The EMBO Journal (2012) 31, 3183–3197 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

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

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

Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

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

The EMBO Journal (2012) **31,** 3183–3197. doi:10.1038/emboj.2012.143; Published online 15 May 2012
Subject Categories: proteins
Keywords: crystal structure; enzymatic mechanism; GDP-fucose; protein *O*-fucosyltransferase 2; thrombospondin type 1 repeat

Introduction

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

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

Received: 12 January 2012; accepted: 23 April 2012; published online: 15 May 2012

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

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

that POFUT2-dependent *O*-fucosylation on TSR might be required for optimal secretion of these proteins. No disorder has yet been directly linked to a genetic defect of the *Pofut2* locus in humans. However, mutations in the *B3GALTL* gene that encodes the β3GlcT enzyme responsible for glucose transfer onto *O*-fucosylated TSR cause the autosomal recessive disorder Peters Plus syndrome (Lesnik Oberstein *et al*, 2006; Hess *et al*, 2008). This disorder is characterized by anterior-eye-chamber abnormalities, disproportionate short stature and developmental delay.

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

Results

Crystal structure of human POFUT2

We have expressed and purified human Δ21-POFUT2 from mammalian cell culture and have determined its crystal structure at 3.0 Å resolution. The protein crystallized in space group $P3_221$ with two molecules in the asymmetric unit (a.u.) and the structure was solved by the single isomorphous replacement with anomalous scattering (SIRAS) method using a platinum derivative. Data collection, phasing and refinement statistics are presented in Table I. The refined POFUT2 crystal structure displays clear electron density for residues 41–429 (out of 22–429) and the two molecules in the a.u. are almost identical with an r.m.s.d. of only 0.49 Å. The structure of POFUT2 is composed of two Rossmann-like domains with β/α/β topology typical of the GT-B fold of glycosyltransferases (Figure 1A). N- and C-terminal domains encompass residues 22–242 and 243–429, respectively. The two domains interact closely with each other (buried surface area of 1416 $Å^2$) forming an extended protein unit. Fully structured loops originating from both the N-terminal (Q141–V156, E158–N189) and C-terminal domain (T407–Y429, L293–L309) form a large central cavity in the molecule with two disulphide bonds stabilizing loop conformations in each domain (C161–C192 and C412–C419). A second narrower cleft is present in the C-terminal domain, formed by helices α13 and α14, loop Q93–Q99 and the N-terminal tip of helix α1 (E54–N57). Electron density for three N-acetylglucosamine (GlcNAc) moieties is present at residues N189, 209, and 259 revealing all predicted N-glycosylation sites occupied. The quality of the electron density allowed model building of GlcNAc moieties at N189 and N259.

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

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

Table I Data collection and refinement statistics

	POFUT2 native	POFUT2 Pt derivative	POFUT2 GDP-fucose complex
Data collection			
Space group	$P3_221$	$P3_221$	$P3_221$
Cell constants *a*, *b*, *c* (Å)	118.6, 118.6, 196.2	118.5, 118.5, 195.0	153.0, 153.0, 185.7
Wavelength λ (Å)	1.000	0.890	1.000
Resolution range (Å)[a]	30.0–3.0 (3.11–3.00)	20.0–5.5 (5.69–5.50)	40.0–3.4 (3.63–3.40)
Unique reflections	31 375	9737	32 751
Completeness (%)[a]	96.4 (71.6)	100 (99.9)	93.3 (93.6)
Multiplicity	11.3	10.4	4.7
R_{sym} (%)[a,b]	12.4 (39.7)	22.4 (43.4)	16.6 (76.6)
$I/\sigma(I)$[a]	21.0 (2.4)	11.3 (3.7)	11.9 (2.1)
Phasing power iso/ano		1.04/0.53	
Refinement			
Resolution range (Å)	30.0–3.0		40.0–3.4
Reflections (all)	31 318		32 745
Reflections (test set)	1593 (5.1%)		1622 (5.0)
R_{crys} (%)	17.4		19.3
R_{free} (%)	23.6		23.8
r.m.s.d.			
Bond lengths (Å)	0.008		0.011
Bond angles (deg)	1.28		1.31
Wilson *B*-factor (Å^2)	50.9		—
Mean B-factor (Å^2)			
Protein	54.6		102.2
Ligand	—		113.4
Ramachandran plot (%)			
Favoured	97.0		97.0
Allowed	3.0		3.0
Outliers	0		0

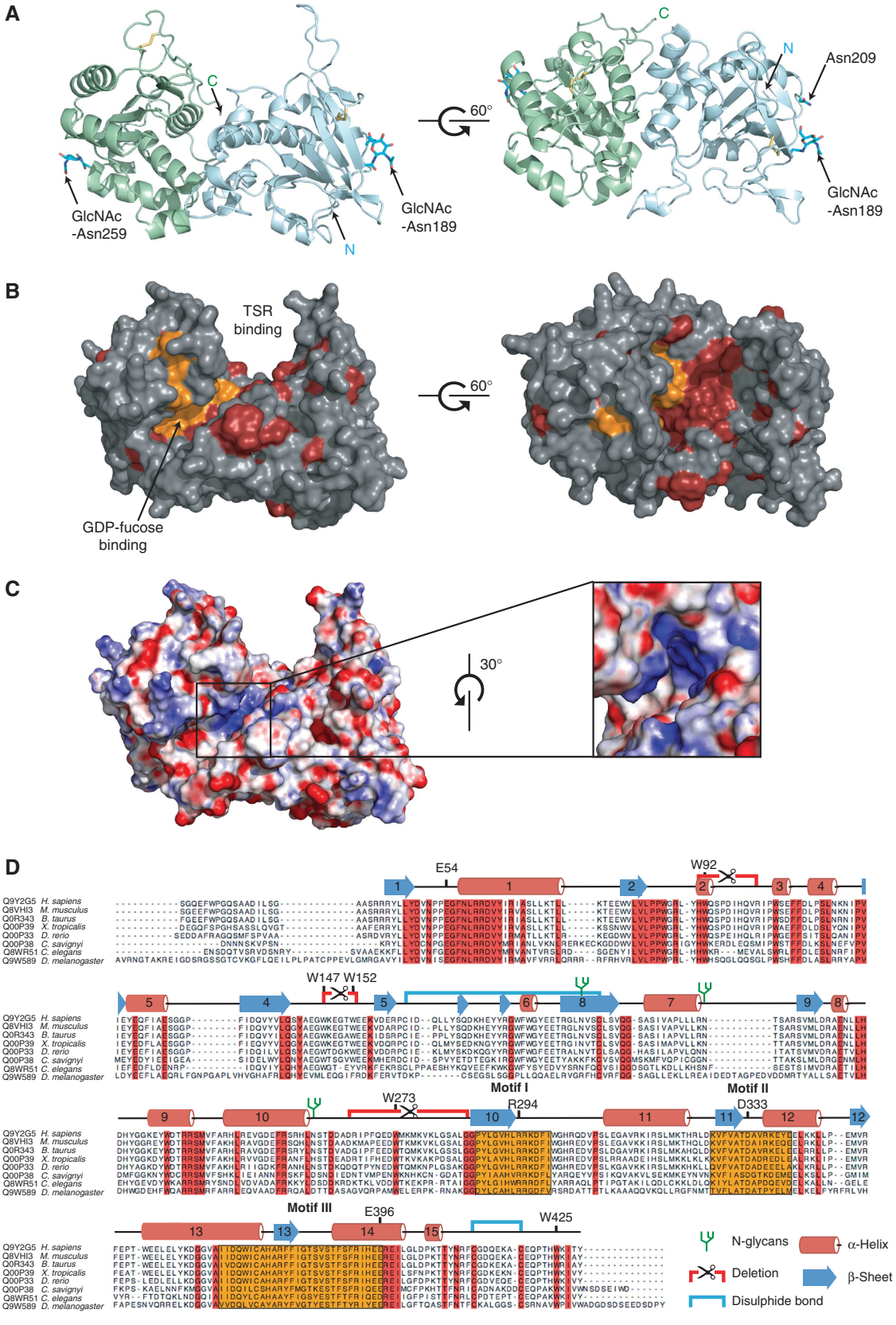
[a]Values in parentheses refer to the highest resolution shell.
[b]$R_{sym} = \Sigma_{hkl}\Sigma_j|I_{j,hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_j I_{j,hkl}$ where $\langle I_{hkl}\rangle$ is the average of the intensity $I_{j,hkl}$ over $j = 1,\ldots, N$ observations of symmetry equivalent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

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

POFUT2 enzymatic activity

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

respectively, and a k_{cat} of 144 per minute. Based on our structural results, we designed 14 mutations targeting specific putative functional residues of the enzyme and tested enzymatic activity (Figure 3D and E). Out of 14 mutations, 2 did not yield any soluble protein pointing to a critical function of these residues in the folding pathway of the protein (D333A and Δ265–285). All other mutants expressed and purified well and equal amounts were used for the activity assay.



From the POFUT2 mutants targeting the catalytic mechanism, E54A and R294A resulted in complete loss of activity while the D297A and E396A mutants remained active (15 and 8%, respectively). A change of the highly conserved W92 to alanine as well as deletion of the entire loop (Δ90–100) abolished enzymatic activity. POFUT2 features a unique loop (265–285) located on the opposite side of the large cleft (Figure 3D), which protrudes from the C-terminal domain and which is attached to the N-terminal domain *via* a completely conserved tryptophan residue. We hypothesized that this residue, W273, is involved in controlling movements of the N- and C-terminal domain relative to each other during the catalytic cycle and indeed lost 90% activity when we mutated W273 to alanine. A series of mutations targeted the large loop forming one wall of the central cavity (residues 147–152) with the aim of disturbing TSR binding. While the point mutations reduced the catalytic activity to ~10–73%, removal of the entire loop (Δ147–152) yielded a completely inactive enzyme. Alanine mutation of another conserved residue (W425) located at the entry of the central cavity and potentially involved in TSR binding also reduced POFUT2 activity to 38%.

Crystal structure of human POFUT2 in complex with GDP-fucose

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

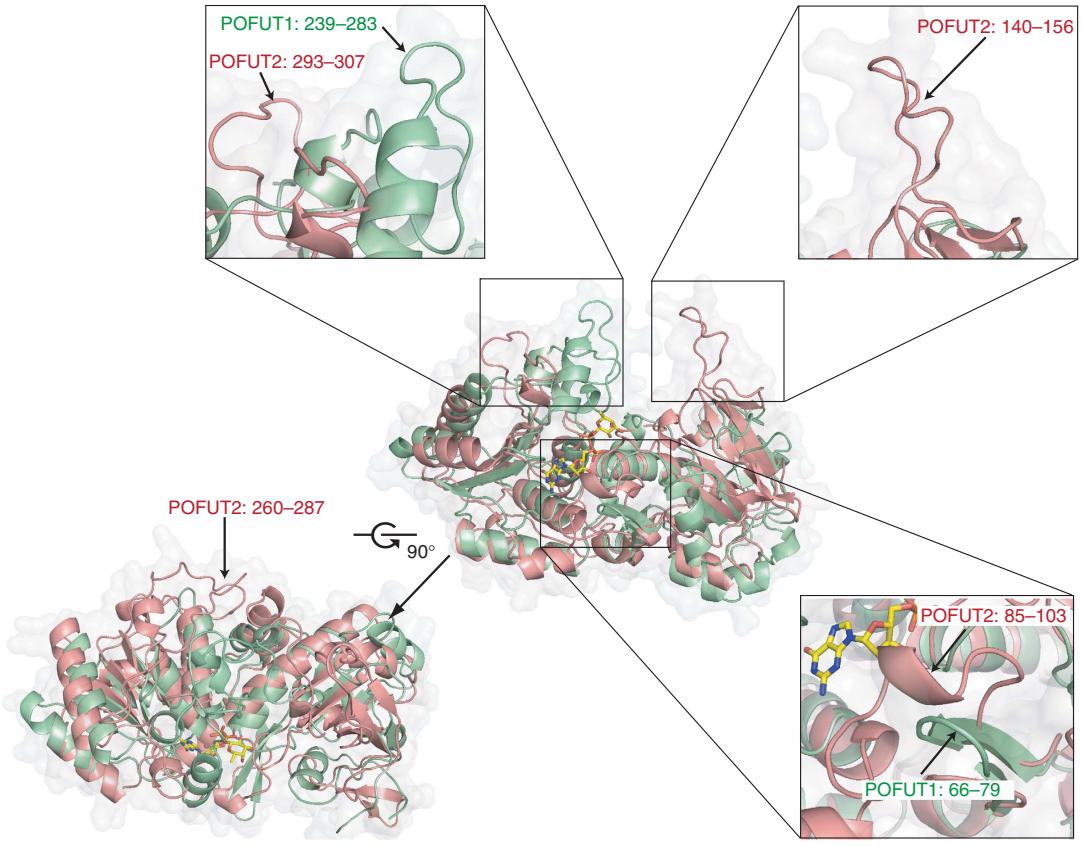


Figure 2 Structural superposition of human POFUT2 and *C. elegans* POFUT1 in two orientations rotated by 90°. POFUT2 (red) and POFUT1 (green) are displayed as cartoon models with transparent surfaces. GDP-fucose bound to POFUT1 is shown as sticks (yellow and atom colours). Main structural differences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

Figure 1 Crystal structure of human POFUT2 and multiple sequence alignment of orthologue sequences. (**A**) The human POFUT2 structure is displayed as cartoon model in two orientations 60° apart. N- and C-terminal domains are in grey and green, respectively, N- and C-termini are labelled. Disulphides and covalently bound GlcNAc molecules are displayed as sticks in atom colours. (**B**) Conserved residues from the multiple POFUT2 sequence alignment (**D**) are mapped onto the surface of the human POFUT2 structure (100% conservation: red, conserved motifs I–III among fucosyltransferases: orange). Substrate binding sites are labelled. (**C**) Mapping of the electrostatic surface potential onto the surface of the POFUT2 structure (scale: −20 to +20 kT/e from red to blue). The highly positive surface patch involved in GDP-fucose binding is boxed, zoomed-in and rotated by 30° in the right image. Computed with the APBS plugin of PyMol (Baker *et al*, 2001). (**D**) Multiple sequence alignment of selected POFUT2 sequences (ClustalW; Larkin *et al*, 2007). Structural features present in the POFUT2 structure and mutated residues are indicated.

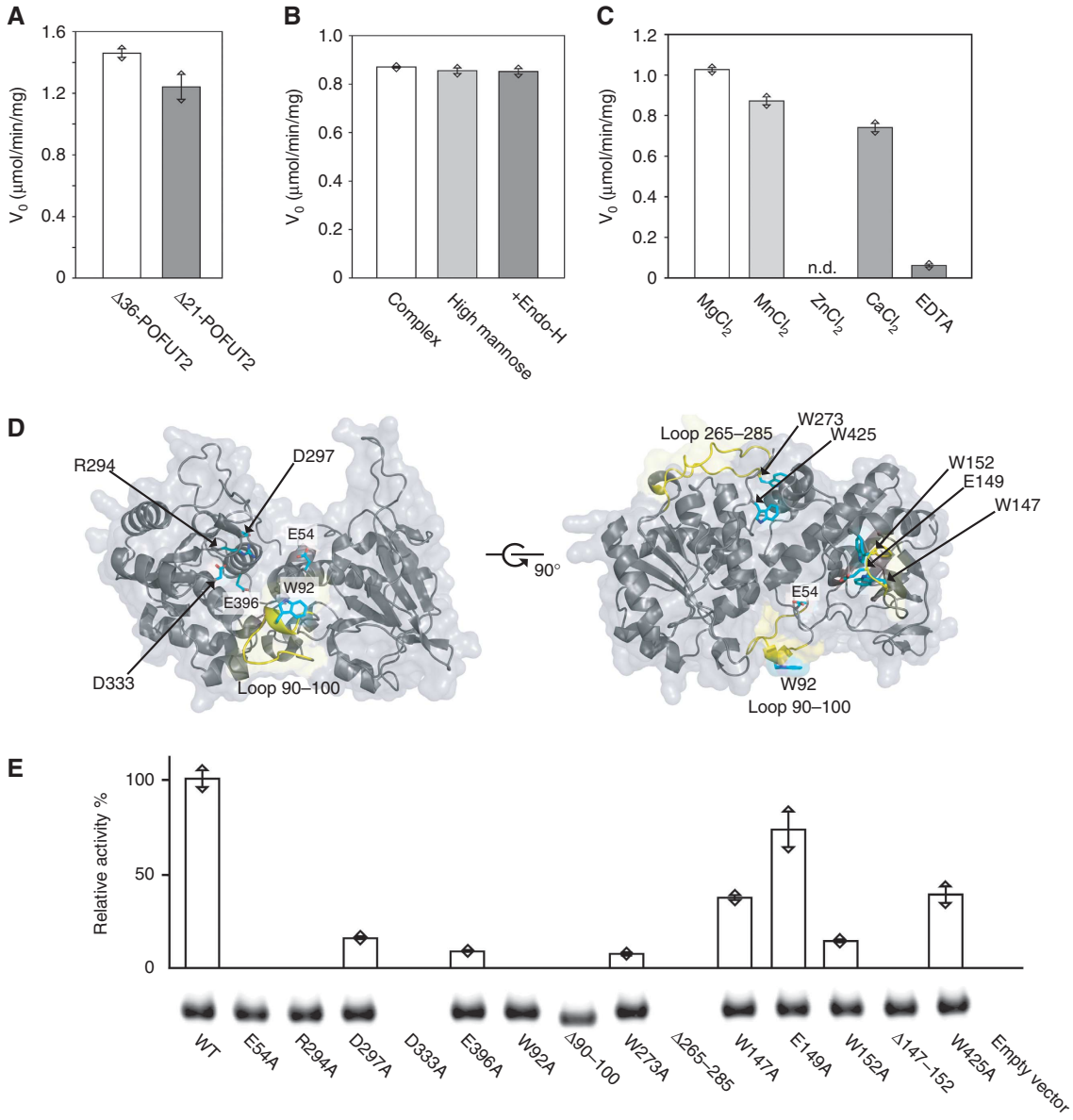


Figure 3 Initial rate of fucosylation of purified wild-type and mutant POFUT2. (**A**) Influence of different N-terminal POFUT2 truncations. (**B**) Influence of the glycan composition (complex, high mannose, + Endo-H). (**C**) Presence or absence of varying divalent metal ions (n.d.: not detected). Columns in (**A–C**) represent the average of two independent measurements while individual measurements are shown as triangles. (**D**) POFUT2 mutations used in the enzymatic *in-vitro* assay. POFUT2 is shown as a cartoon model with transparent surface in two orientations. Point mutations (sticks, cyan) and deletions (yellow) are indicated. (**E**) Relative enzymatic activity of different point and deletion mutations compared with the wild-type (WT) protein. Experiments were repeated twice and the mean of the two measurements and the individual measurements are shown. The western blot signal used for quantification of the enzyme input is indicated.

crystal packing revealed the presence of a crystallographic dimer similarly to the non-crystallographic-symmetry dimer present in the a.u. of the apo structure but with a reduced interface enabling access to the GDP-fucose binding site (Supplementary Figure S3). Clear electron density for GDP-fucose was present in all four molecules in the a.u. (space group $P3_221$, 3.4 Å resolution) and located at the predicted position in the narrow cleft leading from the N- and C-terminal domain interface into the C-terminal domain (Figure 4A). The guanine purine base is mainly held in place by stacking interactions with F389 and hydrogen bonds between the N1 nitrogen and the D371 side chain while other residues of the pocket additionally contribute hydrophobic interactions (Figure 4B). The ribose moiety bulges up from the bottom of the cleft and does not show any tight interaction with the protein. Instead, the main affinity for the sugar donor comes from the interaction of the diphosphate group with the protein. The guanidinium moiety of R294 forms a salt bridge with the β-phosphate while the positive dipole located at the N-terminal end of the last helix (387–400) tightly attaches the diphosphate moiety to the helix tip where it hydrogen bonds side chain (T388) and backbone atoms of residues T388 and F389. Strikingly, the fucose is arranged almost perpendicular to the nucleotide diphosphate *via* hydrogen bonds between the O3 hydroxyl and the P53 carbonyl group and the O2 hydroxyl and the G55 amide nitrogen of the N-terminal domain. This arrangement presents the activated phosphoester bond at the anomeric

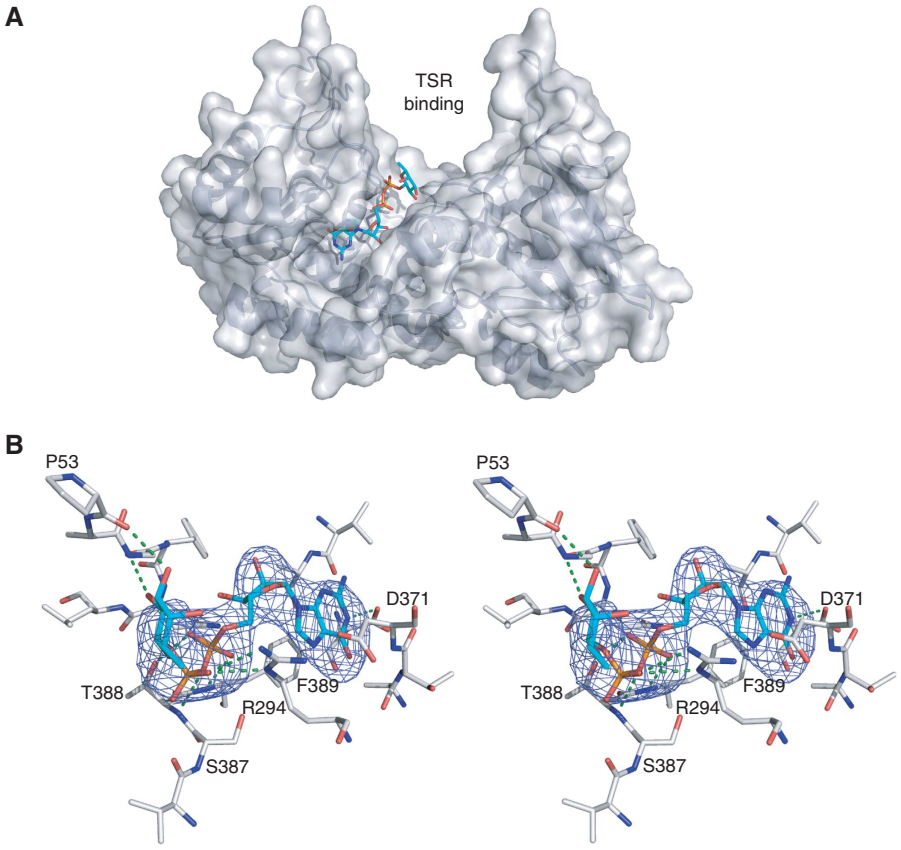


Figure 4 GDP-fucose binding in human POFUT2. (**A**) Structure of the POFUT2 GDP-fucose complex displayed as a cartoon model in grey with transparent surface. GDP-fucose is shown as sticks in atom colours. (**B**) Stereo figure of the GDP-fucose (cyan sticks) binding mode in POFUT2 (white sticks). NCS averaged mF_o-DF_c electron density calculated after molecular replacement and one round of refinement (in absence of GDP-fucose) is shown as mesh in blue (3σ). Hydrogen bonds are displayed as green dotted lines. The orientation is different than in (**A**) for better clarity.

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

Structural restraints in TSR for productive fucose attachment

Our structure of the POFUT2 GDP-fucose complex together with the structures of fucosylated TSR2-TSR3 of human TSP-1 (Tan *et al*, 2002) and fucosylated TSR1 of ADAMTS13 (Akiyama *et al*, 2009) enabled us to build a model of the full enzyme substrate ternary complex. Superposition of anomeric fucose carbons in these structures with the anomeric carbon in the POFUT2 GDP-fucose complex, followed by manual adjustment, yielded the overall TSR position. This initial model was used to overlay the TSR4 structure from rat F-spondin (PDB 1VEX) (Pääkkönen *et al*, 2006) and energy minimization of the full complex in CNS (Brunger *et al*, 1998). We obtained a plausible model of the ternary complex in which the elongated TSR unit lies in the deep interdomain cavity of POFUT2 spanning across the glycosyltransferase (Figure 5A). The TSR module contacts the highly conserved POFUT2 surface *via* its flat hydrophobic side, opposite of the SS-bond pattern and the tryptophan-arginine stacking (CWR layer), where the B and C strand show a regular antiparallel β-sheet. In addition, the rippled A strand contacts the bottom and the side wall of the cleft. The entire AB loop harbouring the $CX_{2-3}(S/T)CX_2G$ motif is in close contact with POFUT2. Only half of the TSR module was predicted to interact with POFUT2 whereas the N-terminal

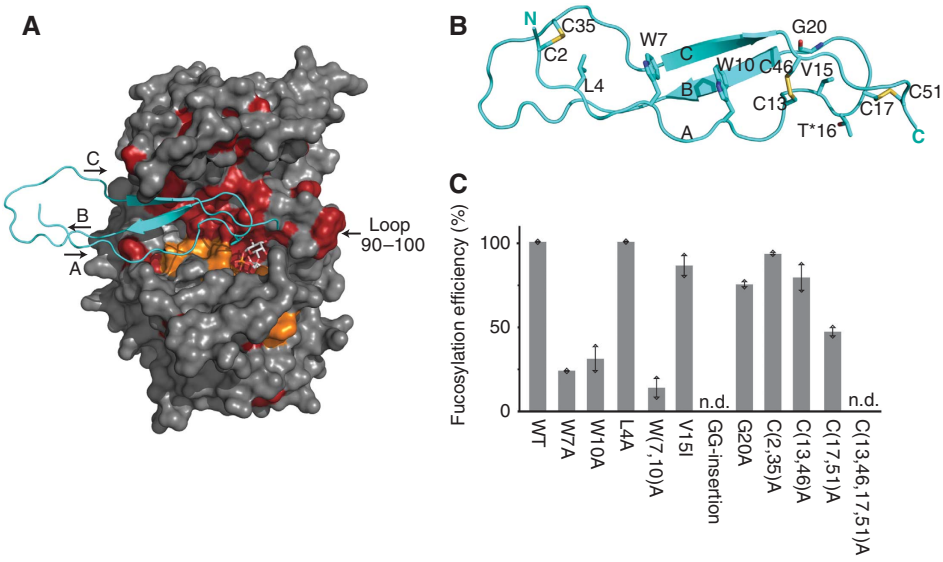


Figure 5 Model of the POFUT2 GDP-fucose TSR complex and POFUT2 enzymatic activity using mutant TSR as substrate. (**A**) Surface representation of the POFUT2 GDP-fucose structure with POFUT2 residues coloured as in Figure 1B and GDP-fucose as sticks. TSR4 from rat F-spondin (PDB 1VEX) is displayed as a cartoon model in cyan. (**B**) Structure of TSR4 from rat F-spondin. Residues that were mutated for fucosylation measurements are numbered. (**C**) Relative fucosylation efficiency of TSR mutants. Wild-type and mutant TSR was expressed and purified from HEK293T cells and analysed by mass spectrometry for fucosylation efficiency. n.d.: not detected. Experiments were repeated twice and the mean of two measurements and the individual measurements are displayed.

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

In order to validate our model of TSR binding, we expressed a series of F-spondin TSR4 mutants in HEK293T cells, purified the secreted protein from the medium and analysed its fucosylation state by mass spectrometry (Figure 5B and C). Changes to the conserved W*XXW*$_7$*XX*W$_{10}$ motif in the A strand (in the case of TSR4 of F-spondin L$_4$*XX*W$_7$*XX*W$_{10}$), the key element of the TSR fold that forms the multi-layered delocalized π-system with the conserved arginine residues from the B strand (WR of CWR), reduced the efficiency of fucosylation significantly. Single mutations of W7 and W10 to alanine produced only 24 and 31% of fucosylated TSR4, double mutations on both sites further reduced the fucosylation to 14% compared with wild type whereas the L4A mutation did not have any influence. Next, we introduced mutations in the AB loop close to the threonine (T*16) that undergoes fucosylation: replacing the valine (V15) next to it with an isoleucine or changing the glycine (G20) at the end of the turn to alanine had only minor effects (86, 75% fucosylation). On the other hand, when we introduced two additional glycine residues right before the threonine (*GG*T*, mimicking the fucosylation sequon in EGF repeats) to create a larger loop between the disulphide forming cysteine and the threonine, fucosylation was completely abolished. The disulphide bond pattern is a hallmark in the fold of TSR therefore we investigated how fucosylation is affected by removing SS-bonds. Mutation of the cysteines forming the SS-bond between the A and C strand, where we postulated no interaction with POFUT2, was well tolerated with no reduction in fucosylation levels (C2,35A). To our surprise, the removal of the second SS-bond between the A and C strand had no dramatic effect with 80% fucosylated product (C13,46A). On the other hand, removal of the SS-bond that connects the AB loop (C17,51A) with the C-terminal end of the C strand resulted in 55% reduced fucosylation levels. When we removed both SS-bonds, which together stabilize the 3D conformation of the AB loop, we could not detect any fucosylated TSR product.

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

Consequently, we investigated whether a minimal TSR module containing only seven conserved structural residues of the TSR hallmark elements (disulphide pattern and side-chain stacking) would be sufficient as a POFUT2 sugar acceptor. We designed and expressed a truncated TSR module based on F-spondin TSR4 consisting of approximately half

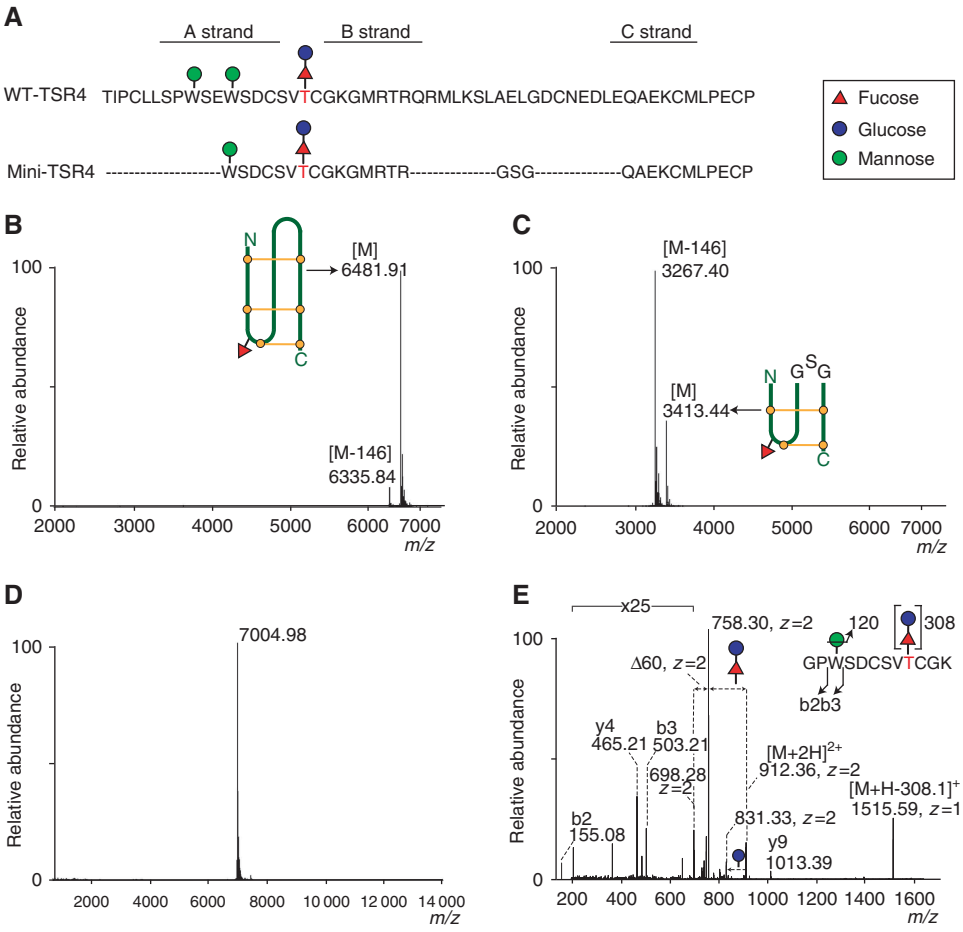


Figure 6 *O*-fucosylation of mini-TSR. (**A**) Amino-acid sequences of wild-type TSR4 from rat F-spondin and engineered mini-TSR are compared. Amino acids that form the three strands are indicated. Wild-type TSR and mini-TSR were expressed and isolated from *E. coli*. These modules were used as acceptor substrates in the POFUT2 *in-vitro* assay and the reaction products analysed by mass spectrometry. Both the wild-type TSR (**B**) and mini-TSR (**C**) show two peaks that differ by 146 Da, indicating the presence or absence of one fucose molecule. Wild-type and mini-TSR are drawn as cartoons with disulphide bonds in yellow and fucose as a red triangle. Mini-TSR is approximately half the size of wild-type TSR, with a truncated A strand and the deleted BC loop replaced by a three residue linker, GSG. (**D**) An N-terminal His_6–FLAG-3C-tagged mini-TSR was expressed in mammalian HEK293T cells, isolated from the medium and analysed by mass spectrometry. The mono-isotopic mass (7004.98) represents the intact, tagged mini-TSR with two hexoses and one deoxyhexose. (**E**) Secreted and purified mini-TSR from HEK293T cell expression was digested with human rhinovirus 3C and Lys-C protease and analysed by mass spectrometry. MS/MS analysis confirmed the sequence (GPWSDCSVTCGK) and revealed the glycan modifications. Loss of one hexose and one hexose-deoxyhexose disaccharide from the parent ion (m/z 912.36) was observed in the MS spectrum (m/z 831.33 and 758.30, respectively) and the mass difference between the b2 and b3 ion corresponded to one hexose molecule attached to the tryptophan. The characteristic −120 Da fragment (m/z 698.28) represents the typical signature motif for *C*-hexosylation.

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

Wild-type TSR4 and mini-TSR were expressed and isolated from *E. coli* and used as acceptor substrates for purified POFUT2 in a qualitative *in-vitro* fucosyltransferase assay. The reaction products were monitored for fucose incorporation by mass spectrometry. An increase in mass of 146 Da corresponding to the addition of one fucose moiety was observed for both the wild-type (Figure 6B) and mini-TSR substrate (Figure 6C). Fucosylation was only observed in reactions that included POFUT2. Thus, POFUT2 recognizes and modifies both wild-type and truncated artificial mini-TSR. We further validated our findings by expressing mini-TSR with an N-terminal His–FLAG tag in HEK293T cells. The secreted mini-TSR protein was purified from the culture medium and analysed by mass spectrometry. We found a homogenous species of 7004.98 Da corresponding to the tagged mini-TSR with two hexoses and one deoxyhexose attached (Figure 6D). To further analyse the glycosylation state of the mini-TSR, secreted and purified protein was digested with human rhinovirus 3C to remove the tag, reduced and alkylated, and digested with Lys-C protease yielding a peptide covering the predicted glycosylation sites. MS/MS analysis confirmed the sequence (GPWSDCSVTCGK) and revealed the glycan modifications (Figure 6E). Loss of one hexose and one hexose-deoxyhexose disaccharide from the parent ion (m/z 912.36) was observed in the MS spectrum (m/z 831.33 and 758.30, respectively). In addition, the mass difference between the b2 and b3 ion corresponded to one

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

Discussion

POFUT2 protein structure and TSR substrate recognition

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

Our data suggest that POFUT2 recognizes key 3D structural TSR elements formed by the disulphide pattern and side-chain stacking common to sequence-wise degenerated TSR modules. A search of the PDB using DALI with coordinates of a minimal TSR poly-alanine module (regions predicted to interact with POFUT2) identified only known structures of TSR domains without discovering this structural motif in any other protein. Therefore, substrate specificity seems to be achieved by the structural complementarity of a part of the TSR fold with the POFUT2 binding site and the wide TSR sequence diversity does not play a role as long as the critical TSR fold motif is intact. From our experimental data and the structural models, we conclude that disrupting the conformation of the rippled A strand (formed by the W*XX*W*XX*W motif, L*XX*W*XX*W in TSR4 from F-spondin) and of the AB loop (defined by the second and third disulphides) impairs TSR substrate recognition and fucosylation efficiency (Figure 5C). Our data of the mini-TSR also show that starting the A strand directly at the third tryptophan of the W*XX*W*XX*W motif has no negative effect on fucosylation as the conformation of the shortened A strand is still intact. Proposed key interactions are located at the entry of the large TSR binding cavity where conserved POFUT2 residues W152 from the N-terminal domain and W425 from the very C-terminal part define the most narrow part of the cleft (∼15 Å) allowing only space for a two stranded β-sheet to enter the cavity (Figure 7A). C-terminal residues seem to lock the position of the bound TSR module by interacting with the backbone bulge formed by the second and third tryptophan of the L*XX*W*XX*W motif of the rippled A strand. Thereby, C-terminal POFUT2 residues act as a ruler to position the S/T residue undergoing modification exactly at the right position for E54-dependent deprotonation and nucleophilic attack at the anomeric GDP-fucose carbon. This model is supported by our observation that fucose attachment is reduced by 90 and 61% for the W152A and W425A mutation, respectively (Figure 3E), and that exchange of the second and third F-spondin TSR4 tryptophan (responsible for bulge formation) to alanine results in 75 and 69% reduction of fucosylated product, respectively (W_7, W_{10} in Figure 5C). The second key interaction is predicted to take place between the AB loop (where the consensus motif $CX_{2-3}(S/T)CX_2G$ is located) and the conserved POFUT2 residues Asn51, Pro52, Pro53, Glu54, Leu58, Asp61, and Glu221 (Figure 7A). These residues, together with L224 that inserts its side chain exactly where the C strand starts to crossover the B strand, ensure *via* TSR backbone interactions that only the unique 3D motif at the very tip of the TSR module can undergo fucosylation. This structural motif is mainly defined by the length and conformation of the AB loop, which is held in place by the two disulphide bonds and which is encoded in the $CX_{2-3}(S/T)CX_2G$ sequon. Our observation explains experimental data where insertion of two additional glycines before the threonine (CX_{2-3}**GG**$(S/T)CX_2G$, changing the length of the loop) as well as removing the two SS-bonds, which together are responsible for pulling the C strand over the B strand and stabilization of the AB-loop conformation, completely abolished fucose attachment (Figure 5C).

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

Having realized that POFUT2 substrate recognition is likely to be driven by the conserved TSR residues responsible for the unique layered TSR fold, we wondered how the substrate binding site is able to accept the large charge and size variation of amino acids on the remaining ∼40 sequence positions. Strikingly, we found in our model of the POFUT2 TSR complex that out of the ∼30 TSR residues building the upper half of the TSR (predicted to interact with POFUT2) 10 are conserved determining the TSR fold (in the central layer of stacking residues and SS-bonds) or are part of the consensus motif for fucosylation. Another nine residues are solvent exposed and likely not involved in POFUT2 interaction. At the 11 remaining TSR positions where wide sequence diversity is present we find large cavities in POFUT2, ready to accommodate side chains of different lengths or with different physicochemical properties (Figure 7B).

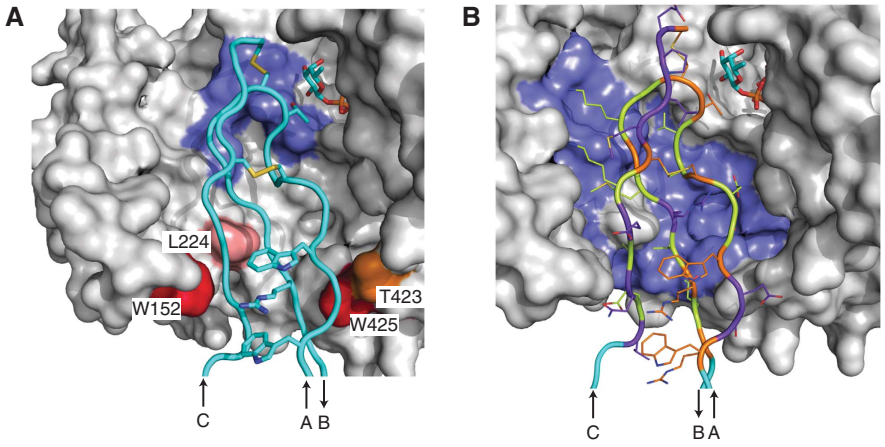


Figure 7 Structural details of the proposed TSR binding mode in human POFUT2. (**A**) Surface representation (grey) of POFUT2 with bound TSR4 (F-spondin) as ribbons (cyan) and the GDP-fucose substrate as sticks. POFUT2 residues predicted to recognize key elements of the TSR fold are coloured: W152 and W425 (red) scan the width of the substrate; T423 (orange) senses the bulge in the A strand and acts as a ruler; L224 (pink) scans the crossing over of the C strand; N51, P52, P53, E54, L58, D61, E221 (all in blue) recognize the correct conformation of the AB loop. Key structural residues of the TSR fold are displayed as sticks in cyan (atom colours). (**B**) The wide TSR sequence diversity of POFUT2 substrates can be explained by the proposed TSR binding mode. POFUT2 provides large cavities (solid surface, blue) for highly variable TSR residues (green) thereby tolerating side chains of different size and physicochemical properties at these positions. TSR side chains can also vary at positions predicted to be solvent exposed (purple) while TSR residues important for the fold and the fucosylation motif are strictly conserved (orange).

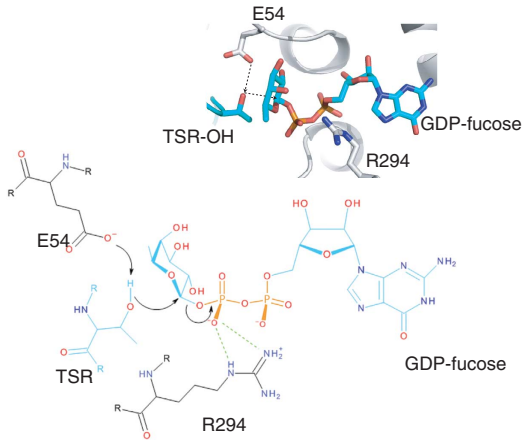


Figure 8 Proposed enzymatic mechanism for POFUT2-dependent TSR fucosylation. Chemical drawing of the enzymatic reaction. The carboxylate side chain of E54 deprotonates the TSR S/T hydroxyl group that attacks as a nucleophile the anomeric carbon of the GDP-fucose sugar donor. R294 activates the labile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

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

Catalytic mechanism of POFUT2

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

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

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

Substrate specificity of protein glycosyltransferases

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

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

Materials and methods

A detailed description of expression and purification of wild-type and mutant POFUT2 and TSR proteins, enzymatic assays and detection of TSR fucosylation states by mass spectrometry is given in Supplementary data.

Crystallization, data collection and structure determination

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

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

Supplementary data

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

Acknowledgements

We thank Ragna Sack from the Protein Analysis Facility for support in mass spectrometry experiments and the staff at the Swiss Light Source (Villigen, Switzerland) for support in X-ray data collection. The Friedrich Miescher Institute for Biomedical Research is a part of the Novartis Research Foundation.

Author contributions: HG, CC, and JH designed the experiments. CC, JK, DK, DH, JH, and HG carried out experiments and analysed the data. HG and CC wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

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

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

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

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

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

Bjoern S, Foster DC, Thim L, Wiberg FC, Christensen M, Komiyama Y, Pedersen AH, Kisiel W (1991) Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine. *J Biol Chem* **266:** 11051–11057

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wang LW, Dlugosz M, Somerville RPT, Raed M, Haltiwanger RS, Apte SS (2007) O-fucosylation of thrombospondin type 1 repeats in ADAMTS-like-1/Punctin-1 regulates secretion: implications for the ADAMTS superfamily. *J Biol Chem* **282:** 17024–17031

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