) the dimensions along the longitudinal and vertical axis can be estimated using the program Coot (*http://www.ysbl.york.ac.uk/~emsley/coot/*). This resulted in 56 Å along the longitudinal axis and 12 Å along the vertical axis. These values vary considerably from the calculated Stokes radius obtained from gel filtration experiments. Since the Stokes radius reflects the compactness of a molecule, it is

possible that TSR4 is not very compact. This could explain why denaturation of wild type TSR4 may not change too much the Stokes radius.

Examination of native ΔDS2,3-TSR4 in analytical gel filtration suggested, that the protein displays significant more heterogeneity than native wild type TSR4. Denatured ΔDS2.3-TSR4 could not be detected. The fact that native ΔDS2,3-TSR4 can be analyzed by gel filtration, but not denatured ΔDS2,3-TSR4, strongly indicates that the former has some degree of structure.

Apart from a large peak at the void volume, two major peaks were detected for native ΔDS2,3-TSR4. It is possible that these two peaks result from the difference in glycosylation observed for ΔDS2,3-TSR4. The molecule was shown to be glycosylated with the disialyl-T-antigen to about 50%. It is possible that the glycosylated portion of the protein interacted differently with the column matrix than the un-glycosylated ΔDS2,3-TSR4. This hypothesis is supported by the fact that an elution buffer with relative low ionic strength was used that could facilitate protein-matrix effects.

In order to further characterize the structure, circular dichroism (CD) measurements on ΔDS2,3-TSR4 could be made and compared to the known CD spectrum of wild type TSR4 (*65*). A potential folding intermediate could also be investigated using the fluorescence of ANSA (*8-Anilino-1-naphthalene Sulfonic Acid)* as a readout. An increase in ANSA fluorescence has been reported to be a criterion to distinguis folded, molten globule and unfolded states (*99*).

4.3.2 Characterization of the disialyl-*T-antigen on ΔDS2,3-TSR4*

Results:

Expression of the disialyl-T-antigen in other cells

The presence of the disialyl-T-antigen on ΔDS2,3-TSR4 was originally discovered in HEK 293T cells (see Chapter 4.2). The modification could be dependant on intrinsic properties of the used cell line or the protein itself. Therefore, it was of interest to determine whether the modification of ΔDS2,3-TSR4 with the disialyl-T-antigen would also occur in other mammalian cell lines. ΔDS2,3-TSR4 was transiently transfected into the hamster cell line CHO-K1, the African green monkey cell line COS 7 and into HEK

293T cells as a control. The proteins were purified as described, digested with Lys-C and analyzed by LC-MS. ΔDS2,3-TSR4 from CHO-K1 cells was hardly modified with the disialyl-T-antigen (see Chapter 4.2), precluding quantification using the A_{280}. In contrast, ΔDS2,3-TSR4 from both HEK 293T and COS 7cells was heavily modified with the disialyl-T-antigen (45% and 54%).This allowed the reliable quantification of the various glycoforms of peptide K2* (see Fig. 11)

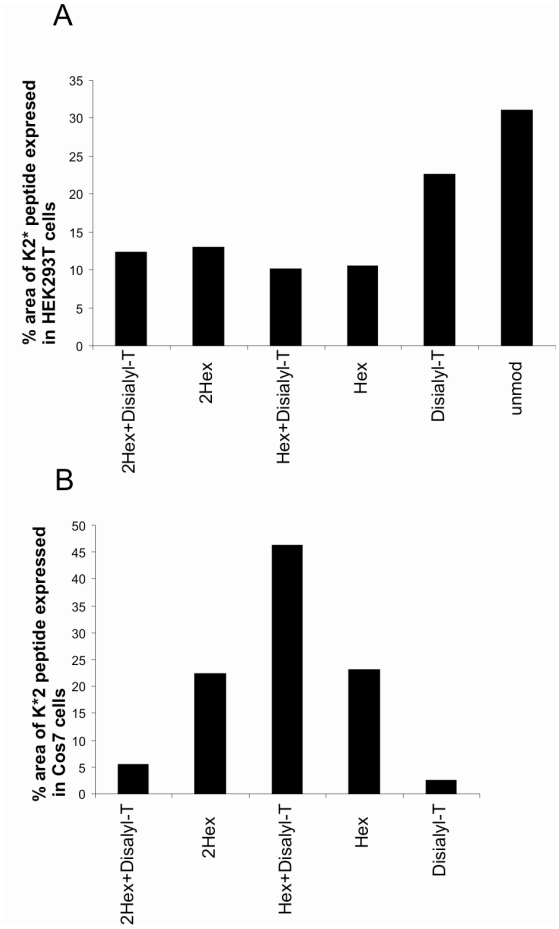


Figure 11: Quantification of the various glycoforms of peptide K2* from ΔDS2,3-TSR4 expressed in HEK 293T or COS 7 cells
ΔDS2.3-TSR4 was expressed in HEK 293T cells (A) and in the African green monkey cell line Cos 7 (B). The protein was purified digested and purified on a HPLC-MS using a C18 reverse phase column. The A_{280} area of various glycoforms of peptide K2* was normalized for two and one *C*-mannosyl tryptophans as described.

Both cell lines were capable to modify the ΔDS2,3-TSR4 protein with the disialyl-T-antigen. These results demonstrated that both cell lines were capable to modify ΔDS2,3-TSR4 with the disialyl-T-antigen. Therefore, the glycosylation pattern of ΔDS2,3-TSR4 results from intrinsic features of the structure or amino acid sequence of the proteins. Differences in the relative amount of each glycoform in COS 7 cells and in HEK 293T cells may result from a variation in the cellular repertoire of involved enzymes, or sugar donors (see discussion chapter 6). Remarkably, we only observed the complete tetrasaccharide and never biosynthetic intermediates.

Analysis of residues required for glycosylation:

The biosynthesis of the disialyl-T-antigen is initiated by the transfer of a GalNAc residue from UDP-GalNAc to the Ser or Thr amino acid chain. Unlike for protein *N*-glycosylation no general recognition sequence for *O*-glycosylation has been defined. The main reason for this is that there exist at least 16 experimentally validated GalNAc transferases, many of which are thought to have their own unique substrate specificity (*28, 132*). Only recently, has insight into this problem been obtained from *in vitro* studies with purified, recombinant transferases and synthetic peptide substrates. For example it was shown with purified GalNAc transferases and a peptide library that GalNAc T1 has the preferred consensus sequence -(F/D)(F/A)(P/V)TP(G/A)P-. In contrast GalNAc T2 was demonstrated to prefer the consensus sequence -(P/I)GPTPGP- (*26*). Interestingly, GalNAc transferases have been described that display activity against already glycosylated peptides. Such a glycopeptide activity has been observed for GalNAc T4 (*124*) GalNAc T7 (*133*) and GalNAc T10 (*134*).

In an attempt to narrow down the possible GalNAc transferases that are involved in the modification of ΔDS2,3-TSR4, we have studied the importance of the residue at position -1 for the synthesis of the disialyl-T-antigen. The Val residue was mutated into Pro, Ala, and Lys and the recombinant protein was expressed in HEK 293T cells. To determine the level of modification with the disialyl-T-antigen, an endoproteinase Lys-C peptide map of the purified protein was made using LC-MS as described in Chapter 4.2. The relevant glycoforms of peptide K2* were quantified from their absorbance at 280 nm and the levels of glycosylation have been summarized in Fig. 12.

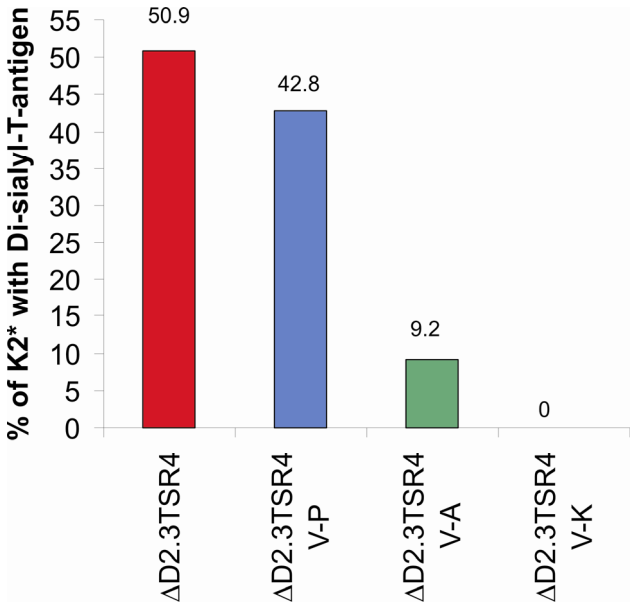


Figure 12: Quantification of the amount of disialyl-T-antigen on ΔDS2.3-TSR4

In the mutants the residue at position -1 was replaced from Val to Pro, Ala or Lys. The amount of disialyl-T-antigen was quantified based on the UV_{280} signal of the K2* peptides.

The results show that Val at position -1 results in the highest level of glycosylation (50.9%). Interestingly, it was found that ΔDS2.3-TSR4 with Pro was nearly equally well modified (42.8%), followed by Ala with 9.2%. The analysis of the protein with the Val-Lys substitution was slightly more complicated, because the mutation created an additional protease cleavage site Therefore, the ΔDS2.3-TSR4 V-K mutant was cleaved with Asp-N. In both digests, no disialyl-T-antigen could be detected, indicating that positive charged residues at position -1 destroy the capability of a GalNAc transferase to glycosylate the module.

Detailed analysis of peptides from in vitro glycosylation experiment:

The peptide K* derived from the sequence of ΔDS2,3-TSR4 was demonstrated to be glycosylated *in vitro* by GalNAc T1 and GalNAc T3. The MS data *in vitro* suggested that GalNac T3 only modified Thr-601, whereas GalNAc T1 modified both Thr-601 and Ser-596. Since the *O*-linked GalNAc residue is easily lost in tandem MS experiments, we

wanted to strengthen these results by isolating the relevant peptides. The glycosylated peptides from the *in vitro* assay were digested with endoproteinase Asp-N. This resulted in an N-terminal peptide K*D1 and a C-terminal peptide K*D2 (Figure 13A). The digests were fractionated by LC-MS using a C18 column. The de-clustering potential in the MS was varied between 25 and 45 V to visualize the neutral loss of 203 Da as an indicator for the GalNAc moiety. The UV chromatograms for K* glycosylated with GalNAc T1 and GalNAc T3 are shown in Fig.13B and Fig.13C, respectively. The peptide peaks were numbered according to the order of appearance in the chromatogram and their assignment has been summarized in Table 5.

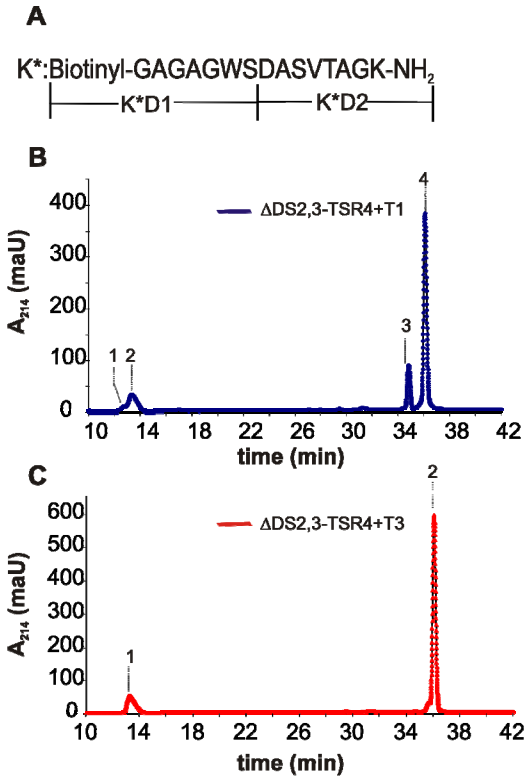


Figure 13: UV chromatograms of peptide ΔDS2,3-TSR4 glycosylated with GalNAc T1 and T3

(A) Peptide ΔDS2,3-TSR4 (K*) glycosylated by GalNAc T1 or GalNAc T3 was digested with Asp-N resulting in peptide K*D1 and K*D2. Subsequently the digest was measured by LC-MS. (B) Chromatogram of ΔDS2,3-TSR4 glycosylated with GalNAc T1. (C) Chromatogram of ΔDS2,3-TSR4 glycosylated with GalNAc T3.

ppGalNAc	Peak	Elution time	mass detected		mass expected		Assignment
		min	MH+	M2H+	MH+	M2H+	
T1	1	12.85	747.6		747.83		K*D2
	2	13.4	950.4		951.2		K*D2+GalNAc
	3	34.85	1034.4		1035.1		K*D1+GalNAc
	4	36.11	831.2		831.93		K*D1
		35.59		881.6		882.46	K*+GalNAc
		36.66		881.6		882.46	
T3	1	12.34	950.4		951.2		K*D2+GalNAc
	2	36.1	831.2		831.93		K*D1
		35.59		881.6		882.46	K*+GalNAc

Table 5: Assignments of the peaks identified in LC-MS

K* glycosylated by GalNAc T1 represents a mixture of two glycoforms. One glycoform is modified on Ser-596 in peptide K*D1 and one on Thr-601 in peptide K*D2. Two peptides were found to co-elute with peak 4. These peptides represent the unmodified peptide K*D1 and the uncleaved peptide K* containing one GalNAc moiety. The results obtained for GalNAc T3 indicate that exclusively Thr-601 in peptide K*D2 is modified, since unmodified peptide K*D2 or modified peptide K*D1 peptides could not be detected. Similar to the results obtained for GalNAc T1, unmodified peptide K*D1 and undigested co-eluting glycosylated ΔDS2.3-TSR4 were detected in peak 2. Extraction of the *m/z* value of the undigested material, revealed a minor peak at 35.59 min and the major part eluting at 36.66 min for the sample treated with GalNAc T1. Interestingly, this peptide eluted predominantly at 36.59 in the sample glycosylated with GalNAc T3, indicating that the localization of the modified residue had an influence on the elution behavior of the ΔDS2.3-TSR4.

Discussion:

Modification of ΔDS2,3-TSR4 with the disialyl-T-antigen:

The modification of ΔDS2,3-TSR4 with NeuNAcα2,3Galβ1,3[NeuNAcα2,6]GalNAc-*O*- requires a complex interplay of enzymes located throughout the Golgi apparatus (*32*). It is known that the involved enzymes modify the substrate in a sequential process. The first enzyme the GalNAc transferase is of particular importance because it needs to recognize the polypeptide chain directly. It is generally assumed that the primary sequence is the major factor that defines substrate specificities for GalNAc transferases (*135*). In a recent peptide screen for the requirements of GalNAc T1 and T2, positive enhancement factors were defined for hydrophobic residues at -1 and + 3 relative to the glycosylated residue (*26*). In order to study the structural requirements for enzyme substrate recognition the Val at position -1 of Thr-601 in ΔDS2,3-TSR4 was replaced by Pro, Ala and Lys (Figure 12). Quantification of the glycoforms modified with the disialyl-T-antigen revealed that the level of modification decreased in the order Val>Pro>Ala>Lys. These results suggest that large hydrophobic residues like Val or Pro positively influence mucin-type *O*-glycosylation in ΔDS2,3-TSR4. A Val rather than a Pro at position -1 was found to be favored for *O*-glycosylation. Small uncharged residues like Ala at position -1 seem to play already an inhibitory role, since only about $^1/_5$ of the glycoforms receive the disialyl-T-antigen. Introduction of a Lys at position -1 blocked the modification of Thr-601 with the disialyl-T-antigen completely. This result is consistent with earlier reports that charged residues at position -1 and +3 have an inhibitory effect on *O*-glycosylation (*136*). ΔDS2,3-TSR4 has a positively charged Lys residue at position +3. Therefore, it would be interesting to determine whether this residue already exerts a negative effect on ΔDS2,3-TSR4 or ΔDS2,3-TSR4 V-P and whether the total amount of disialyl-T-antigen could be increased by replacing this residue with a hydrophobic residue.

Using ΔDS2,3-TSR4-derived peptides, it was found that mucin type *O*-glycosylation can be initiated by GalNAc T1 and GalNAc T3 (see Chapter 4.2 or figure 13). GalNAc transferase 1 was found to glycosylate both Ser-596 and Thr-601 after 24h (Chapter 4.2 and table 5). GalNAc transferase 3 was shown to glycosylate solely Thr-601 (Chapter 4.2, Figure 13 and table 5). This transferase could also be confirmed *in vivo* in co-

expression experiments in ChO-K1 cells (Chapter 4.2 figure 8). These results indicate that in CHO-K1 cells ΔDS2,3-TSR4 can be efficiently glycosylated by GalNAc T3. Table 6 shows further peptides glycosylated by GalNAc T3 and where the site had been mapped.

Protein	Alignment	Reference
ΔDS2,3-TSR4	DASV**T**AGKG	this study
HIV gp120 protein	RGPGRAFV**T**IGKIGNMR	(*132*)
Muc1a	AHGV**T**SAPDTR	(*137*)
Erythropoietin	PPDAA**T**AAPLR	
Fibronectin	Ac-PFV**T**HPGYD	(*28*)
FGF23	PIPRRH**T**RSAEDDSERDP	(*35*)

Table 6: Peptides glycosylated by GalNAc T3

Peptides taken from literature: Only those peptides with mapped sites were considered. The glycosylated residue is highlighted in bold.

The peptides glycosylated by GalNAc T3 show some similarities. Apart from the FGF23 peptides, all selected peptides contain a Val residue at position -1 and small uncharged residues at position +1 and +2 are frequently observed. GalNAc T3 has been implicated in familial tumoral calcinosis (*29*) and secretion of FGF23 was shown to be dependent on a specific GalNAc T3 site (*35*). Additionally, GalNAc T3 has been described to be over expressed in a variety of epithelial cancers of breast, colon, stomach and pancreas origin (*138*). In breast cancer, GalNAc T3 expression was found to correlate with malignancy (*139*). These results render the identification of ΔDS2,3-TSR4 as a substrate for GalNac T3 a valuable tool for further analysis of this enzyme. Interestingly, GalNAc T3 cannot be the only polypeptide GalNAc transferase that modifies ΔDS2,3-TSR4 *in vivo.* In this study the disialyl-T-antigen on ΔDS2,3-TSR4 was also identified in COS 7 cells. The kidney cell line derived from the African green monkey *Cercopithecus aethiops* was described to express no detectable transcript of GalNAc T3 (*136*). Consequently, an additional GalNAc homologue must be present in COS 7 cells that glycosylate ΔDS2,3-TSR4. A potential candidate is GalNAc T6. In humans, GalNAc T3 and GalNAc T6 are highly conserved and show similar specificity but are expressed in

different tissues (*28*). Alternatively, GalNAc T1 which modifies Ser-596 and Thr-601 *in vitro* could mediate O-glycosylation of ΔDS2,3-TSR4 in COS 7 cells.

In summary, the data provides compelling evidence that ΔDS2,3-TSR4 can be glycosylated *in vitro,* and *in vivo* in the hamster cell line CHO-K1 by human GalNAc T3. Glycosylation seems to be strongly dependent on the Val at position -1, which is also present in a variety of other specific GalNAc T substrates. Data from COS 7 cells, suggests however that this enzyme is not the only one capable of glycosylating ΔDS2,3-TSR4.

Potential functions of the disialyl-T-antigen:

The synthesis of disialyl-T-antigen requires the successive action of four different enzymes that modify the substrate. As described, GalNAc T3 is one possible candidate out of at least 16 different GalNAc Ts that can initiate *O*-glycosylation. Subsequently the T- antigen is generated by transfer of Gal by the Core 1 β1,3 galactosyltransferase. The sialic acid in α2,3 linkage onto Gal can be transferred by ST3Gal-I,-II or IV. The sialic acid in α2,6 linkage onto the GalNAc can be transferred by ST6GalNAcI-IV (*37*). This transfer has to occur after the addition of galactose, since the NeuNAc moiety α2,6-linked to GalNAc is inhibitory for the Core 1 β1,3 galactosyltransferase (*32*). Thus the transfer of the disialyl-T-antigen tetrasaccharide onto ΔDS2,3-TSR4 requires a complex interplay of glycosyltransferases in the right order. Interestingly, so far always only the disialyl-T-antigen and not other core structures or shorter glycans were observed on ΔDS2,3-TSR4. The observation that the disialyl-T-antigen is also generated in cells different from HEK 293T indicates a specific function of the glycan.

As indicated in table 7 the presence of sialyl-T-antigen is observed in various proteins ranging from surface receptors to secreted proteins. The function of these proteins is often determined by lectin binding. On receptors such a binding can lead to specific activation of downstream molecules (see references (*25, 140*) as an example). On secreted proteins like podoplanin the sialic acids on the disialyl-T-antigen are critical for the platelet-aggregation activity (*24*). Plasminogen 2 contains a sialyl-T-antigen on Thr-345 and on Ser-248. The sialic acids of this protein have been implicated in decreasing the catalytic efficiency of tissue-type plasminogen activator and urinary-type

plasminogen activator (*141*). These results demonstrate that especially the negatively charged sialic acids can mediate or decrease binding to other proteins.

So far the disialyl-T-antigen was observed on ΔDS2,3-TSR4. This raises the question whether wild type TSR4 also undergoes the same modification under specific cellular conditions that affect disulfide formation. Correct disulfide bridge formation is dependent amongst other things on availability of O_2 and on disulfide isomerases. In order to study these effects on glycan expression of wild type TSR4, transiently transfected HEK 293T cells were incubated under hypoxic conditions or with the protein disulfide isomerase inhibitor bacitracin for 48 h. The protein was purified and the glycosylation pattern was analyzed by mass spectrometry as described before. In both cases, glycosylation of wild type TSR4 was not changed. These preliminary experiments were hampered by the fact that positive controls were missing. Further analysis is required to see if the glycosylation could be “induced” on wild type TSR4.

Another possible function of the glycan on ΔDS2,3-TSR4 would be that the disialyl-T-antigen stabilizes the module and facilitates secretion. In order to verify that hypothesis secretion of the ΔDS2,3-TSR4 mutants was measured. However, it was observed that the Val-Lys mutant exhibited increased secretion levels in HEK 293T cells compared to ΔDS2,3-TSR4 or the Val-Pro mutant (not shown).These results argue against a stabilizing function of the glycan.

Protein	Reference	Proposed Function
Podoplanin	(*24*)	Necessary for platelet aggregation
von Willebrand factor	(*142*)	Inverse relation between PNA binding and healthy individuals compared to liver cirrhosis or von Willebrand disease
Plasminogen 2	(*141*)	Trisaccharide *O*-glycan located between kringle 2 and 3 proposed to disturb the interaction between plasminogen and tPA[a]
CD45 (protein tyrosine phosphatase)	(*140*)	Induction of IL2 production by Jacalin[b] binding
alpha-dystroglycan	(*143*)	Inhibition of lamin-induced AChR[c] clustering by Jacalin
human milk bile salt-activated lipase (BAL)	(*144*)	May contribute to adhesive activity in the physiological function of BAL.
CD44v6,	(*25*)	PNA stimulates colon cancer cells by interaction with C-met

a: Tissue specific plasminogen activator

b: Lectin from the Jackfruit, *Artocarpus integrifolia* binds to sialylated T-antigens

c: Acetylcholine receptor

Table 7: Proteins modified with sialyl-T-antigen

5 Identification of an *O*-glycan in the reespo domain of F-spondin

5.1 Introduction

The axonal guidance molecule F-spondin has been shown to be important for axonal guidance. Studies have revealed that the molecule is highly glycosylated on its TSRs (*61*). This study extends the knowledge of glycosylation of this protein. It will be shown for the first time that the tetrasaccharide NeuNAcα2,3Galβ1,3[NeuNAc]GalNAc-*O* is linked to a residue located in the N-terminal reelin/spondin (reespo) domain of rat F-spondin (P35446). The presentation of the results will be preceded by a brief introduction about the protein.

5.1.1 The axonal guidance protein F-spondin

F-spondin is an extracellular matrix (ECM) protein highly expressed in the floor plate during early spinal cord development. It contains 807 amino acids and has an apparent molecular weight of 116 kDa in SDS polyacrylamide gels (*145*) (Figure 14).

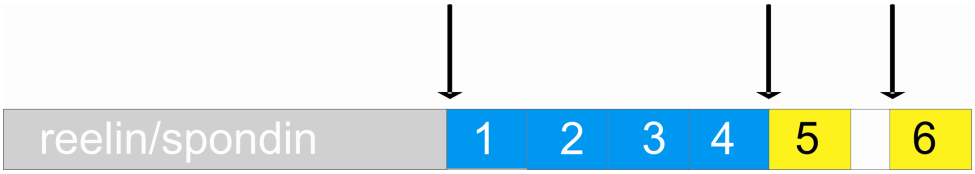
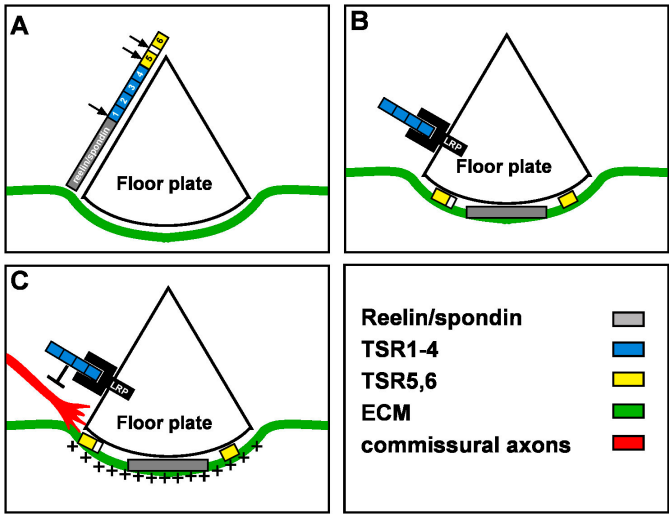


Figure 14: Graphical illustration of vertebrate F-spondin
Vertebrate F-spondin consists of an N-terminal reelin domain, a spondin domain six class two thrombospondin type 1 repeats (TSRs). Three consecutive proteolytic events (indicated with arrows) release the reelin/spondin domain (grey box), TSR1-4 (blue box) and the basic TSRs TSR5 and TSR6 (yellow boxes).

F-spondin contains two potential *N*-glycosylation sites, TSR1-5 have been shown to be *C*-mannosylated and TSR1-4 contain the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr (*61*). Proteolytic processing has been described to generate a reelin/spondin fragment (reespo) of about 60 kDa. Additionally, F-spondin can be cleaved by the serine protease plasmin between TSR5 and TSR6 and within TSR5. Upon cleavage by plasmin, F-spondin loses its capacity to bind to the ECM, which is presumably mediated through

basic domains in the second and third antiparallel strand of TSR5 and TSR6. This releases a soluble protein containing TSR1-4 (*146*). F-spondin has been shown to promote neurite outgrowth of neuronal subpopulations like commissural neurons (*147*), hippocampal neurons (*148*) or sensory neurons like dorsal root ganglions (*149*). Depending on cellular context, different domains of F-spondin were shown to promote neurite outgrowth. Sensory neuron outgrowth is promoted by the reespo domain. Outgrowth of hippocampal neurons were only promoted by TSR1-4. These domains were demonstrated to be inhibitory for commissural neurons (*150*). Therefore, it was proposed that F-spondin mediates negative and positive cues depending on cellular context and cell (*151*). In the floor plate, F-spondin is involved in the turning of commissural neurons in the basement membrane that underlines the floor plate. The TSR1-4 fragment has been shown to bind to Apolipoprotein E receptor (ApoEr2) (*152*) and other members of the low density lipoprotein receptor-related protein family (e.g. megalin, LPR4) and to exert repulsive cues on commissural neurons. This leads to a squeezing of the axon between floor plate and basement membrane. At the same time, the reespo domain, together with the plasmin-released TSR5 and TSR6, bind to the ECM of the floor plate and exert attractive cues on the commissural axon (Figure 15) (*150*).

Comment [m2]: Is it apolipoprotein?



Adapted from (*150*)

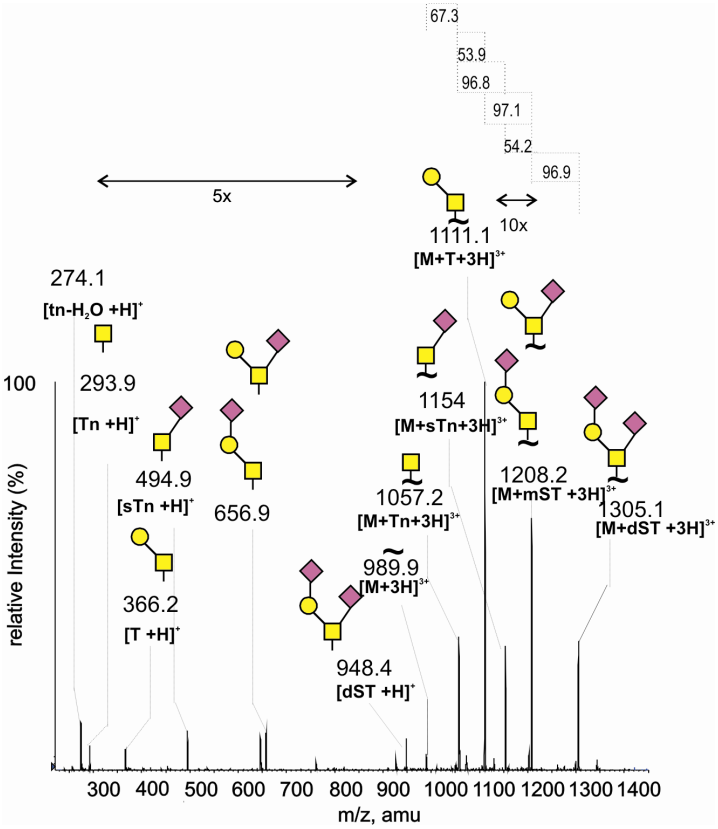
Figure 15: Proposed function of F-spondin in the floor plate during embryonic development

F-spondin is secreted from the floor plate and cleaved between the reelin/spondin domain and by a serine protease within TSR5 and between TSR5 and TSR6. (A) The released soluble TSR1-4 binds to receptors of the LRP family. TSR5 and TSR6 and the reelin/spondin domain bind the ECM of the basement membrane underlining the floor plate (B). Commissural axons are repelled by TSR1-4 and attracted by TSR5, TSR6 and the reelin/spondin domain. This leads to a squeezing of the axon between floor plate and basement membrane and prevents lateral drifting as soon as the commissural neuron has crossed the floor plate.

5.2 Results

It has been shown that rat F-spondin (P35446) expresses *C*-mannosyl tryptophans on its TSRs and the disaccharide Glcβ1,3-Fucα-*O*- on serines or threonines (*61*). Preliminary results from the laboratory indicated that F-spondin produced in COS 7 might be modified with another harbor glycan. Two tryptic peptides at m/z 1304.8 and 1348 were found to exhibit glycan specific fragments in LC-MS (D.Klein unpublished). Edman sequencing of the triply charged peptide at *m/z* 1304.8 demonstrated that it corresponded to Lys-C peptide K9 (^{153}LCEQDPTLDGVTDRPILDCCACGTAK^{178}). Residues 153-166 are located in the reelin domain and residues 167-178 represent the N-terminus of the spondin domain. The peptide with *m/z* 1348 was found to correspond to the same peptide containing a uncleaved Lys residue at the N-terminus (designated

K8-9). The observed mass of peptide K9 (3911.4 Da) and peptide K8-9 (4041 Da) was approximately 947 Da higher than expected from the amino acid sequence. This together with the observed glycan-specific fragmentation strongly suggested that a disialyl-T-antigen was attached to this part of the molecule. In order to examine this, F-spondin was produced in COS 7 and in HEK 293T cells, purified and digested with endoproteinase Lys-C. The relevant peptides (m/z *1305* and *1348)* were identified in LC-MS experiments. In source decay experiments at a de-clustering potential of 120 V demonstrated successive losses of *m/z* 97.2, 54, 98.2, 54.8 and 66.8 (not shown), confirming the initial observation that these peptides are glycosylated. Due to overlapping losses of the precursor ions at m/z 1305 and m/z 1348, the spectrum contained various unidentified peaks. First fragmentations at high CE were performed which confirmed the identity of the peptide as K9 and K8-9 respectively (not shown). To obtain sequence information on the structure of the glycan, tandem MS experiments at relatively low CE (30V) were carried out. As expected we observed a stepwise fragmentation of the glycan in a pattern very similar to that of ΔDS2,3-TSR4 (see Fig 3 Chapter 4.2). The tandem MS fragmentation pattern of the glycan was identical for K9 and K8-9 peptides from F-spondin expressed in COS 7 and HEK 293T cells (not shown). In figure 16 the glycan fragmentation of peptide K9 from COS 7 cells is illustrated. The peptide K8-9 exhibited a similar fragmentation pattern (not shown).



Abbreviations: dST: Disialyl-T-antigen; mST: Monosialyl-T-antigen; sTn: Sialyl-Tn-antigen; T: T-antigen; Tn:Tn-antigen

Figure 16: Tandem MS at low collision energy of Lys-C peptide K9 of rat F-spondin

Zoomed ions are indicated with a double-tipped arrow. Peptide-linked carbohydrates are indicated with "~". The triple charged precursor ion at *m/z* 1305.1 experiences sequential losses of a NeuNAc (loss of 96.9) resulting in the ion at *m/z* 1208.2, presumably corresponding to the sialyl-T-antigen. Simultaneous losses of a hexose (54.2) or a NeuNAc (97.1), result in a triple charged peptide plus a T-antigen (m/z or 1111.1) or a sialyl-Tn antigen (*m/z* 1154). These ions lose another hexose (53.9) or NeuNAc (96.8) resulting in a fragment containing the Tn-antigen (*m/z* 1057.2). Finally, a loss of a HexNAc yields the unmodified peptide at m/z 989.9. The only known structure consistent with the spectrum is the structure of a disialyl-T-antigen (NeuNAc-Hex-[NeuNAc]-HexNAC-*O*). In parallel to the peptide-linked glycan, also the glycan (fragment ion *m/z 948.4)* undergoes fragmentation. These ions exhibit an identical fragmentation pattern as observed for the triple charged peptide-linked glycans

The fragmentation pattern was completely consistent with the presence of a disialyl-T-antigen. Unfortunately, the modified amino acid could not be identified, because it was not possible to establish conditions under which glycosylated fragments were generated

that yield information on the amino acid sequence. In order, to further confirm the sugar identity and their conformation, the glycosylated peptides were digested with endo- and exo-glycosidases using a very similar approach as described in Chapter 4.2. Peptides K9 and K8-9 were isolated from F-spondin produced in COS 7, neutralized and digested with neuraminidases. Probably due to the additional positive charge at its N-terminus peptide K8-9 was found to display better ionization efficiency in LC-MS than peptide K9. Therefore, the data obtained with this peptide are presented here, although peptide K9 gave very similar results. Digestion with neuraminidase from *Clostridium perfringens* resulted in a peptide with a mass of 3459 Da (Fig. 17A) indicative of removal of both sialic acids (Figure 17A). Full-length F-spondin was digested with this neuraminidase in the same way, and analyzed by western blot, using the Core 1-specific lectin, PNA. F-spondin only bound the lectin when the sialic acids had been removed, confirming that the presence of the core 1 structure (Galβ1,3-GalNAc-*O*), which in the untreated peptide is masked by two sialic acids (Figure 17B).

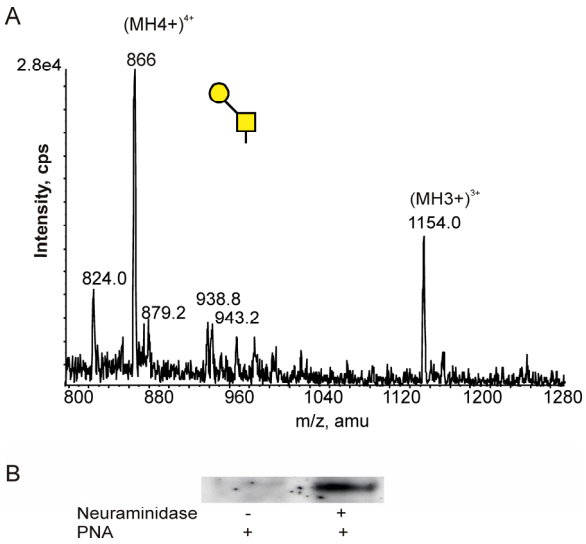


Figure 17: Digestion of peptide K8-9 with neuraminidase from Clostridium perfrigens and immunoblot of F-spondin before and after treatment with neuraminidase.

(A) K8-9 was purified by HPLC and digested with neuraminidase from *Clostridium perfringens*. Subsequent LC-MS revealed that two sialic acids had been removed and only the peptide backbone and a disaccharide remained. (B) The disaccharide was determined to be Galβ1,3-GalNAc-*O* with the core 1-specific lectin peanut agglutinin (PNA) on an immunoblot of full-length F-spondin.

Comment [MM3]: Abbreviations are enough in legends

In order to determine the linkage position of the sialic acids peptide K8-9 was digested with α2,3 neuraminidase. This yielded a peptide with a mass of 3749.4 Da. This is consistent with the peptide K8-9 containing a sialyl-T-antigen (Figure 18) and results prove the presence of a Core-1 glycan with α2,3-linked sialic acid on peptide K8-9.

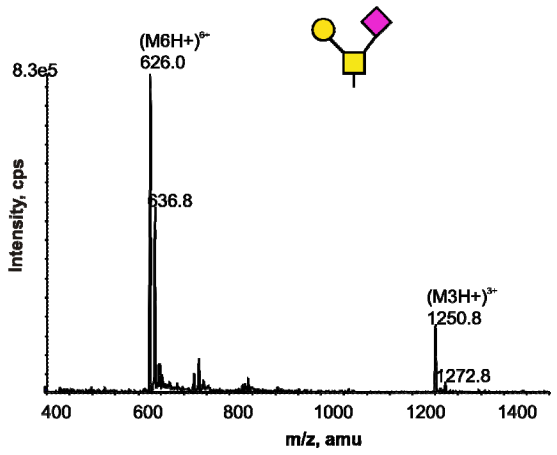


Figure 18: Digestion of peptide K8-9 Da with α2,3 Neuraminidase

Peptide K8-9 was digested with a neuraminidase specific for the α2,3 linkage. This yielded a peptide in LC-MS, which was found to contain a trisaccharide, confirming the presence of a α2,3-linked sialic in K8-9

In summary, the tandem MS spectra and the digestion with neuraminidases proved the existence of the NeuNAcα2,3Galβ1,3[NeuNAc]GalNAc-*O*- tetrasaccharide on F-spondin in the reespo domain. The exact amino acid residue remains to be determined.

5.3 Discussion

These results firmly establish the presence of a disialyl-T-antigen in the reespo domain of rat F-spondin. Besides *C*-mannosylation and *O*-fucosylation (*61*), this demonstrates another glycosylation on this protein. The glycosylated peptides K9 and K8-9 were identified in F-spondin from either COS 7 or HEK 293T cells. The results suggest that this modification is due to intrinsic properties of the protein. The glycosylated peptides contain three Thr residues (Thr^{159}, Thr^{164} and Thr^{176}), which could potentially carry the tetrasaccharide. In tandem MS experiments, no prominent ion could be identified which corresponds to a modified Thr. The glycans of the disialyl-T-antigen in particular the sialic acids were demonstrated to be very labile in tandem MS experiments and showed facile fragmentation even under low collision energies. At low collision energies, the peptide backbone often stays intact and no information about the modified residue can be obtained. At higher collision energies only low abundant peptide fragments containing a Galβ1,3GalNAc-*O* or only a GalNAc-*O* can be observed. In ΔDS2,3-TSR4 the modified residue was determined in MS^3 experiments on diagnostic tandem MS ions containing a di-or monosaccharide. The reduction of peptide size by successive digestion with endoproteinase Lys-C followed by Asp-N greatly facilitated this identification. A similar approach was also tried on the glycosylated Lys-C peptides K9 and K8-9 from F-spondin. Theoretically an Asp-N digest on these peptides yields the tetrapeptides $DPT^{159}L$, $DGVT^{164}$ and the nonapeptide $DCCACGT^{176}AK$. Unfortunately, none of these peptides could be identified in static nanospray experiments. The tandem MS fragmentation of the Lys-C peptides K9 and K8-9 yielded low abundant fragments that would correspond to a modified Thr^{159} (and a modified Thr^{176}). No ions corresponding to a modified Thr^{164} were observed. The putative glycosylated ions would fit to a triple charged peptide connected with the Galβ1,3GalNAc-*O* disaccharide. Since these ions were not examined in MS^3 experiments for a loss of a hexose and a HexNAc (view chapter 4.2) it is unknown if these ions really represent a modified Thr ion or are simply background ions. Further experiments are necessary to prove the site of modification. The identification of the modified residue would also give an indication about the involved GalNAc T that recognizes the sequence. Based on the specificity of

different GalNAc T on peptides it was postulated that Pro, or Val at position – 1 relative to the glycosylated Thr/Ser and a Pro at position + 3 are preferred residues (*26, 31*). Merely, Thr^{164} fulfills these criteria having a Val (Val^{163}) at position -1 and a Pro at position +3 (Pro^{167}). Neither this amino acid nor the remaining two threonines are predicted to be modified with a mucin-type *O*-glycan using the online prediction tool NetOGlyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc-3.1/). *O*-glycosylation of F-spondin is consequently another example of a protein, which is predicted not to be glycosylated, yet is experimentally demonstrated to contain a mucin-type *O*-glycan. This illustrates the difficulty of accurately predicting *O*-glycosylation sites.

A Blastp of Lys-C peptide K8-9 against a complete database revealed a high conservation among different species. In mouse human and bovine F-spondin the threonines are conserved. Additionally, the first and the third threonines are present in F-spondin from *Xenopus laevis* and the second and third Thr residues are present in chicken and in zebrafish. Interestingly, chicken and zebrafish also have a conserved Pro^{167} at position +3 relative to the second Thr implicating this residue as a possible candidate for the glycosylation. Determination of the glycosylation status of F-spondin from different species would further answer the question whether not only the residue but also the glycosylation is conserved in evolution.

It has been described before that F-spondin mediates attractive and repulsive cues depending on tissue and domains examined. The reespo domain mediates attractive cues on commissural neurons in the developing floor plate (*150*) and in dissociated cortical neurospheres (*153*). In cultures of rat dorsal root ganglions (DRG) from embryonic day 14 (E14), addition of F-spondin resulted in neurite outgrowth in a dose dependent manner. The outgrowth could be blocked by a spondin specific antibody indicating that this effect was mediated by the reespo domain (*149*). TSRs on the other hand, were shown to be attractive (TSR5 and TSR6) as well as repulsive (TSR1-4) to commissural neurons and TSRs alone are sufficient to promote outgrowth in hippocampal neurons (*148*). The distinct effects of various domains could result from expression of different target receptors. TSR1-4 binds to receptors of the low density lipoprotein receptor-related (LRP) protein family (ApoER2, VLDLR, LRP4, megalin), which was described to mediate repulsive cues on commissural neurons in the floor

plate (*150, 152*). A possible mechanism would be the immobilization of the TSRs that they can be presented to the growth cone of the commissural neuron. The repulsive effect mediated by TSR-LRP binding could be mediated by a similar effect than the Slit-Robo repulsion on commissural neurons. Slit is an evolutionary conserved axon guidance molecule first described in Drosophila as a midline axon repellent. It binds to a receptor of the Roundabout (Robo) family expressed on commissural neurons before and during the midline crossing of the axon. Binding of Slit to the Robo receptor leads to repulsion of the commissural neuron. Slit is immobilized by heparan sulfate in the extracellular matrix providing a gradient that prevents the commissural axon to cross the midline (*154*). In analogy to the Slit-Robo system TSRs could be immobilized by receptors of the LRP family. Binding to a counter receptor expressed on commissural neurons could then mediate the repulsive effect of the TSR1-4. In combination with the attractive cues mediated by TSR5, TSR6 and the reespo domain this could lead to the squeezing of the commissural neuron under the floor plate. A receptor for the reespo domain has not been described. A potential candidate could be the myelin-associated glycoprotein MAG (Siglec-4a). This assumption is based on several lines of evidence. It was shown that this transmembrane receptor shows a high affinity for sialylated *O*-glycans including the disialyl-T-antigen. Inhibition experiments showed that MAG could be inhibited by sialyl-T-antigen or disialyl-T-antigen in the low micromolar range (*155*). MAG is highly expressed in Schwann cells, which form the myelin sheath of axons. Interestingly, it was found that expression of F-spondin was significantly upregulated in Schwann cells after axotomy (i.e. severing) of the sciatic nerve in embryonic rats. It is noteworthy that F-spondin co-localized in this case with the so-called HNK-1 carbohydrate (*149*). This structure has been described to occur on the MAG protein (*156*).

In summary, it was demonstrated that F-spondin expresses a disialyl-T-antigen on its reespo domain. This domain has been implicated in the promotion of outgrowth of various neuronal sub-populations. Currently no receptor has been described which could bind to this domain. A potential candidate could be the myelin-associated glycoprotein MAG that has a high affinity for the disialyl-T-antigen and its temporal as well as spatial expression partially overlaps with F-spondin.

6 Expression and Purification of TSRs

6.1 Introduction

TSRs are present in a variety of different proteins performing distinct functions. Currently, it is unknown how the glycosylations of TSRs are involved in these processes. In order to obtain tools to understand the function of glycosylation, TSRs of the rat axonal guidance protein F-spondin were selected for purification. These TSRs have been described to be glycosylated (*61*) and provide unique model proteins to study *C*-mannosylation and the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr. TSR4 and TSR1-4 were expressed as fusion proteins at high amounts in HEK-EBNA cells. Analysis of glycosylation reveled the presence of (C^2-Man)Trp and Glcβ1,3Fuc-*O*-Ser/Thr. The same fusion proteins were expressed in the methylotrophic yeast *Pichia pastoris*. However, only the single TSR-containing fusion protein could be successfully purified. Although this protein was demonstrated to lack glycosylation on the TSR, indirect evidence indicated that the protein is glycosylated in the Fc portion of the fusion protein.

6.2 Results

6.2.1 Analysis of Thrombospondin-Type 1 repeats

6.2.1.1 Strategy and cloning

In order to obtain TSRs with and without glycosylation, the proteins were expressed in two different species. HEK-EBNA cells were selected as an expression system capable to modify tryptophans with mannose and to synthesize the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr. The methyloptrophic yeast *Pichia pastoris* was selected to express TSRs in its non-glycosyated form. As model proteins, the fourth TSR (TSR4) and four successive repeats of F-spondin (TSR1-4) were chosen. These proteins were expressed as secreted proteins containing a tandem Fc-His_6 tag for purification preceded by a tobacco Etch Virus cleavage site (TEV). Proteins expressed in mammalian systems contained an IgK secretion signal and proteins expressed in yeast were secreted using the α-factor

prepro-signal present in the yeast expression vector pPicZα. Due to the cloning strategy, 14 additional amino acids derived from the vector pSecTAgB were added to the N-terminus (DAAQPARRARRTKL**)** of the proteins expressed in the mammalian system. This resulted in two different N-termini starting with DAA or TKL respectively. In order to have comparable proteins these 14 amino acids were also introduced in the TSRs expressed in yeast. As an example, the PCR products of TSR1-4Fchis and TSR4Fchis are illustrated in Figure 19. These products were ligated into the mammalian vector IRES-GFP-pRS5a and into the *Pichia pastoris* secretion vector pPICZα

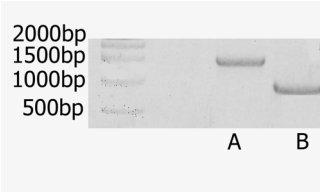


Figure 19: Fusion PCR resulting in TSR1-4fchis and TSR4fchis

TSRs were amplified from the rat F-spondin cDNA using primers adding a tobacco-etch-virus (TEV) at the 5' site. The Fc tag was amplified from human IgG using primers adding an additional His_6 sequence at the 5' end. PCR products were fused using the strategy of overlap extension as described (*157*) (A) PCR product of TSR1-4Fchis. (B) PCR product of TSR4Fchis.

6.2.1.2 Purification and analysis of mammalian expressed TSRs

To obtain stably transformed HEK-EBNA cells, expressing TSR-fusion proteins, transfected cells were selected using Zeocin. Expression levels were monitored using the GFP signal produced from the bicistronic vector IRES-GFP-pRS5a. After two weeks, Zeocin resistant cells were sorted by flow cytometry using the GFP fluorescence as readout (Figure 20).

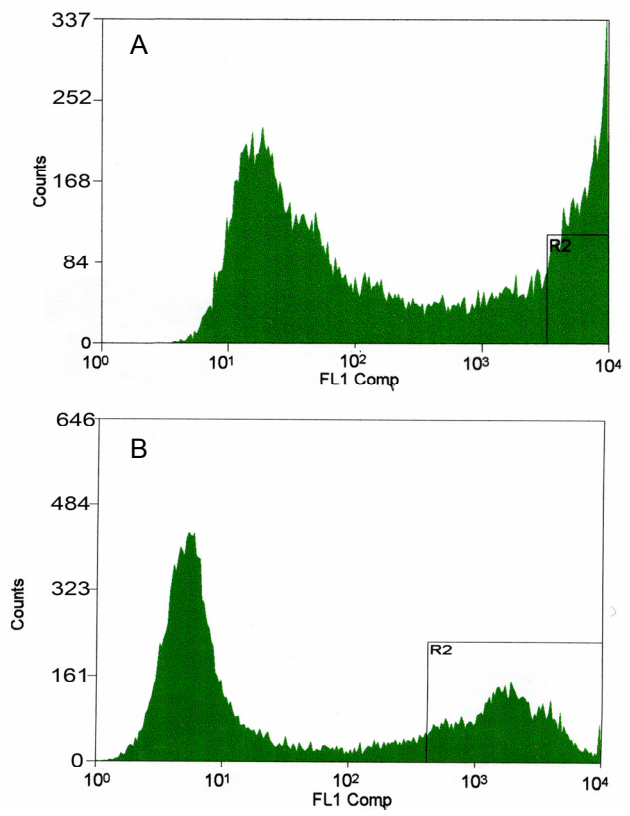


Figure 20: Cell sorting of HEK EBNA cells expressing TSR4fchis or TSR1-4fchis.

(A) Sorting of cells stably expressing TSR4fchis. (B) Sorting of cells stably expressing TSR1-4fchis. Only cells exhibiting the highest GFP fluorescence (R2) were collected.

The cells were sorted as single or multiple clones and further expanded. The single cell clones exhibited higher protein amounts of the target TSR4Fchis and TSR1-4Fchis (not shown). TSR4Fchis and TSR1-4Fchis were purified using a tandem purification strategy. In the first step, the supernatant containing the secreted proteins was loaded onto a NiNTA column, where the TSR-fusion proteins bound through their attached His_6-tag. Upon elution, relevant fractions were subjected to the second dimension. In the second dimension, proteins were purified by binding to the Fc tag using protein G beads. Purity was observed on SDS gels and on HPLC (Figure 21).

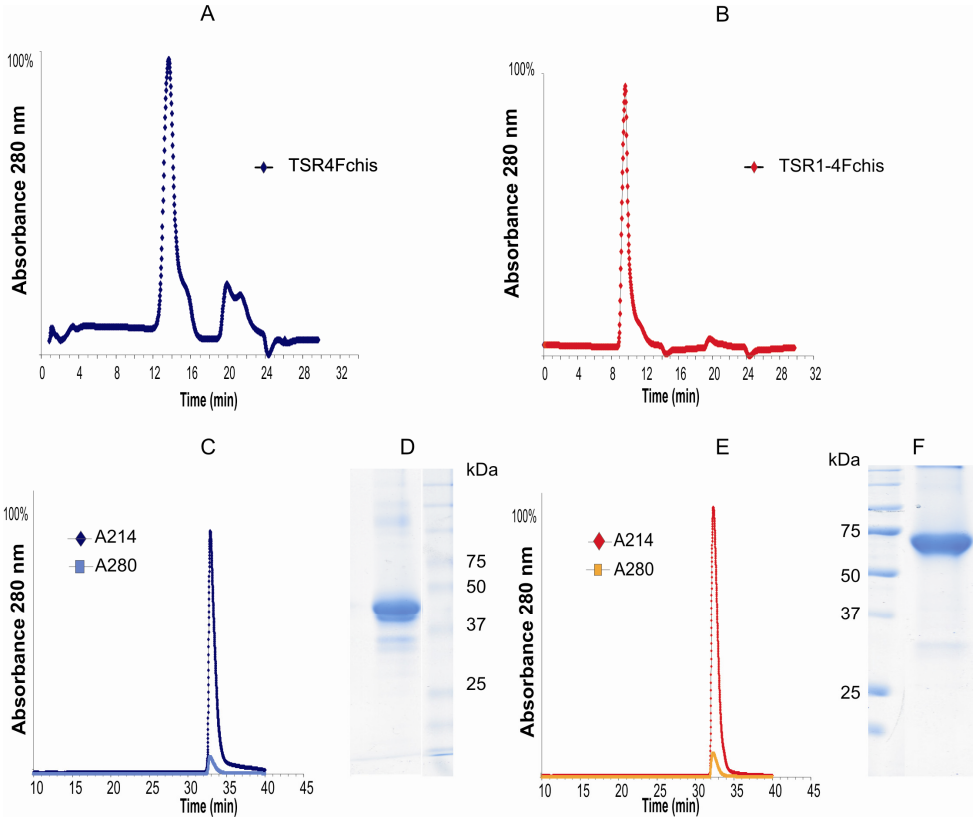


Figure 21: Two step purification approaches for TSR4fchis and TSR1-4fchis produced in mammalian cells

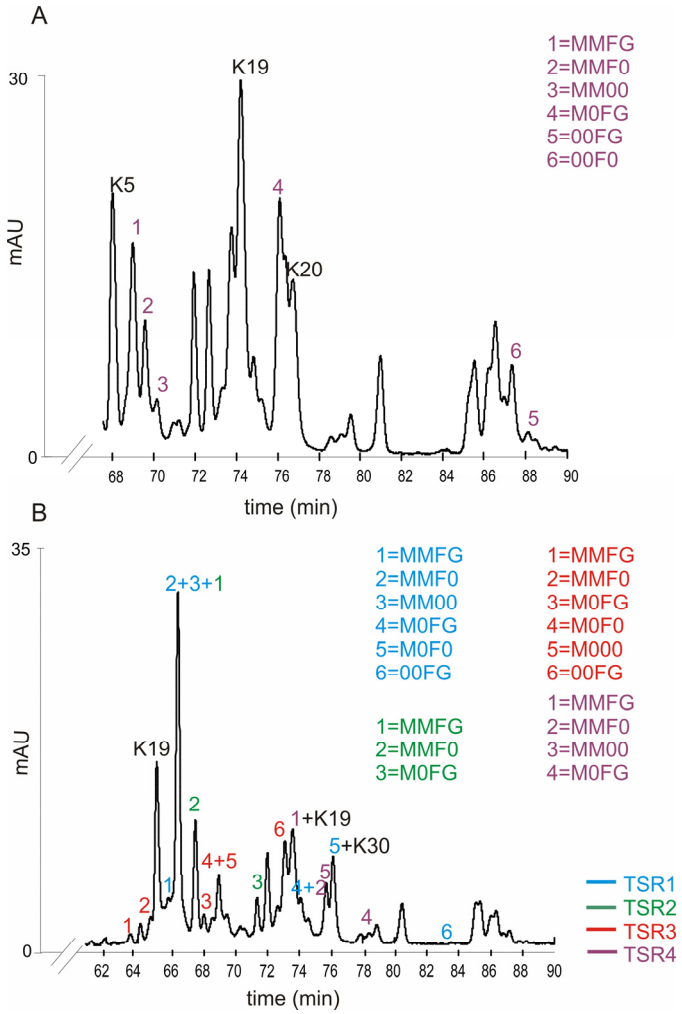
(A) Immobilized metal affinity chromatography (IMAC) of TSR4fchis (B) IMAC of TSR1-4Fchis. (C) UV_{280} trace of 5 µg TSR4fchis protein. (D) SDS-PAGE after elution of TSR4Fchis from protein G beads. (E) UV_{280} trace of 5 µg TSR1-4fchis protein. (F) SDS-PAGE after elution of TSR1-4Fchis from protein G.

The tandem purification strategy resulted in high protein amounts of TSR4fchis (approximately 11.2 mg/l conditioned medium) and TSR1-4Fchis (approximately 9.2 mg/l conditioned medium). The HPLC profile revealed only one major peak for TSR4fchis and TSR1-4fchis. In SDS polyacrylamide gels some contaminants were visible.

In order to investigate the glycosylation of the TSR-fusion proteins, peptides of TSR4fchis and TSR1-4fchis were analyzed by LC-MS. The purified TSR-fusion proteins were reduced, cysteines carboxyamidomethlylated and digested by Lys-C. This

endoproteinase generated peptides encompassing C-mannosylation as well as the disaccharide Glc-Fuc-*O*. To assign the glycopeptides to the measured peaks, the total ion count (TIC) of the LC-MS chromatogram was extracted for the theoretical mass-to-charge ratio (*m/z*) of the glycopeptides. *C*-mannnosylation as well as the disaccharide Glc-Fuc-O were considered for peptide maps. Peptides were only regarded as identified when theoretical and observed masses differed by less than 1 Da. In total, nine different glycoforms per TSR have previously been observed, which result from differential glycosylation of the involved tryptophans, serines or threonines.

Peptide T4+MMFG, for example, represents the fully glycosylated TSR4 having two *C*-mannosylation sides at tryptophan 592 and 595 (nomenclature based on mature F-spondin) and the disaccharide Glc-Fuc-*O*- on threonine 601. Absence of a glycosylation site is indicated with 0. Consequently, the peptide T4+M0FG describes TSR4 plus a *C*-mannosyl residue at W-595 and Glc-Fuc-*O*- at a T-601. This nomenclature was applied to all peptide glycoforms observed in the LC-MS runs of TSR1-4fchis and in TSR4fchis. From previous experiments, it was known that the elution time of *C*-mannosylated peptides strongly decreases with the number of (C^2-Man)Trp residues (*80, 83, 89*). It was found that a peptide containing two (C^2-Man)Trp residues eluted approximately 5-9 minutes earlier than a glycoform having one (C^2-Man)-Trp residues. In contrast, a glycoform containing the disaccharide Glc-Fuc-*O* eluted approximately one minute earlier than the glycoform containing only the *O*-fucose moiety. Using this knowledge, the glycoforms could be grouped according to elution time and *m/z* value (Figure 22).



Abbreviations: M-Mannose, F-Fucose, G-Glucose, 0-non modified

Figure 22: UV chromatogram and assigned peptides from TSR4fchis and TSR1-4fchis

(A) 5 µg TSR4fchis and (B) 5 µg TSR1-4Fchis digested with Lys-C and fractionated by LC-MS. Glycopeptides were assigned according to their *m/z* value and their elution time. Naming was done according to the presence of the glycosylated residues in the relevant TSR peptide. The first M represents *C*-mannosylation of the first tryptophan and the second M *C*-mannosylation of the second tryptophan in the consensus sequence WXXW. The disaccharide Glc-Fuc-O in the consensus sequence $C^1X_{2-3}S/TC^2X_2G$ is indicated by FG. Glycoforms lacking a carbohydrate at a particular position are indicated with 0. Further, assigned peptides belonging to TSR4fchis, TSR1-4fchis are indicated with K plus the number of appearance in the proteins.

The glycosylation analysis revealed that TSR4fchis and TSR1-4fchis expressed in HEK-EBNA cells consist of a large set of different glycoforms. *C*-mannosylation of both

tryptophans was detected in the TSR4fchis protein and in the four consecutive TSR repeats in TSR1-4fchis. Furthermore, all the examined TSRs carried the disaccharide Glc-Fuc-*O*-Ser/Thr. Compared to the single TSR in TSR4fchis fewer peptide glycoforms were observed in the fourth TSR of the TSR1-4 protein. In TSR4fchis the A_{280} area under the glycoforms could be integrated. However, in the peptide map from TSR1-4fchis, various glycoforms and non-glycosylated Lys-C peptides co-eluted. Therefore, an estimation of the degree of modification was not possible in those cases. The results of the integration of TSR4fchis are shown in Table 8. Glycoforms are indicated as a fraction of the total area. The percentage of the individual modified residues is also marked.

identified glycoform	% total of K2 area	Residue modified	Total % of glycosylation
2Man +Fuc-Glc	22.3	(C^2-Man-)Trp^{592}	84.6
2Man +Fuc	11.6	(C^2-Man-)Trp^{595}	37.4
2Man	3.6	Thr^{601}	96.4
Man+Fuc-Glc	47.2	(O-Fuc-Glc-)Thr^{601}	82.3
Fuc-Glc	12.8		
Fuc	2.6		

Table 8: Distribution of glycoforms in the TSR4fchis protein

The area under the peaks in figure 12A was integrated To be able to calculate molar ratios, the areas were corrected for the difference in absorbance between (C^2-Man)Trp and Trp .The areas of peptides with one and two (C^2-Man)Trp residues were divided by 1.3 and 1.6, respectively.

The integration revealed that Trp^{592} (nomenclature according to majored F-spondin) in TSR4fchis is almost completely *C*-mannosylated. Residue Trp^{595} shows only partial modification and Thr^{601} is modified in almost all of the identified glycoforms, predominantly with the disaccharide Glc-Fuc-*O*.

In summary, TSRs can be efficiently expressed in high amount as Fchis fusion proteins in HEK-EBNA cells. Analysis of the glycosylation demonstrated that the cell line has the

capability to modify both tryptophans in the WXXW motif and also possesses POFUT2 and β3Glc-T activity that synthesize the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr. No unglycosylated TSR peptide could be identified. Integration of the areas in TSR4fchis pointed to almost complete glycosylation of the first tryptophan in the WXXW motif and the threonine in the $C^1X_{2-3}S/TC^2X_2G$ consensus sequence.

6.2.1.3 Purification and analysis of TSRs expressed in yeast

We chose the yeast *Pichia pastoris* as the expression system for unglycosylated TSRs. Although *E. coli* also lacks the relevant enzymes for glycosylation, expression of TSRs in this system is complicated due to the formation of intermolecular disulfide bridges. TSR1-4 has not been successfully expressed in *E. coli*. In order to obtain a system having fewer problems with the disulfide bridge formation, TSR-fusion proteins were expressed in *Pichia pastoris*. The yeast was made competent and plasmids for the expression of TSR4fchis and TSR1-4fchis were introduced through electroporation. After selection on agarose plates containing the antibiotic Zeocin™, resistant clones were screened for expression of the transgene by dot blot analysis (not shown). High expression clones were cultured in baffled Erlenmeyer flasks. Protein production was induced by adding methanol. After culturing for 48 hours in the presence of methanol cells were spun and the supernatant was chromatographed over a protein G column. Although Zeocin resistant clones were obtained, unfortunately no expression of TSR1-4fchis was achieved. TSR4fchis, however was successfully expressed and purified with a yield of 250 µg purified protein from 1L culture. Analysis of the protein G-purified fusion protein TSR4fchis is shown in Fig. 23. For comparison, also the TSR4fchis expressed in HEK-EBNA cells is loaded.

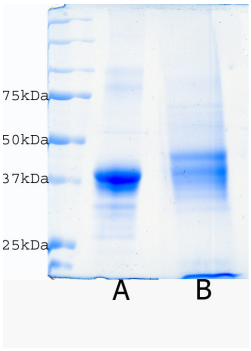


Figure 23: Expression of TSR4fchis in HEK-EBNA cells and in *P. pastoris*

TSR4fchis can be expressed in the methylotrophic yeast *P. pastoris*. Protein production was induced and 48h after induction, the supernatant was chromatographed over protein G columns. Lane A shows TSR4Fchis isolated from HEK-EBNA cells as a control and lane B shows the protein isolated from *P. pastoris* (2 µg of total protein /lane was loaded).

TSR4fchis from *P. pastoris* migrated as two major bands at a molecular weight of about 37 kDa and about 42 kDa. Compared to mammalian TSR4fchis control, the apparent molecular weight in SDS-PAGE of the yeast expressed protein was slightly higher. Western blot analysis with antibodies against the His_6 tag showed bands in this area and additional bands at about 50 kDa (not shown). The bands at 37 kDa and 42 kDa were excised, digested by Lys-C and analyzed by LC-MS on the 4000 Q trap mass spectrometer using a linear gradient of 0-45%B in buffer system 2. This indicated that both bands contained TSR4fchis (not shown). In order to verify that the TSR4 lacked all glycosylation, the peptide map was searched for peptide K2. This peptide contains all potentially glycosylated residues. We only could identify the non-glycosylated peptide (1143.2 Da). Its identity was confirmed by CID tandem MS (Figure 24).

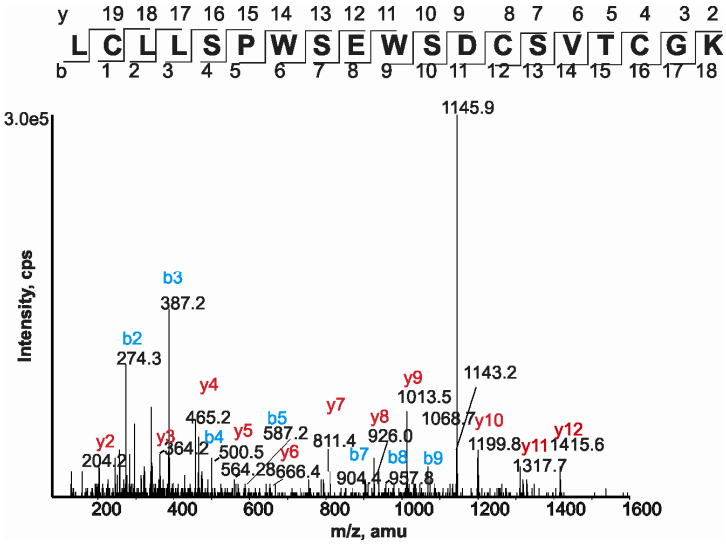


Figure 24: MSMS spectrum of K2 peptide of TSR4fchis produced in *P.pastoris*

Peptide K2 obtained from a Lys-C digest ofTSR4fchis produced in *P.pastoris* was sequenced by MSMS. An almost complete set of b -and -y-ions was obtained showing no indication of glycosylation

Taken together, the data indicate, that TSR4fchis produced in *P. pastoris* was negative for C-mannosylation and modification with the disaccharide Glcβ1,3Fuc-*O*. However, the observation of bands with a higher molecular weight than found in TSR4fchis from mammalian cells is an important concern. It may indicate the presence of different forms of *N*-glycans or *O*-glycans presumably in the Fc portion of the molecule.

6.3 Discussion

6.3.1 *C*-Mannosylation of TSRs expressed in HEK-EBNA cells

The HEK-EBNA system has been used for the recombinant expression of various proteins. The high titers of recombinant proteins are due to a stably integrated EBNA-1 gene that drives the episomal replication of vectors containing the oriP origin of replication (*97*). Similar to other cell lines HEK-EBNA can be selected with a suitable antibiotic to obtain stable expression of the transgene. In HEK-EBNA cells, the transgene is not integrated randomly in the genome, but stays extra-chromosomal (S. Geisse personal communication). Consequently, recombinant proteins can be expressed at high levels even in stable systems. In this study, HEK-EBNA cells were successfully used to express high titers of the recombinant proteins TSR4fchis and TSR1-4fchis. These proteins were demonstrated to be *C*-glycosylated on tryptophans in the WXXW motif and to contain the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr in the consensus sequence $C^1X_{2-3}S/TC^2X_2G$. However, the glycosylation of both TSR4fchis and TSR1-4fchis was heterogeneous. This was mainly caused by partial C-mannosylation of the Trp residues. The exact reason for this is presently unclear, but previous studies have revealed various factors affecting the efficiency of C-mannosylation. Initially, *C*-mannosylation was only observed on the first tryptophan of the WXXW repeat in RNAse (*76, 77, 80*). Later, it was shown in the four terminal components of complement that also the second tryptophan can be *C*-mannosylated. Moreover, even proteins and peptides lacking a WXXW motif were identified to be *C*-mannosylated (*80, 81*). These findings prompted the suggestion of a secondary *C*-mannosylation signal located in the three-dimensional structure of the molecule (*81*). Currently, the structural requirements of this alternative *C*-mannosylation signal are not known.

C-mannosylation has been described to be highly variable. In properdin the tryptophans were demonstrated to be modified almost quantitatively (*70*). In thrombospondin 1, the complement factors, and in this study, variable degrees of *C*-mannosylation between TSRs were observed (*70, 81*). *In vitro* studies with synthetic peptides have shown that

the degree of c-mannosylation is dependent on the amino acid sequence around the Trp residues. For example, the peptide Ac-KWAQW-NH_2 was only modified to a degree of 30% and the peptide Ac-WAQWAK-NH_2 was *C*-mannosylated to about 110% compared to a control peptide. The peptide Ac-WSEW-NH_2 was found to be a substrate showing 130% modification (*62*). These results suggest that amino acids surrounding the WXXW motif can directly influence the level of modification. Some amino acids inside the WXXW motif may decrease or abolish *C*-glycosylation. This is illustrated in the WMNW sequence in β-fibrinogen or in the WFHW sequence in tenacin C. Both proteins have been shown to lack *C*-mannosylation (*62*). The rate of folding of an individual *C*-mannosylated protein could also influence its glycosylation. It was shown that *C*-mannosylation is an early event that is likely to happen when the protein has not adopted its tertiary structure (*80*). This is supported by the observation that in pulse chase experiments the mannose-binding lectin GNA detects putative *C*-mannosylated cysteine-rich sub-domains of mucin 5A and 5B after five minutes of labeling (*84*). The rate of folding amongst proteins can vary considerably, since it is dependent on various factors, like amino acids composition or availability of chaperones or lectins. It is possible that for fast-folding proteins sufficient access of protein *C*-mannosyltransferase to its recognition sequence is reduced. Another factor, which might influence the amount of C-mannosylation, is the availability of sugar donor Dol-P-Man. In *N*-glycans it has been shown that the amount of *N*-glycosylation depends on the size of the pool of Dol-P-Man (*158*). In a similar way, the availability of sugar donors or responsible enzymes could influence the amount of *C*-mannosylation.

Only little is known about the function of *C*-mannosylation on TSRs. Modeling the *C*-mannosyltryptophans into the known crystal structure determined by Tan and colleagues suggested that the sugars protrude from the surface of the molecule (*102*). It is conceivable that these moieties interact with other proteins. Single TSR modules can be expressed in *C*-mannosylation deficient systems like in *E.coli* or *P. pastoris*. Multiple repeats without *C*-mannosylation could not be expressed. The hexoses could therefore play an important role in the correct assembly of multi-TSR-containing proteins.

In the erythropoietin receptor, a structural function for *C*-mannosylation had been suggested. (*87*). Conversely, Perez-Vilar and co-workers proposed a crucial role of *C*-

mannosylation for secretion of cysteine-rich subdomains (*84*) (see Chapter 1.4.1). A putative role for secretion was also obtained from my own experiments. A TSR4-GFP fusion protein was expressed in CHO-K1 wild type cells, in CHOlec15 (deficient in Dol-P-Man synthase) or in CHOlec35.1 cells (deficient in Dol-P-Man utilization). It was observed that TSR4 secretion levels were lower in CHOlec15 and almost absent in CHOlec35.1 cells. In CHOlec35.1 cells, accumulation of TSRs was observed inside the cells (data not shown). A possible explanation is that absence of *C*-mannosylation leads to a misfolded TSR. Alternatively, the absence of *C*-mannosylation abolishes a signal for an unknown protein required for secretion. Future experiments using CHOlec35.1 and CHOlec15 cells and *C*-mannosylated proteins having point mutations in the WXXW repeat could help to define a precise role of this modification on proteins.

6.3.2 *Pichia pastoris* as an expression system for TSRs

In order to generate TSRs lacking C-mannosylation and the disaccharide Glc-Fuc-O, the methylotrophic yeast *Pichia pastoris w*as used for expression studies. This eukaryotic expression system secretes low amounts of endogenous proteins facilitating the purification of recombinant secreted proteins (*98*). Proteins containing a high number of disulphide bridges, like for example the snake venom rhodostomin were demonstrated to be active and correctly folded in *P. pastoris (159).* In order to use *P. pastoris* as an expression system, TSR4fchis and TSR1-4fchis were cloned in a vector containing the *S. cerevisiae* α-factor prepro-signal. Induction of the transgene was driven by the AOX1 promoter. This promoter binds methanol and induces the expression of alcohol oxidase 1, which allows the yeast to grow exclusively on methanol as a carbon source. TSR4fchis was successfully expressed in this system. Mass spectrometric analysis of the Lys-C peptide K2 revealed no glycosylation. In total 250 µg/l TSR4fchis was obtained from *Pichia pastoris*. This was achieved by culturing in baffled Erlenmeyer flasks. Erlenmeyer cultures of TSR4 in *E. coli* yield in average 150 µg/l correctly folded TSR4 (Chun-I Chen personal communication and (*65*)). It is not known if the TSR4fchis produced in the yeast contains a correctly folded TSR. To check this, circular dichroism (CD) analysis would be required. Bacteria and *P. pastoris* generate comparable final

amounts of the TSR4fchis. In *E. coli* a large amount of expressed protein is lost in aggregated TSRs due to formation of inter-molecular disulfide bridges. This problem can be overcome by production in large batch cultures. In *P. pastoris*, protein production in shaking flasks is influenced by various factors resulting in a reduction of recombinant protein amount. The Mut^+ strain of *Pichia pastoris* used in this study easily becomes oxygen deprived in batch cultures. Therefore, the use of controlled bioreactors is preferable for heterologous protein expression in this yeast (*98*). The expressed TSR4fchis protein displayed an apparent molecular weight on an SDS polyacrylamide gel that was higher than that of the recombinant mammalian homologue. The TSR4fchis protein from *P. pastoris* was identified by MS in various bands, indicating that the recombinant protein was potentially differently glycosylated in the Fc portion of the moelcule. The observed effect could also result from inefficient processing of the α-factor prepro-signal. This secretion signal undergoes three consecutive proteolytic steps. The pre signal is recognized by the signal peptidase in the ER to translocate the nascent protein into the ER lumen. The pro signal is recognized by Kex-2 endopeptidase cleaving between Arg and Lys in the pro leader sequence. The third proteolytic cleavage is performed by Ste13 recognizing Glu-Ala repeats. Particularly the third cleavage process can be disturbed by the three-dimensional folding of the protein or by the close proximity of Pro residues (*160*). The expressed TSR4fchis has a Pro residue at amino acid five, which could theoretically perturb the efficient cleavage. Another factor that is often observed in recombinant proteins is hyperglycosylation. *P. pastoris* is capable to add *N*-and *O*-linked glycans onto secreted proteins. *O*-linked glycans in yeast consist of multiple mannoses added to serines or threonines. The mannoses are linked in α1,2-linkage and the glycans can be three to five sugar moieties in length. Similar to mammalian *O*-glycosylation, no consensus sequence is known in yeast. In *P. pastoris,* protein *O*-glycosylation can occur on residues, which are not glycosylated when the protein is expressed in mammalian cells. In some cases hyper-*O*-glycosylation can account for up to 15% of the total molecular weight of the expressed protein (*160*). The TSR4fchis protein contains 19 threonine and 26 serine residues. *O*-glycosylation on yeast expressed TSR4fchis has not been observed in mass spectrometry. However, the fact that *O*-glycosylation was not detected on TSR4fchis

does not mean it is non-existent. Since O-glycans are labile in CID tandem MS experiments, it is possible that potential *O*-glycosylation sites have been missed.
N-glycans are often observed as high mannose-type glycans ($Man_{8-9}GlcNAc_2$-Asn) on secreted proteins, which can occur on sequons that are unused in mammalian cells. The used fusion construct contains one potential *N*-glycosylation site in the CH2 domain of the Fc tag. PNGase F experiments with mammalian Fc-fusion proteins revealed no obvious shift in SDS polyacrylamide gels indicating that no *N*-glycans are present. However, it is possible that in yeast, this site has been used and an *N*-glycan could have been added.

The reason why TSR1-4fchis could not be expressed in *P. pastoris* is presently unclear. Further studies with TSR-containing proteins expressed in *P. pastoris* are necessary to elucidate the underlying reasons. Yeast of the *Saccharomycetaceae* family lack TSR-containing proteins (*103*). It is possible that single TSRs can be efficiently folded in the yeast ER. Proteins like TSR1-4fchis with a large number of cysteins (24 in TSR1-4, 7 in Fc tag) may exhaust the folding machinery leading to a misfolded protein. All these assumptions are made on the basis that the protein was misfolded and therefore degraded. Additional analysis on the transcript level would further confirm this hypothesis.

6.3.3 Suitability of expressed TSRs for neuronal assays

The functions of the different TSRs of F-spondin have been examined by using them as substratum for cultured neuronal cells (*145, 147, 149*). To elucidate the precise role of the glycosylation of TSRs, we intended to perform similar experiments with fully glycosylated and non-glycosylated recombinant TSR-fusion proteins. Whereas the former could be easily obtained from a mammalian expression system only TSR4fchis lacking glycosylation of the TSR could be isolated from *P. pastoris.* In that case, however, indirect evidence suggests that hyperglycosylation of the fusion protein had occurred. Since this could lead to artifacts in cellular assays, it was decided to abort this project at this point.

7 Outlook and summary

7.1 Outlook

The results presented in this thesis raise various questions which will be discussed here:

ΔDS2,3-TSR4 as a model protein

The study of mucin-type *O*-glycans is challenging because of the lack of consensus sequence, the large number of polypeptide GalNAc transferases and the redundant substrate specificities. Enzyme specificity is often determined based on *in vitro* assays using peptides. The ΔDS2,3-TSR4 represents a small protein where the glycosylation has been intensively studied. Modification *in vivo* has been observed with GalNAc T3. The available data from the *in vitro* peptide assay indicates that ΔDS2,3-TSR4 might be a limited substrate that is efficiently modified only by a number of GalNAc transferases. The large number of GalNAc Ts, however suggests functional redundancy. Currently it is not known if redundancy *in vivo* can also be achieved by up-regulation of a less specific enzyme. ΔDS2,3-TSR4 could be used to answer the question if different GalNAc transferases could also functionally complement a missing enzyme. It would be conceivable to inhibit a GalNAc that modifies ΔDS2,3-TSR4 and observe if over-expression of another transferase could rescue the modification. The fact that ΔDS2,3-TSR4 expressed in different cells was found to be modified only with the tetrasaccharide allows the detailed study of the involved enzymes. Besides the elucidation of the GalNAc T, the question can also be answered which sialyltransferase can add the sialic acids in α2,3 and in α2,6 linkage to ΔDS2,3-TSR4. These analysis could help to reveal how different glycosyltransferases contribute to the formation of the tetrasaccharide on ΔDS2,3-TSR4.

F-spondin as a model protein for detection of CDGs

An increasing number of patients suspected to suffer from a congenital disorder of glycosylation has to be grouped as CDG-IIx, because their defects can not be detected using the classical CDG marker proteins transferrin or α1-antitrypsin. Therefore, novel marker proteins are required which would not only detect defects in *N*-glycosylation but reveal the global impact of glycans on the cell. F-spondin could largely fulfill these criteria. F-spondin has two predicted *N*-glycosylation sites, which require further detailed analysis. Preliminary experiments indicate, however, that they contain hybrid and complex *N*-glycans (personal information Chun-I Chen unpublished). Together with the identification of the disialyl-T-antigen (this work) and the previously reported C-mannosylation and the disaccharide Glcβ1,3-Fuc-*O*-Ser/Thr (*61*) F-spondin contains at least four different types of glycosylation. These glycosylations occur in the ER (*N*-glycans, C-mannosylation and Glcβ1,3-Fuc-*O*-Ser/Thr) but also in the Golgi (disialyl-T-antigen, *N*-glycan) and involve a variety of different monosaccharides and sugar donors. Whether glycosylation of F-spondin could be used as a novel disease marker for CDGs has to be demonstrated experimentally. Therefore, further detailed analysis of F-spondin from healthy individuals versus F-spondin from patients is required to validate this approach. The glycan analysis of F-spondin could be based on mass spectrometric methods similar to those applied in this dissertation, which would provide sensitive and also quantitative information about the glycosylation pattern.

Function of glycosylation on TSRs of F-spondin

The TSRs of F-spondin have been implicated to exert positive as well as negative cues on various neuronal subpopulations. Currently it is not known how glycosylation is involved in these processes. Recently it was shown that soluble TSR1-4 fused to a Myc-tag can be immobilized on the surface of COS cells by co-expression with ApoeR. The TSRs were detected with the 9E10 anti-myc antibody and visualized by microscopy (*150*). In order to assess a function of the glycan this binding could be competed with glycosylated peptides. (C^2-Man-*)*Trp containing peptides can either by synthesized chemically (*161*) or peptides containing *C*-mannosylation as well as Glcβ1,3-Fuc-*O*-Ser/Thr can be purified by HPLC using TSR4cfhis and TSR1-4fchis from HEK-EBNA

cells. As controls, peptides obtained from peptide synthesis could be used. In order to extend the study for function of the glycans on thrombospondin-type 1 repeats, single TSR repeats could be expressed in mammalian cells and in a system incapable to perform the required glycosylation (preferably bacteria). These proteins could be purified and plated as substratum for cells. To minimize the experimental error the glycosylated and unglycosylated protein could be plated in stripes next to each other using a specific silicon matrix. This matrix was applied in a neuronal outgrowth assay where it was shown that neurons extend their neurites only on substrate that promotes outgrowth (*162*). In a similar experiment the effect of glycosylation on TSRs on neuronal outgrowth could be studied by plating a mammalian TSR next to the homologous protein expressed in bacteria. This assay could be used to study the effect of TSR glycosylation on neurite outgrowth of various neuronal subpopulations.

7.2 Summary

In the course of mutagenesis of the forth thrombospondin type 1 repeat (TSR4) of the axonal guidance protein F-spondin, one mutant designated ΔDS2,3-TSR4 missing two out of three cystein bridges was found to lack the disaccharide Glcβ1,3-Fuc-*O*. This protein was demonstrated to be secreted as efficiently as wild type TSR4. In gel filtration experiments unusual features of folding could not be observed. Analysis of the protein by mass spectrometry indicated that the module had acquired an additional glycosylation. Using a combination of tandem CID MS and digestion with glycosidases, it was demonstrated that the tetrasaccharide corresponds to a biantennary NeuNAcα2,3Galβ1,3[NeuNAcα2,6]GalNAc-*O*- glycan. This carbohydrate structure is also known as the disialyl-T-antigen. The glycan could be demonstrated to be attached exclusively to the same residue (Thr-601) that carries the disaccharide Glcβ1,3-Fuc-*O* in wild type TSR4. ΔDS2,3-TSR4 is consequently an experimental proof that one residue can be modified with two different types of glycosylation. Since the disaccharide Glcβ1,3-Fuc-*O* is transferred in the Endoplasmatic Reticulum and the disialyl-T-antigen is added in the Golgi, ΔDS2,3-TSR4 has undergone a glycosylation shift involving different cellular organelles. In order to identify the enzymes that recognized the module and modify the Thr-601 residue ΔDS2,3-TSR4-derived peptides were screened in an in *vitro assay*. It was found that GalNAc T1 and GalNAc T3 are capable to add GalNAc moieties. GalNAc T1 was found to glycosylate as well Thr-601 and Ser-596. GalNAc T3 was identified to be specific for Thr-601, thus mimicking the situation observed *in vivo*. The role of this transferase could be further confirmed *in vivo* by co-expression experiments of GalNAc T3 together with ΔDS2,3-TSR4 in CHO-K1 cells. These experiments resulted in significant increase of the disialyl-T-antigen on glycopeptides of ΔDS2,3-TSR4. The results further demonstrated that after initiation of mucin-type *O*-glycosylation by a glycosyltransferase such as GalNAc T3 the carbohydrate is extended to the tetrasaccharide of a disialyl-T-antigen type in various cell types. Interestingly the same O-glycan was also identified in the rat axonal guidance protein F-spondin. F-spondin had previously been described to be *C*-mannosylated on tryptophans and to contain the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr. Using low energy CID tandem MS in

combination with glycosidases and specific sugar-binding lectins, it was possible to map the disialyl-T-antigen to a peptide located in the N-terminal reelin/spondin domain. F-spondin is therefore the first protein known to contain C-mannosylation, the disaccharide Glcβ1,3Fuc-*O*-Ser/Thr, N-glycosylation and also the tetrasacchride NeuNAcα2,3Galβ1,3 [NeuNAcα2,6]GalNAc-*O*-.

In order to study the glycosylation of thrombospondin type 1 repeats of F-spondin in more detail, the fourth (TSR4) and four consecutive thrombospondin type 1 repeats (TSR1-4) were expressed in the mammalian cell line HEK-EBNA. In order to facilitate purification, the proteins were expressed as fusion proteins designated TSR4fchis and TSR1-4fchis. The proteins were purified in high amounts and *C*-mannosylation as well as the modification with the disaccharide Glcβ1,3-Fuc-*O* were examined by LC-MS. It was found that the TSR4fchis and TSR1-4fchis proteins carried *C*-mannosylation on their tryptophans and Glcβ1,3-Fuc-*O* on their serines or threonines in the consensus sequence $C^1X_{2-3}S/TC^2X_2G$. Quantification of the amount of glycosylation on the TSR4fchis protein revealed that the first tryptophan in the WXXW motif is almost quantitatively *C*-mannosylated, the second tryptophan shows partial *C*-mannosylation and the threonine in the consensus sequence $C^1X_{2-3}S/TC^2X_2G$ was predominantly modified with the disaccharide Glcβ1,3-Fuc-*O*. In order to have non-glycosylated TSRs, TSR4fchis and TSR1-4fchis were expressed in the methylotrophic yeast *Pichia pastoris.* TSR4fchis could be successfully expressed and purified in amounts comparable to correctly folded TSR4 expressed in *E. coli*. It was found that this host is negative for *C*-mannosylation and *O*-fucosylation. However, indirect evidence indicated, that TSR4fchis expressed in yeast had undergone hyperglycosylation of the fusion protein. Therefore a direct comparison of TSR4fchis from HEK-EBNA and TSR4fchis from *P. pastoris* was not possible.

References

1. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease, *Cell 126*, 855-867.
2. Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A., and Sipila, L. (2007) Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3), *J Cell Physiol 212*, 323-329.
3. Olsen, E. H., Rahbek-Nielsen, H., Thogersen, I. B., Roepstorff, P., and Enghild, J. J. (1998) Posttranslational modifications of human inter-alpha-inhibitor: identification of glycans and disulfide bridges in heavy chains 1 and 2, *Biochemistry 37*, 408-416.
4. West, C. M., Van Der Wel, H., Sassi, S., and Gaucher, E. A. (2004) Cytoplasmic glycosylation of protein-hydroxyproline and its relationship to other glycosylation pathways, *Biochim Biophys Acta 1673*, 29-44.
5. Kieliszewski, M. J., and Shpak, E. (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins, *Cell Mol Life Sci 58*, 1386-1398.
6. Smythe, C., Caudwell, F. B., Ferguson, M., and Cohen, P. (1988) Isolation and structural analysis of a peptide containing the novel tyrosyl-glucose linkage in glycogenin, *Embo J 7*, 2681-2686.
7. Lowe, J. B., and Marth, J. D. (2003) A genetic approach to Mammalian glycan function, *Annu Rev Biochem 72*, 643-691.
8. Varki, A. C., Richard; Esko, Jeffrey; Freeze, Hudson; Hart, Gerald; Marth, Jamey, (1999) *Essentials of Glycobiology* Cold Spring Harbor Laboratory Press
9. Szymanski, C. M., and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens, *Nat Rev Microbiol 3*, 225-237.
10. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum, *Annu Rev Biochem 73*, 1019-1049.

11. Helenius, J., Ng, D. T., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M. (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein, *Nature 415*, 447-450.
12. DeFrees, S., Wang, Z. G., Xing, R., Scott, A. E., Wang, J., Zopf, D., Gouty, D. L., Sjoberg, E. R., Panneerselvam, K., Brinkman-Van der Linden, E. C., Bayer, R. J., Tarp, M. A., and Clausen, H. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli, *Glycobiology 16*, 833-843.
13. Molinari, M. (2007) N-glycan structure dictates extension of protein folding or onset of disposal, *Nat Chem Biol 3*, 313-320.
14. Caramelo, J. J., Castro, O. A., de Prat-Gay, G., and Parodi, A. J. (2004) The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates, *J Biol Chem 279*, 46280-46285.
15. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology 12*, 43R-56R.
16. Wiertz, E. J., Tortorella, D., Bogyo, M., Yu, J., Mothes, W., Jones, T. R., Rapoport, T. A., and Ploegh, H. L. (1996) Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction, *Nature 384*, 432-438.
17. Ghosh, P., Dahms, N. M., and Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale, *Nat Rev Mol Cell Biol 4*, 202-212.
18. Ioffe, E., and Stanley, P. (1994) Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates, *Proc Natl Acad Sci U S A 91*, 728-732.
19. Ungar, D., Oka, T., Vasile, E., Krieger, M., and Hughson, F. M. (2005) Subunit architecture of the conserved oligomeric Golgi complex, *J Biol Chem 280*, 32729-32735.
20. Freeze, H. H. (2006) Genetic defects in the human glycome, *Nat Rev Genet 7*, 537-551.
21. Freeze, H. H., and Aebi, M. (2005) Altered glycan structures: the molecular basis of congenital disorders of glycosylation, *Curr Opin Struct Biol 15*, 490-498.

22. Kornak, U., Reynders, E., Dimopoulou, A., van Reeuwijk, J., Fischer, B., Rajab, A., Budde, B., Nurnberg, P., Foulquier, F., Lefeber, D., Urban, Z., Gruenewald, S., Annaert, W., Brunner, H. G., van Bokhoven, H., Wevers, R., Morava, E., Matthijs, G., Van Maldergem, L., and Mundlos, S. (2008) Impaired glycosylation and cutis laxa caused by mutations in the vesicular H+-ATPase subunit ATP6V0A2, *Nat Genet 40*, 32-34.

23. Sasaki, H., Bothner, B., Dell, A., and Fukuda, M. (1987) Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA, *J Biol Chem 262*, 12059-12076.

24. Kaneko, M. K., Kato, Y., Kameyama, A., Ito, H., Kuno, A., Hirabayashi, J., Kubota, T., Amano, K., Chiba, Y., Hasegawa, Y., Sasagawa, I., Mishima, K., and Narimatsu, H. (2007) Functional glycosylation of human podoplanin: glycan structure of platelet aggregation-inducing factor, *FEBS Lett 581*, 331-336.

25. Singh, R., Subramanian, S., Rhodes, J. M., and Campbell, B. J. (2006) Peanut lectin stimulates proliferation of colon cancer cells by interaction with glycosylated CD44v6 isoforms and consequential activation of c-Met and MAPK: functional implications for disease-associated glycosylation changes, *Glycobiology 16*, 594-601.

26. Gerken, T. A., Raman, J., Fritz, T. A., and Jamison, O. (2006) Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates, *J Biol Chem 281*, 32403-32416.

27. Hennet, T., Hagen, F. K., Tabak, L. A., and Marth, J. D. (1995) T-cell-specific deletion of a polypeptide N-acetylgalactosaminyl-transferase gene by site-directed recombination, *Proc Natl Acad Sci U S A 92*, 12070-12074.

28. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy, *J Biol Chem 274*, 25362-25370.

29. Topaz, O., Shurman, D. L., Bergman, R., Indelman, M., Ratajczak, P., Mizrachi, M., Khamaysi, Z., Behar, D., Petronius, D., Friedman, V., Zelikovic, I., Raimer, S., Metzker, A., Richard, G., and Sprecher, E. (2004) Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis, *Nat Genet 36*, 579-581.
30. Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, *Glycobiology 13*, 1R-16R.
31. Elhammer, A. P., Kezdy, F. J., and Kurosaka, A. (1999) The acceptor specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases, *Glycoconj J 16*, 171-180.
32. Brockhausen, I. (2006) Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions, *EMBO Rep 7*, 599-604.
33. Hang, H. C., and Bertozzi, C. R. (2005) The chemistry and biology of mucin-type O-linked glycosylation, *Bioorg Med Chem 13*, 5021-5034.
34. Benet-Pages, A., Orlik, P., Strom, T. M., and Lorenz-Depiereux, B. (2005) An FGF23 missense mutation causes familial tumoral calcinosis with hyperphosphatemia, *Hum Mol Genet 14*, 385-390.
35. Kato, K., Jeanneau, C., Tarp, M. A., Benet-Pages, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., and Clausen, H. (2006) Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation, *J Biol Chem 281*, 18370-18377.
36. Ju, T., and Cummings, R. D. (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase, *Proc Natl Acad Sci U S A 99*, 16613-16618.
37. Dall'Olio, F., and Chiricolo, M. (2001) Sialyltransferases in cancer, *Glycoconj J 18*, 841-850.
38. Harris, R. J., and Spellman, M. W. (1993) O-linked fucose and other post-translational modifications unique to EGF modules, *Glycobiology 3*, 219-224.

39. Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F., and Peter-Katalinic, J. (2001) C-mannosylation and O-fucosylation of the thrombospondin type 1 module, *J Biol Chem 276*, 6485-6498.
40. Nakakura, N., Hietter, H., Van Dorsselaer, A., and Luu, B. (1992) Isolation and structural determination of three peptides from the insect Locusta migratoria. Identification of a deoxyhexose-linked peptide, *Eur J Biochem 204*, 147-153.
41. Wang, Y., and Spellman, M. W. (1998) Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells, *J Biol Chem 273*, 8112-8118.
42. Luo, Y., Koles, K., Vorndam, W., Haltiwanger, R. S., and Panin, V. M. (2006) Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats, *J Biol Chem 281*, 9393-9399.
43. Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006) Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats, *J Biol Chem 281*, 9385-9392.
44. Rampal, R., Luther, K. B., and Haltiwanger, R. S. (2007) Notch signaling in normal and disease States: possible therapies related to glycosylation, *Curr Mol Med 7*, 427-445.
45. Hamada, Y., Kadokawa, Y., Okabe, M., Ikawa, M., Coleman, J. R., and Tsujimoto, Y. (1999) Mutation in ankyrin repeats of the mouse Notch2 gene induces early embryonic lethality, *Development 126*, 3415-3424.
46. Swiatek, P. J., Lindsell, C. E., del Amo, F. F., Weinmaster, G., and Gridley, T. (1994) Notch1 is essential for postimplantation development in mice, *Genes Dev 8*, 707-719.
47. Shi, S., and Stanley, P. (2003) Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways, *Proc Natl Acad Sci U S A 100*, 5234-5239.
48. Haltiwanger, R. S., and Stanley, P. (2002) Modulation of receptor signaling by glycosylation: fringe is an O-fucose-beta1,3-N-acetylglucosaminyltransferase, *Biochim Biophys Acta 1573*, 328-335.

49. Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K. D., Haltiwanger, R. S., and Vogt, T. F. (2000) Fringe is a glycosyltransferase that modifies Notch, *Nature 406*, 369-375.
50. Chen, J., Moloney, D. J., and Stanley, P. (2001) Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1, *Proc Natl Acad Sci U S A 98*, 13716-13721.
51. Xu, A., Haines, N., Dlugosz, M., Rana, N. A., Takeuchi, H., Haltiwanger, R. S., and Irvine, K. D. (2007) In vitro reconstitution of the modulation of Drosophila Notch-ligand binding by Fringe, *J Biol Chem 282*, 35153-35162.
52. Okajima, T., Xu, A., Lei, L., and Irvine, K. D. (2005) Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding, *Science 307*, 1599-1603.
53. Okajima, T., Reddy, B., Matsuda, T., and Irvine, K. D. (2008) Contributions of chaperone and glycosyltransferase activities of O-fucosyltransferase 1 to Notch signaling, *BMC Biol 6*, 1.
54. Lu, L., and Stanley, P. (2006) Roles of O-fucose glycans in notch signaling revealed by mutant mice, *Methods Enzymol 417*, 127-136.
55. Yang, L. T., Nichols, J. T., Yao, C., Manilay, J. O., Robey, E. A., and Weinmaster, G. (2005) Fringe glycosyltransferases differentially modulate Notch1 proteolysis induced by Delta1 and Jagged1, *Mol Biol Cell 16*, 927-942.
56. Arboleda-Velasquez, J. F., Rampal, R., Fung, E., Darland, D. C., Liu, M., Martinez, M. C., Donahue, C. P., Navarro-Gonzalez, M. F., Libby, P., D'Amore, P. A., Aikawa, M., Haltiwanger, R. S., and Kosik, K. S. (2005) CADASIL mutations impair Notch3 glycosylation by Fringe, *Hum Mol Genet 14*, 1631-1639.
57. Visan, I., Tan, J. B., Yuan, J. S., Harper, J. A., Koch, U., and Guidos, C. J. (2006) Regulation of T lymphopoiesis by Notch1 and Lunatic fringe-mediated competition for intrathymic niches, *Nat Immunol 7*, 634-643.
58. Rabbani, S. A., Mazar, A. P., Bernier, S. M., Haq, M., Bolivar, I., Henkin, J., and Goltzman, D. (1992) Structural requirements for the growth factor activity of the amino-terminal domain of urokinase, *J Biol Chem 267*, 14151-14156.

59. Yan, Y. T., Liu, J. J., Luo, Y., E, C., Haltiwanger, R. S., Abate-Shen, C., and Shen, M. M. (2002) Dual roles of Cripto as a ligand and coreceptor in the nodal signaling pathway, *Mol Cell Biol 22*, 4439-4449.

60. Shi, S., Ge, C., Luo, Y., Hou, X., Haltiwanger, R. S., and Stanley, P. (2007) The threonine that carries fucose, but not fucose, is required for Cripto to facilitate Nodal signaling, *J Biol Chem 282*, 20133-20141.

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

62. Hartmann, S. (2000) Protein C-Mannosylation, in *Philosophisch-Naturwissenschaftliche Fakultät*, p 148, University of basel, Basel.

63. Ricketts, L. M., Dlugosz, M., Luther, K. B., Haltiwanger, R. S., and Majerus, E. M. (2007) O-fucosylation is required for ADAMTS13 secretion, *J Biol Chem 282*, 17014-17023.

64. Wang, L. W., Dlugosz, M., Somerville, R. P., Raed, M., Haltiwanger, R. S., and Apte, S. S. (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.

65. Kozma, K., Keusch, J. J., Hegemann, B., Luther, K. B., Klein, D., Hess, D., Haltiwanger, R. S., and Hofsteenge, J. (2006) Identification and characterization of abeta1,3-glucosyltransferase that synthesizes the Glc-beta1,3-Fuc disaccharide on thrombospondin type 1 repeats, *J Biol Chem 281*, 36742-36751.

66. Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F., and Peter-Katalinic, J. (2001) C-mannosylation and O-fucosylation of the thrombospondin type 1 module, *J Biol Chem 276*, 6485-6498.

67. Sato, T., Sato, M., Kiyohara, K., Sogabe, M., Shikanai, T., Kikuchi, N., Togayachi, A., Ishida, H., Ito, H., Kameyama, A., Gotoh, M., and Narimatsu, H. (2006) Molecular cloning and characterization of a novel human beta1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates O-linked fucosylglycan on thrombospondin type 1 repeat domain, *Glycobiology 16*, 1194-1206.

68. Canevascini, S., Klein, D., Grob, M., Keusch, K., Hess, D., Althaus, J., Chiquet-Ehrismann, R., and Hofsteenge, J. (2007) The O-fucosyltransferase PAD-2/POFUT2 regulates distal tip cell migration in C. elegans., Basel.
69. Lesnik Oberstein, S. A., Kriek, M., White, S. J., Kalf, M. E., Szuhai, K., den Dunnen, J. T., Breuning, M. H., and Hennekam, R. C. (2006) Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase, *Am J Hum Genet 79*, 562-566.
70. Hartmann, S., and Hofsteenge, J. (2000) Properdin, the positive regulator of complement, is highly C-mannosylated, *J Biol Chem 275*, 28569-28574.
71. Hess, D., Keusch, J. J., Lesnik Oberstein, S. A., Hennekam, R. C., and 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*.
72. Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W., and Marth, J. D. (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny, *Proc Natl Acad Sci U S A 97*, 5735-5739.
73. Hanover, J. A., Forsythe, M. E., Hennessey, P. T., Brodigan, T. M., Love, D. C., Ashwell, G., and Krause, M. (2005) A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout, *Proc Natl Acad Sci U S A 102*, 11266-11271.
74. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins, *Nature 446*, 1017-1022.
75. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W., and Gong, C. X. (2004) O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease, *Proc Natl Acad Sci U S A 101*, 10804-10809.
76. Hofsteenge, J., Muller, D. R., de Beer, T., Loffler, A., Richter, W. J., and Vliegenthart, J. F. (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.

77. de Beer, T., Vliegenthart, J. F., Loffler, A., and 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.

78. Loffler, A., Doucey, M. A., Jansson, A. M., Muller, D. R., de Beer, T., Hess, D., Meldal, M., Richter, W. J., Vliegenthart, J. F., and Hofsteenge, J. (1996) Spectroscopic and protein chemical analyses demonstrate the presence of C-mannosylated tryptophan in intact human RNase 2 and its isoforms, *Biochemistry 35*, 12005-12014.

79. Li, Q. M., van den Heuvel, H., Delorenzo, O., Corthout, J., Pieters, L. A., Vlietinck, A. J., and Claeys, M. (1991) Mass spectral characterization of C-glycosidic flavonoids isolated from a medicinal plant (Passiflora incarnata), *J Chromatogr 562*, 435-446.

80. Krieg, J., Hartmann, S., Vicentini, A., Glasner, W., Hess, D., and 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.

81. Hofsteenge, J., Blommers, M., Hess, D., Furmanek, A., and Miroshnichenko, O. (1999) The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues, *J Biol Chem 274*, 32786-32794.

82. Furmanek, A., and Hofsteenge, J. (2000) Protein C-mannosylation: facts and questions, *Acta Biochim Pol 47*, 781-789.

83. Doucey, M. A., Hess, D., Cacan, R., and Hofsteenge, J. (1998) Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor, *Mol Biol Cell 9*, 291-300.

84. Perez-Vilar, J., Randell, S. H., and Boucher, R. C. (2004) C-mannosylation of MUC5AC and MUC5B CYS subdomains, *Glycobiology*.

85. Munte, C. E., Gade, G., Domogalla, B., Kremer, W., Kellner, R., and Kalbitzer, H. R. (2008) C-mannosylation in the hypertrehalosaemic hormone from the stick insect Carausius morosus, *FEBS J*.

86. Ihara, Y., Manabe, S., Kanda, M., Kawano, H., Nakayama, T., Sekine, I., Kondo, T., and Ito, Y. (2004) Increased expression of protein C-mannosylation in the aortic vessels of diabetic Zucker rats, *Glycobiology*.

87. Furmanek, A., Hess, D., Rogniaux, H., and Hofsteenge, J. (2003) The WSAWS motif is C-hexosylated in a soluble form of the erythropoietin receptor, *Biochemistry 42*, 8452-8458.

88. Ervin, L. A., Ball, L. E., Crouch, R. K., and Schey, K. L. (2005) Phosphorylation and Glycosylation of Bovine Lens MP20, *Invest Ophthalmol Vis Sci 46*, 627-635.

89. Falzarano, D., Krokhin, O., Van Domselaar, G., Wolf, K., Seebach, J., Schnittler, H. J., and Feldmann, H. (2007) Ebola sGP--the first viral glycoprotein shown to be C-mannosylated, *Virology 368*, 83-90.

90. Nishikawa, T., Kajii, S., Sato, C., Yasukawa, Z., Kitajima, K., and Isobe, M. (2004) Alpha-C-mannosyltryptophan is not recognized by conventional mannose-binding lectins, *Bioorg Med Chem 12*, 2343-2348.

91. Li, J. S., Cui, L., Rock, D. L., and Li, J. (2005) Novel glycosidic linkage in Aedes aegypti chorion peroxidase: N-mannosyl tryptophan, *J Biol Chem 280*, 38513-38521.

92. Hilton, D. J., Watowich, S. S., Katz, L., and Lodish, H. F. (1996) Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor, *J Biol Chem 271*, 4699-4708.

93. Anand, M., Rush, J. S., Ray, S., Doucey, M. A., Weik, J., Ware, F. E., Hofsteenge, J., Waechter, C. J., and Lehrman, M. A. (2001) Requirement of the Lec35 gene for all known classes of monosaccharide-P-dolichol-dependent glycosyltransferase reactions in mammals, *Mol Biol Cell 12*, 487-501.

94. Acar, M., Jafar-Nejad, H., Takeuchi, H., Rajan, A., Ibrani, D., Rana, N. A., Pan, H., Haltiwanger, R. S., and H.J., B. (2008) Rumi Is a CAP10 Domain Glycosyltransferase that Modifies Notch and Is Required for Notch Signaling, *Cell 132*, 247-258.

95. Endo, T. (2004) Structure, function and pathology of O-mannosyl glycans, *Glycoconj J 21*, 3-7.

96. Cachianes, G., Ho, C., Weber, R. F., Williams, S. R., Goeddel, D. V., and Leung, D. W. (1993) Epstein-Barr virus-derived vectors for transient and stable expression of recombinant proteins, *Biotechniques 15*, 255-259.

97. Geisse, S., and Henke, M. (2005) Large-scale transient transfection of mammalian cells: a newly emerging attractive option for recombinant protein production, *J Struct Funct Genomics 6*, 165-170.

98. Macauley-Patrick, S., Fazenda, M. L., McNeil, B., and Harvey, L. M. (2005) Heterologous protein production using the Pichia pastoris expression system, *Yeast 22*, 249-270.

99. Sutovsky, H., and Gazit, E. (2004) The von Hippel-Lindau tumor suppressor protein is a molten globule under native conditions: implications for its physiological activities, *J Biol Chem 279*, 17190-17196.

100. Chaudhuri, T. K., and Sinha, N. K. (1996) Refolding of trypsin-subtilisin inhibitor from marine turtle eggwhite, *J Protein Chem 15*, 315-320.

101. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.

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

103. Tucker, R. P. (2004) The thrombospondin type 1 repeat superfamily, *Int J Biochem Cell Biol 36*, 969-974.

104. Huwiler, K. G., Vestling, M. M., Annis, D. S., and Mosher, D. F. (2002) Biophysical characterization, including disulfide bond assignments, of the anti-angiogenic type 1 domains of human thrombospondin-1, *Biochemistry 41*, 14329-14339.

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

106. Asch, A. S., Silbiger, S., Heimer, E., and Nachman, R. L. (1992) Thrombospondin sequence motif (CSVTCG) is responsible for CD36 binding, *Biochem Biophys Res Commun 182*, 1208-1217.

107. Simantov, R., Febbraio, M., and Silverstein, R. L. (2005) The antiangiogenic effect of thrombospondin-2 is mediated by CD36 and modulated by histidine-rich glycoprotein, *Matrix Biol 24*, 27-34.

108. Jimenez, B., Volpert, O. V., Crawford, S. E., Febbraio, M., Silverstein, R. L., and Bouck, N. (2000) Signals leading to apoptosis-dependent inhibition of neovascularization by thrombospondin-1, *Nat Med 6*, 41-48.

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

110. Tortorella, M., Pratta, M., Liu, R. Q., Abbaszade, I., Ross, H., Burn, T., and Arner, E. (2000) The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for aggrecan substrate recognition and cleavage, *J Biol Chem 275*, 25791-25797.

111. Oriol, R., Martinez-Duncker, I., Chantret, I., Mollicone, R., and Codogno, P. (2002) Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate, *Mol Biol Evol 19*, 1451-1463.

112. Baldus, S. E., Wienand, J. R., Werner, J. P., Landsberg, S., Drebber, U., Hanisch, F. G., and Dienes, H. P. (2005) Expression of MUC1, MUC2 and oligosaccharide epitopes in breast cancer: prognostic significance of a sialylated MUC1 epitope, *Int J Oncol 27*, 1289-1297.

113. Campbell, B. J., Yu, L. G., and Rhodes, J. M. (2001) Altered glycosylation in inflammatory bowel disease: a possible role in cancer development, *Glycoconj J 18*, 851-858.

114. Santos-Silva, F., Fonseca, A., Caffrey, T., Carvalho, F., Mesquita, P., Reis, C., Almeida, R., David, L., and Hollingsworth, M. A. (2005) Thomsen-Friedenreich antigen expression in gastric carcinomas is associated with MUC1 mucin VNTR polymorphism, *Glycobiology 15*, 511-517.

115. Canevascini, S., Kozma, K., Grob, M., Althaus, J., Klein, D., Chiquet-, and Ehrismann, R., and Hofsteenge, J. . (2006) in *European Worm Meeting,*

April 29-May 3, 2006, p p. 71, Hersonissos, Crete, Greece.

116. Paakkonen, K., Tossavainen, H., Permi, P., Rakkolainen, H., Rauvala, H., Raulo, E., Kilpelainen, I., and Guntert, P. (2006) Solution structures of the first and fourth TSR domains of F-spondin, *Proteins 64*, 665-672.

117. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene 77*, 51-59.

118. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry, *Nature 379*, 466-469.

119. Sorensen, A. L., Reis, C. A., Tarp, M. A., Mandel, U., Ramachandran, K., Sankaranarayanan, V., Schwientek, T., Graham, R., Taylor-Papadimitriou, J., Hollingsworth, M. A., Burchell, J., and Clausen, H. (2006) Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance, *Glycobiology 16*, 96-107.

120. Dwek, R. A., Edge, C. J., Harvey, D. J., Wormald, M. R., and Parekh, R. B. (1993) Analysis of glycoprotein-associated oligosaccharides, *Annu Rev Biochem 62*, 65-100.

121. Dall'Olio, F., and Chiricolo, M. (2001) Sialyltransferases in cancer, *Glycoconj J 18*, 841-850.

122. Hulsmeier, A. J., Paesold-Burda, P., and Hennet, T. (2007) N-Glycosylation Site Occupancy in Serum Glycoproteins Using Multiple Reaction Monitoring Liquid Chromatography-Mass Spectrometry, *Mol Cell Proteomics 6*, 2132-2138.

123. Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., and Clausen, H. (2007) The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation, *Glycobiology 17*, 374-387.

124. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-

acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities, *J Biol Chem 275*, 38197-38205.

125. Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., and Mandel, U. (2007) Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer, *Acta Histochem 109*, 273-284.
126. Sutherlin, M. E., Nishimori, I., Caffrey, T., Bennett, E. P., Hassan, H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M. A. (1997) Expression of three UDP-N-acetyl-alpha-D-galactosamine:polypeptide GalNAc N-acetylgalactosaminyltransferases in adenocarcinoma cell lines, *Cancer Res 57*, 4744-4748.
127. Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., and Mandel, U. (2007) Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer, *Acta Histochem 109*, 273-284.
128. Vogt, G., Vogt, B., Chuzhanova, N., Julenius, K., Cooper, D. N., and Casanova, J. L. (2007) Gain-of-glycosylation mutations, *Curr Opin Genet Dev 17*, 245-251.
129. Santucci, R., Sinibaldi, F., and Fiorucci, L. (2008) Protein Folding, Unfolding and Misfolding: Role Played by Intermediate States, *Mini Rev Med Chem 8*, 57-62.
130. Morrow, J. A., Hatters, D. M., Lu, B., Hochtl, P., Oberg, K. A., Rupp, B., and Weisgraber, K. H. (2002) Apolipoprotein E4 forms a molten globule. A potential basis for its association with disease, *J Biol Chem 277*, 50380-50385.
131. Farrell, H. M., Jr., Qi, P. X., Brown, E. M., Cooke, P. H., Tunick, M. H., Wickham, E. D., and Unruh, J. J. (2002) Molten globule structures in milk proteins: implications for potential new structure-function relationships, *J Dairy Sci 85*, 459-471.
132. Bennett, E. P., Hassan, H., and Clausen, H. (1996) cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D-galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3, *J Biol Chem 271*, 17006-17012.
133. Bennett, E. P., Hassan, H., Hollingsworth, M. A., and Clausen, H. (1999) A novel human UDP-N-acetyl-D-galactosamine:polypeptide N-

acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates, *FEBS Lett 460*, 226-230.

134. Cheng, L., Tachibana, K., Zhang, Y., Guo, J., Kahori Tachibana, K., Kameyama, A., Wang, H., Hiruma, T., Iwasaki, H., Togayachi, A., Kudo, T., and Narimatsu, H. (2002) Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T10, *FEBS Lett 531*, 115-121.
135. Tarp, M. A., and Clausen, H. (2007) Mucin-type O-glycosylation and its potential use in drug and vaccine development, *Biochim Biophys Acta*.
136. Nehrke, K., Ten Hagen, K. G., Hagen, F. K., and Tabak, L. A. (1997) Charge distribution of flanking amino acids inhibits O-glycosylation of several single-site acceptors in vivo, *Glycobiology 7*, 1053-1060.
137. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3, *J Biol Chem 272*, 23503-23514.
138. Sutherlin, M. E., Nishimori, I., Caffrey, T., Bennett, E. P., Hassan, H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M. A. (1997) Expression of three UDP-N-acetyl-alpha-D-galactosamine:polypeptide GalNAc N-acetylgalactosaminyltransferases in adenocarcinoma cell lines, *Cancer Res 57*, 4744-4748.
139. Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., and Mandel, U. (2007) Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer, *Acta Histochem 109*, 273-284.
140. Baba, M., Yong Ma, B., Nonaka, M., Matsuishi, Y., Hirano, M., Nakamura, N., Kawasaki, N., and Kawasaki, T. (2007) Glycosylation-dependent interaction of Jacalin with CD45 induces T lymphocyte activation and Th1/Th2 cytokine secretion, *J Leukoc Biol 81*, 1002-1011.

141. Pirie-Shepherd, S. R., Stevens, R. D., Andon, N. L., Enghild, J. J., and Pizzo, S. V. (1997) Evidence for a novel O-linked sialylated trisaccharide on Ser-248 of human plasminogen 2, *J Biol Chem 272*, 7408-7411.
142. van Schooten, C. J., Denis, C. V., Lisman, T., Eikenboom, J. C., Leebeek, F. W., Goudemand, J., Fressinaud, E., van den Berg, H. M., de Groot, P. G., and Lenting, P. J. (2007) Variations in glycosylation of von Willebrand factor with O-linked sialylated T antigen are associated with its plasma levels, *Blood 109*, 2430-2437.
143. McDearmon, E. L., Combs, A. C., and Ervasti, J. M. (2003) Core 1 glycans on alpha-dystroglycan mediate laminin-induced acetylcholine receptor clustering but not laminin binding, *J Biol Chem 278*, 44868-44873.
144. Wang, C. S., Dashti, A., Jackson, K. W., Yeh, J. C., Cummings, R. D., and Tang, J. (1995) Isolation and characterization of human milk bile salt-activated lipase C-tail fragment, *Biochemistry 34*, 10639-10644.
145. Klar, A., Baldassare, M., and Jessell, T. M. (1992) F-spondin: a gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension, *Cell 69*, 95-110.
146. Tzarfaty-Majar, V., Lopez-Alemany, R., Feinstein, Y., Gombau, L., Goldshmidt, O., Soriano, E., Munoz-Canoves, P., and Klar, A. (2001) Plasmin-mediated release of the guidance molecule F-spondin from the extracellular matrix, *J Biol Chem 276*, 28233-28241.
147. Burstyn-Cohen, T., Tzarfaty, V., Frumkin, A., Feinstein, Y., Stoeckli, E., and Klar, A. (1999) F-Spondin is required for accurate pathfinding of commissural axons at the floor plate, *Neuron 23*, 233-246.
148. Feinstein, Y., Borrell, V., Garcia, C., Burstyn-Cohen, T., Tzarfaty, V., Frumkin, A., Nose, A., Okamoto, H., Higashijima, S., Soriano, E., and Klar, A. (1999) F-spondin and mindin: two structurally and functionally related genes expressed in the hippocampus that promote outgrowth of embryonic hippocampal neurons, *Development 126*, 3637-3648.

149. Burstyn-Cohen, T., Frumkin, A., Xu, Y. T., Scherer, S. S., and Klar, A. (1998) Accumulation of F-spondin in injured peripheral nerve promotes the outgrowth of sensory axons, *J Neurosci 18*, 8875-8885.

150. Zisman, S., Marom, K., Avraham, O., Rinsky-Halivni, L., Gai, U., Kligun, G., Tzarfaty-Majar, V., Suzuki, T., and Klar, A. (2007) Proteolysis and membrane capture of F-spondin generates combinatorial guidance cues from a single molecule, *J Cell Biol 178*, 1237-1249.

151. Debby-Brafman, A., Burstyn-Cohen, T., Klar, A., and Kalcheim, C. (1999) F-Spondin, expressed in somite regions avoided by neural crest cells, mediates inhibition of distinct somite domains to neural crest migration, *Neuron 22*, 475-488.

152. Hoe, H. S., Wessner, D., Beffert, U., Becker, A. G., Matsuoka, Y., and Rebeck, G. W. (2005) F-spondin interaction with the apolipoprotein E receptor ApoEr2 affects processing of amyloid precursor protein, *Mol Cell Biol 25*, 9259-9268.

153. Schubert, D., Klar, A., Park, M., Dargusch, R., and Fischer, W. H. (2006) F-spondin promotes nerve precursor differentiation, *J Neurochem 96*, 444-453.

154. Inatani, M. (2005) Molecular mechanisms of optic axon guidance, *Naturwissenschaften 92*, 549-561.

155. Blixt, O., Collins, B. E., van den Nieuwenhof, I. M., Crocker, P. R., and Paulson, J. C. (2003) Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein, *J Biol Chem 278*, 31007-31019.

156. Quarles, R. H. (2007) Myelin-associated glycoprotein (MAG): past, present and beyond, *J Neurochem 100*, 1431-1448.

157. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, *Gene 77*, 61-68.

158. Spiro, M. J., and Spiro, R. G. (1986) Control of N-linked carbohydrate unit synthesis in thyroid endoplasmic reticulum by membrane organization and dolichyl phosphate availability, *J Biol Chem 261*, 14725-14732.

159. Guo, R. T., Chou, L. J., Chen, Y. C., Chen, C. Y., Pari, K., Jen, C. J., Lo, S. J., Huang, S. L., Lee, C. Y., Chang, T. W., and Chaung, W. J. (2001) Expression in Pichia pastoris and characterization by circular dichroism and NMR of rhodostomin, *Proteins 43*, 499-508.

160. Cereghino, J. L., and Cregg, J. M. (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris, *FEMS Microbiol Rev 24*, 45-66.

161. Manabe, S., Marui, Y., and Ito, Y. (2003) Total synthesis of mannosyl tryptophan and its derivatives, *Chemistry 9*, 1435-1447.

162. Vielmetter, J., Stolze, B., Bonhoeffer, F., and Stuermer, C. A. (1990) In vitro assay to test differential substrate affinities of growing axons and migratory cells, *Exp Brain Res 81*, 283-287.

Acknowledgements

I would like to thank my supervisor PD Dr. Jan Hofsteenge who gave me the opportunity to do my thesis in his laboratory. I would like to express my gratitude for his technical assistance, the help and confidence in my work. Additionally I am also grateful to Mrs. Nancy Hynes and Mr. Kurt Ballmer-Hofer for serving in my thesis committee.

A special thanks to all the people from the Hofsteenge lab for their friendship, helpfulness, discussions and the wonderful time: Constanze Heinrich, Dominique Klein, Claudia Keller, Chun-I Chen, Matthias Schmid, Krisztina Keusch, Jeremy Keusch, Ragna Sack, Daniel Hess, Reto Portmann, Stefano Carnevascini

Also I would like to acknowledge all people who have contributed to the success of this work and who are not listed in detail.

Finally I am indebted to my family and to my wife who encouraged me during all the years and made my PhD possible.

Curriculum Vitae

PERSONAL DATA

Name:	Carsten Krantz
Address:	Muesmattweg 112a 4123 Allschwil Switzerland
Phone:	mobile: 0049-176 674 104 74 residence: 0041-61 711 04 66
E-mail:	carsten.krantz@fmi.ch
Date of birth:	22nd February 1974
Nationality:	German
Marital status:	Married

EDUCATION

PhD in Cell Biology:	September 2003 – March 2008 Friedrich-Miescher Institute for Biomedical Research, Basel Switzerland (supervised by PD Dr. Jan Hofsteenge, Prof. Dr. Nancy Hynes and Prof Dr. Kurt Ballmer) Topic: Glycosylation of Thrombospondin type I repeats
Masters in Biotechnology:	October 2001- April 2003: University of Applied Sciences, Mannheim, Germany Topic Master Thesis: Soluble $A\beta_{1-42}$ oligomers characterization and interacting proteins (supervised by Dr. Heinz Hillen and ProF Winfried Storhas)
Diplom Ingenieur (FH) in Biotechnology:	March 1996 – June 2001 Fachhochschule Mannheim, Germany Topic Diploma Thesis: Investigations of unexpected proteins in the surface proteom of the human pathogen *Listeria monocytogenes (*supervised by Dr. Lothar Jaensch and Prof Dr. Heinz Trasch)
Emergency Medical Technician (EMT):	September 1994 – March 1996 Social service at Johanniter Unfallhilfe Giessen and training as paramedic
Secondary School:	August 1985 – June 1994 Theo-Koch-Schule Grünberg Graduation with A-levels

PUBLICATIONS AND PATENTS

1.) Amyloid-β(l-42) oligomers, derivatives thereof, antibodies for the same, method for production and use thereof. Hillen H., Striebinger A., Krantz C., Moeller A. Mueller R. *International Patent WO, 2004/067561 A1*

2.) Globular amyloid β-peptide$_{1-42}$ oligomer- a homogenous and stable neuropathological protein in Alzheimer's disease. Barghorn S., Nimmrich V., Striebinger A., Krantz C., Keller P., Janson B., Bahr M., Schmidt M., Bitner RS., Harlan J., Barlow E., Ebert U. and Hillen H. *Journal of Neurochemistry, 2005, 95, 834–847*

3.) Alternative glycosylation of Thrombospondin Type I repeats. Krantz C., Klein D., Sack R., Karo K., Hess D., Clausen H., and Hofsteenge J.; *in preparation*