ARTICLE
PUBLISHED ONLINE: XX XX 2016 | DOI: 10.1038/NCHEMBIO.2019

A proactive role of water molecules in acceptor recognition by protein *O*-fucosyltransferase 2

Jessika Valero-González[1,10], Christina Leonhard-Melief[2,10], Erandi Lira-Navarrete[1], Gonzalo Jiménez-Osés[1,3,4], Cristina Hernández-Ruiz[1], María Carmen Pallarés[5], Inmaculada Yruela[6], Deepika Vasudevan[2], Anabel Lostao[5,7], Francisco Corzana[4], Hideyuki Takeuchi[2,9], Robert S Haltiwanger[2,9] & Ramon Hurtado-Guerrero[1,7,8]*

Protein *O*-fucosyltransferase 2 (POFUT2) is an essential enzyme that fucosylates serine and threonine residues of folded thrombospondin type 1 repeats (TSRs). To date, the mechanism by which this enzyme recognizes very dissimilar TSRs has been unclear. By engineering a fusion protein, we report the crystal structure of *Caenorhabditis elegans* POFUT2 (*Ce*POFUT2) in complex with GDP and human TSR1 that suggests an inverting mechanism for fucose transfer assisted by a catalytic base and shows that nearly half of the TSR1 is embraced by *Ce*POFUT2. A small number of direct interactions and a large network of water molecules maintain the complex. Site-directed mutagenesis demonstrates that POFUT2 fucosylates threonine preferentially over serine and relies on folded TSRs containing the minimal consensus sequence C-X-X-S/T-C. Crystallographic and mutagenesis data, together with atomic-level simulations, uncover a binding mechanism by which POFUT2 promiscuously recognizes the structural fingerprint of poorly homologous TSRs through a dynamic network of water-mediated interactions.

Protein glycosylation is the most frequent and complex post-translational modification in nature. The reaction in eukaryotes is carried out by a limited number of glycosyltransferases (GTs) that specifically recognize acceptor substrates and catalyze transfer from different activated sugar donors[1,2]. Whereas many of these GTs have the capacity to glycosylate unstructured peptides, only four GTs have been proven to require disulfide bridge–containing folded domains to achieve glycosylation[3,4]. Of these GTs, POFUT1 and POFUT2 have been extensively studied. These inverting enzymes are classified as GT65 and GT68, respectively, in the CAZy database[5]. POFUT1 and POFUT2, located in the endoplasmic reticulum (ER), fucosylate epidermal growth factor-like (EGF) repeats and TSRs, respectively, which are present in the extracellular domains of a large number of proteins (such as Notch, thrombospondin 1 and ADAMTS13)[6]. O-fucosylation is an essential biological modification that is important not only as an ER quality-control mechanism (owing to its influence in the correct folding and stability of EGF repeats and TSRs) but also for protein-protein interaction[7,8]. In particular, glycosylation of Notch EGF repeats is important for Notch-ligand interactions[9,10]. The molecular aspects of this interaction were recently uncovered, showing that *O*-fucose of EGF12 acts as a crucial amino acid surrogate interacting with Notch-ligand delta-like 4 (ref. 7).

Crystal structures of *Ce*POFUT1 and human POFUT2 (*Hs*POFUT2) have been solved in complex with nucleotides, revealing that both enzymes adopt the typical GT-B fold[11,12]. Despite similarities at the structural level, their proposed catalytic mechanisms differed. Whereas *Ce*POFUT1 was suggested to follow an S_N1-like mechanism with the GDP β-phosphate acting as the catalytic base, *Hs*POFUT2 was proposed to follow an inverting S_N2-like mechanism in which a glutamic acid acts as the catalytic base[11,12]. Site-directed mutagenesis on *Hs*POFUT2 suggests that the specificity of POFUT2 on TSR domains is based on the 3D structure of these domains.

Both EGF repeats and TSRs are small domains containing three disulfide bridges, and consensus sequences for fucosylation of each have been identified: C^2-X_4-S/T-C^3 (C^2 and C^3 represent the second and third cysteines, respectively; underlining indicates the residues that are potentially fucosylated) for POFUT1 modification of EGF repeats, and C^1-X_{2-3}-S/T-C^2-X_2-G for POFUT2 modification of TSRs[6]. However, the molecular details of how these fucosyltransferases recognize these repeats and catalyze reactions are not currently known. The lack of cristallized ternary complexes has impeded answering the long-standing question of what strategies these broad-substrate enzymes employ to recognize and fucosylate such a vast number of protein substrates. In fact, up to ~75% of human TSRs, which are characterized by highly variable sequence identities and a very low number of conserved residues (identities oscillate between ~19% and ~37%; **Supplementary Results**, **Supplementary Fig. 1a**,**b**), contain the above consensus sequence and are fucosylated by POFUT2. Here, by engineering a fusion protein, we have captured a ternary complex formed between POFUT2, GDP and TSR1 by X-ray crystallography, uncovering the molecular basis of POFUT2 recognition on multiple protein substrates.

RESULTS

Strategy for producing a crystallizable construct

A multiple alignment of *Ce*POFUT2 with higher eukaryote POFUT2s clearly indicated the high level of sequence identity among species, with values ranging from 40% to 44% (**Supplementary Fig. 2**).

[1]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Zaragoza, Spain. [2]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York, USA. [3]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, USA. [4]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, Logroño, Spain. [5]Laboratorio de Microscopias Avanzadas, Instituto de Nanociencia de Aragón, Universidad de Zaragoza, Zaragoza, Spain. [6]Estación Experimental de Aula Dei (EEAD-CSIC), Zaragoza, Spain. [7]Fundación Agencia Aragonesa para la Investigación y Desarrollo (ARAID), Zaragoza, Spain. [8]Instituto de Investigaciones Sanitarias de Aragón (IIS-A), Zaragoza, Spain. [9]Present address: Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, USA. [10]These authors contributed equally to this work. *e-mail: rhurtado@bifi.es

POFUT2s from several species were expressed in *Pichia pastoris*, but only *Ce*POFUT2 was successfully secreted, though the enzyme was partly degraded into two forms (**Supplementary Fig. 3**). We carried out trypsin digestion of the wild-type form and the degraded products followed by MALDI-TOF-MS analysis, which helped us elucidate the potential cleavage site. Single and double mutants were designed and purified, but only the double mutant R298K-R299K (for simplicity, this double mutant will be henceforth referred to as *Ce*POFUT2) showed no degradation products (**Supplementary Fig. 3**). The stable enzyme was active at a very similar rate to that of *Hs*POFUT2 and showed a similar decrease in activity in the presence of manganese (**Supplementary Fig. 4**). We isolated a complex formed between *Ce*POFUT2, GDP and the first TSR of human thrombospondin 1 (*Hs*TSR1) (**Supplementary Fig. 5**). However, this complex did not render crystals containing the ternary complex. Of note, most proteins containing TSRs have multiple tandem repeats; this led us to consider a complex of the enzyme with TSRs in tandem, such as *Hs*TSR1-2-3, which also did not result in crystals. Another strategy was to try to produce a complex of the enzyme with a fucosylated form of *Hs*TSR1, though in this case the complex was not formed (**Supplementary Fig. 5**). This was further supported by isothermal titration calorimetry (ITC) analysis, which showed that no energy was released upon titration of the fucosylated *Hs*TSR1 under the same conditions measured for GDP-fucose, GDP or *Hs*TSR1 (Online Methods, **Supplementary Fig. 6** and **Supplementary Table 1**).

Given that *Hs*TSR1 was able to moderately bind to *Ce*POFUT1 in the absence or presence of saturated GDP (K_d of *Hs*TSR1 = 0.95 or 1.03 μM, respectively; **Supplementary Table 1**) and we could purify a complex by gel filtration (**Supplementary Fig. 5**), we envisioned a different strategy, based on docking studies. We modeled the structure of *Ce*POFUT2 on the basis of the *Hs*POFUT2 crystal structure[11] and docked *Hs*TSR1 into the modeled enzyme (**Supplementary Fig. 7a**). The complex showed that the N terminus of *Hs*TSR1 was ~30 Å away from the *Ce*POFUT2 C terminus (**Supplementary Fig. 7a**; there is a strong resemblance between our modeled complex and one described previously[11]). Consequently, we designed a fusion protein in which *Ce*POFUT2, located at the N terminus, and *Hs*TSR1, located at the C terminus, are coupled by insertion of a flexible linker consisting of 14 glycine and 8 serine residues. The rationale was to increase the local concentration of *Ce*POFUT2-*Hs*TSR1 complex to increase the likelihood of obtaining crystals with a functional ternary complex.

Architecture of the ternary complex

We expressed a synthetic DNA construct encoding the fusion protein in *P. pastoris* and found that the purified complex was able to transfer fucose from GDP-fucose to itself but not to a mutant lacking the O-fucosylation site (Online Methods and **Supplementary Fig. 7b**). We then crystallized the fusion protein in the presence of GDP. The monoclinic crystals diffracted at 1.98 Å (Online Methods and **Supplementary Table 2**), enabling the structure to be solved and the electron density maps to be unambiguously interpreted (**Fig. 1** and **Supplementary Fig. 8**; details in Online Methods). The asymmetric unit (AU) contained two independent ternary complexes with the expected 1:1:1 stoichiometry (one molecule each of *Ce*POFUT2, *Hs*TSR1 and GDP; **Fig. 1a** and **Supplementary Fig. 9**), which was also supported by quaternary structure prediction in PISA, gel filtration chromatography and atomic force microscopy (AFM) (**Supplementary Figs. 5** and **10** and Online Methods). The *Ce*POFUT2 structure adopts the typical GT-B fold formed by two Rossmann-like domains that face each other, as reported for *Hs*POFUT2 and *Ce*POFUT1 (refs. 11,12). *Ce*POFUT2 contains the two conserved disulfide bridges and one N-glycosylation site (N205, according to *Ce*POFUT2 numbering; **Fig. 1a** and **Supplementary Fig. 2**). GDP is located in a shallow cavity of the C terminus, whereas half of *Hs*TSR1 is embraced by a pocket formed between the *Ce*POFUT2 domains (**Fig. 1a**,**b**). Each side of the interface buries an area of 789 Å^2, which agrees with the typical range of 600–800 Å^2 shown for other heterocomplexes[13].

POFUT2 is capable of fucosylating TSRs not only of group 1 but also of group 2, which contains the consensus sequence mentioned above[11,14]. These two types of TSRs share a similar 3D fold and a common core structure (a cysteine-tryptophan-arginine (CWR)-layered structure) but differ in their disulfide bonding patterns[14] (**Fig. 1c**,**d**). For example, only the second disulfide bridge (C18-C51, discussed below) is conserved (**Fig. 1c**,**d**). POFUT2 not only recognizes and reacts with TSRs showing slightly different structures but also accepts TSRs with very low sequence identity. In fact, a sequence conservation analysis of 63 human TSRs showed a very limited number of conserved residues among these repeats (**Fig. 1c**,**e**, and **Supplementary Fig. 1**), emphasizing the versatility of this enzyme in recognizing multiple TSRs. By contrast, a sequence conservation analysis of different POFUT2s showed ~100% identity at both the binding site for the sugar nucleotide and the TSRs (**Fig. 1e** and **Supplementary Fig. 2**).

As mentioned above, POFUT2 fucosylates proteins containing TSRs in tandem repeats and not as single entities (**Supplementary Fig. 11**), raising the question of whether this fucosyltransferase acts processively on these repeats. To address this question, we incubated the enzyme with a fixed concentration of fully folded *Hs*TSR1-2-3 and various concentrations of GDP-fucose (Online Methods and **Supplementary Fig. 12**). The results showed that the TSRs are randomly fucosylated, suggesting that this enzyme is nonprocessive *in vitro* (**Supplementary Fig. 12**). Our prior results suggested that POFUT2 modifies tandem TSRs sequentially from N to C termini *in vivo*[8]. The *in vitro* data presented here suggest that this is not an inherent property of POFUT2 but is probably due to sequential folding of newly synthesized TSRs in the ER. Previous work showed that POFUT2 recognizes a newly folded TSR and adds a fucose, stabilizing the folded TSR and accelerating the net rate of folding, consistent with the quality-control function of POFUT2 (ref. 8).

E52 of *Ce*POFUT2 is the catalytic base

A close-up view of the intersection between the donor and acceptor substrate binding sites shows that GDP is very close to the acceptor S17 of *Hs*TSR1 (**Fig. 2a**) and illuminates mechanistic features of the glycosylation reaction. Specificity for GDP is defined by several hydrogen bonds with the conserved residues N55, H288, R290, D329, D366, S382 and T383, and a stacking interaction between the guanine moiety and F384 (**Fig. 2a** and **Supplementary Fig. 2**). Some of these conserved residues are mutated in *Hs*POFUT2 and show the essential role of R294 in catalysis (R290 in the *Ce*POFUT2)[11]. This residue might be involved in facilitating the cleavage of the glycosidic bond and the stability of the product GDP, as was previously reported for the equivalent R240 of *Ce*POFUT1 (ref. 12). Similar roles have been attributed to metals in GTs that adopt a GT-A fold[15]. Though POFUT2 adopts a GT-B fold, Mn^{2+} has been reported to have a critical role in catalysis for the human enzyme[11]. However, in our study *Hs*POFUT2 showed a ~50% decrease in activity in the absence of this metal (**Supplementary Fig. 4**). Furthermore, the role of the Mn^{2+} in *Ce*POFUT2 catalysis was less important, supporting a nonessential role for Mn^{2+} in catalysis (**Supplementary Fig. 4**). These data were reinforced by our crystallized ternary complex, which shows no metal in the structure, and the general observation that GTs with GT-B folds do not usually bind metals[15]. In our case, R290 replaced Mn^{2+} function by establishing electrostatic and hydrogen bond interactions with β-phosphate (**Fig. 2a**).

The acceptor serine, S17, is engaged in a hydrogen bond with E52, reinforcing the role of this residue as the catalytic base.

This is supported by previous site-directed mutagenesis of E54 (in the human enzyme) (**Supplementary Fig. 2**), revealing that this residue is essential for catalysis[11]. A 0.5 µs molecular dynamics (MD) simulation performed on the enzyme in complex with GDP-fucose and *Hs*TSR1 (Online Methods) showed that the hydroxyl group of S17 is very close to the anomeric carbon of GDP-fucose (3.28 Å) and in an optimal orientation for nucleophilic attack assisted by E52 (**Fig. 2b**). All these features suggest a classical inversion mechanism facilitated by an enzymatic general base catalyst (a glutamate residue in this case)[1,11].

A comparison of the two ternary complexes present in the AU also revealed the presence of a loop (residues 88–94 of *Ce*POFUT2) (**Supplementary Fig. 2**) burying the pyrophosphate moiety of GDP in one of the complexes (**Fig. 1b** and **Supplementary Fig. 13a**). The loop is disordered in one complex of the AU (**Fig. 1b**) and ordered in the other (**Supplementary Fig. 13b**), revealing its dynamic character. In our MD simulation with GDP-fucose (**Fig. 2b**), the sugar was hydrogen bonded to H89 in this loop, implicating this motif in the entrance of GDP-fucose and release of GDP during the catalytic cycle. This is in agreement with previous observations of the importance of this loop in catalysis[11].

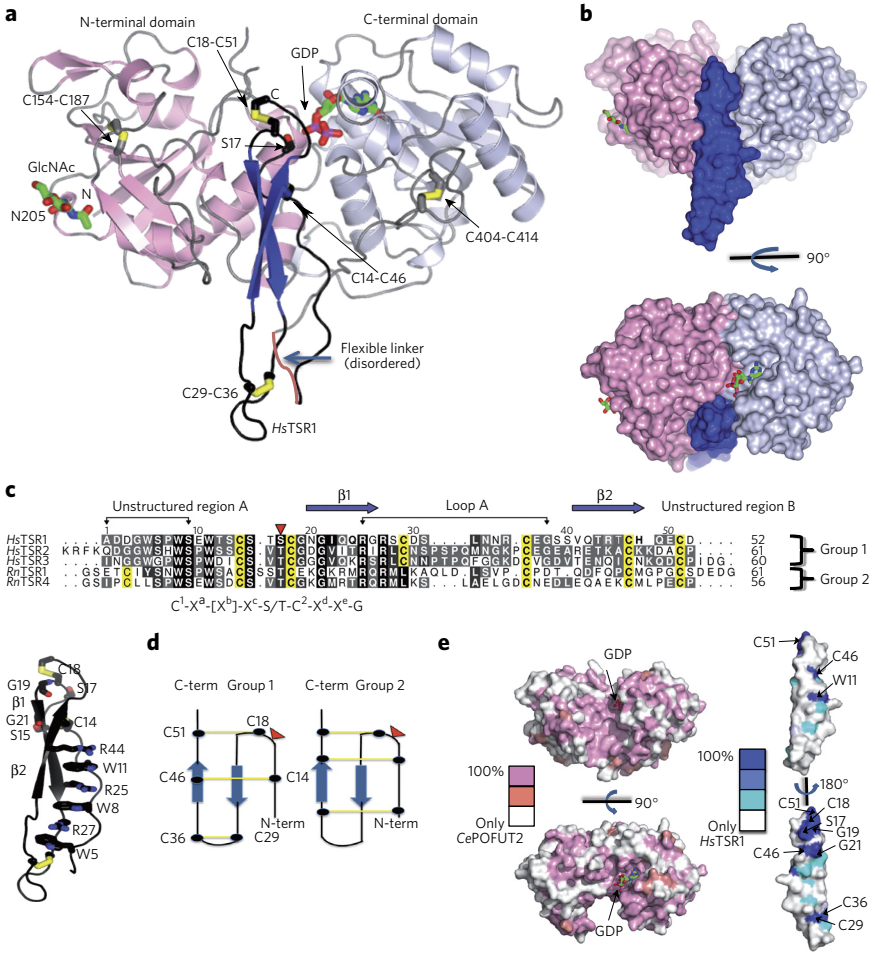

Figure 1 | Structure of *Ce*POFUT2 in complex with GDP and *Hs*TSR1. (**a**) Cartoon representation of the complex. The N and C termini of *Ce*POFUT2 are shown in pink and gray, respectively. Secondary structures of *Hs*TSR1 are shown in blue, loops and unstructured regions in black. Disulfide bridges are indicated in yellow. Carbon atoms of the GlcNAc moiety covalently bound to N205 and GDP are shown in green; carbon atoms of N205 are shown in pink; carbon atoms of the acceptor S17 of *Hs*TSR1 are shown in black. The flexible linker is shown in coral. (**b**) Surface representation of the complex in two views. (**c**) Top, multiple sequence alignment of *Hs*TSR1, *Hs*TSR2 and *Hs*TSR3 of thrombospondin 1 and *Rattus norvegicus* F-spondin 1 and 4 (*Rn*TSR1 and *Rn*TSR4). Red arrowhead indicates fucosylated serine and threonine residues; conserved C residues are highlighted in yellow. The numbering for each TSR does not correspond to its location in thrombospondin 1 or F-spondin. Brackets indicate residues exposed to the bulk solvent. The consensus sequence C^1-X^a-[X^b]-X^c-S/T-C^2-X^d-X^e-G is also shown. Bottom, cartoon representation of *Hs*TSR1. Residues belonging to the CWR-layered structure and other conserved residues between human TSR1, TSR2 and TSR3 are shown as sticks with carbon atoms in black. (**d**) Arrangements of disulfide bridges found in TSRs of group 1 and group 2. Red arrowheads indicate the fucosylation sites. (**e**) Surface representation of *Ce*POFUT2 (left) and *Hs*TSR1 (right), color-coded by degree of sequence conservation.

Preference of threonine over serine residues

POFUT2 has been shown to fucosylate both serine and threonine residues, but whether it 'prefers' one over the other is not known. To address this point, we mutated the *Hs*TSR3 T17 to serine (T17S) and alanine (T17A) (**Fig. 1c**, Online Methods and **Supplementary Tables 3** and **4**). We compared the abundance of secreted *Hs*TSR3 and that of the mutants by quantifying *Hs*TSR3 in cell lysates and medium from transiently transfected HEK293T cells (Online Methods). Both mutations decreased secretion significantly ($P < 0.05$, **Supplementary Fig. 14**), consistent with the relationship between O-fucosylation and secretion previously observed for a disintegrin and metalloproteinase with thrombospondin motifs, member 13 (ADAMTS13) and ADAMTS-like 1 (ADAMTSL1) (refs. 8,16,17). We next evaluated the activity of *Hs*POFUT2 on *Hs*TSR3 and its mutants using a cell-based fucosylation assay and observed that the enzyme fucosylates T17 approximately two-fold as efficiently as S17 (**Fig. 2c** and **Supplementary Fig. 15**). A possible explanation for this behavior might be the higher affinity of the enzyme for T17 owing to the hydrophobic interactions between its methyl group and highly conserved residues such as L56 and T383 (**Fig. 2c** and **Supplementary Fig. 2**), discussed below. The better secretion of TSRs containing threonine-*O*-fucose and the binding preference of the enzyme for threonine might explain why a major population of threonine-containing TSRs has been selected throughout evolution (~64% and ~36% of the human TSRs contain acceptor threonine and serine residues, respectively) (**Supplementary Fig. 1a**,**b**). Similar preferences have been reported for the distant family of retaining polypeptide GalNAc transferases[18].

Specificity of POFUT2 on TSR domains

The interface between *Ce*POFUT2 and *Hs*TSR1 shows relevant features for the specificity of POFUT2 on TSR repeats. *Ce*POFUT2 binding domain contains large cavities that are filled by a large network of water molecules along the interface (**Fig. 3a**). The complex is maintained by a limited number of direct hydrogen bonds and stacking interactions between *Ce*POFUT2 and *Hs*TSR1 that are complemented by water-mediated contacts (**Fig. 3b–d** and **Supplementary Table 5**). Of the ten residues in *Hs*TSR1 directly involved in the binding, only W11, S15 and S17 are conserved. The side chain of S17 establishes a hydrogen bond with E52 (**Fig. 2a**), and W11, S15 and S17 backbones are engaged in hydrogen bond interactions with W420, R63 and E52, respectively (**Fig. 3c**,**d**, and **Supplementary Table 5**). In contrast, the interaction of nonconserved *Hs*TSR1 residues

with *Ce*POFUT2 is exclusively through *Hs*TSR1 side chains (**Supplementary Table 5**). In particular, E10, T12, S13, S39, T43 and Q49 side chains interact with K419, Y225, Y225, E144, W141 and D293, respectively (**Fig. 3b–d** and **Supplementary Table 5**), and I22 interacts with W141, Y145 and L221 (W141/Y145/L221). The central part of *Hs*TSR1 is tethered at both sides by two hydrophobic pockets formed by conserved residues, such as W141/Y145/L221 and Y225/W420, in POFUT2 (**Fig. 3b**). The results also imply that each TSR might adopt a small number of unique direct interactions through their nonconserved residues. Unlike *Hs*TSR1, *Ce*POFUT2 engages most of its conserved residues in direct interactions (**Fig. 1e**, **Supplementary Fig. 2** and **Supplementary Table 5**).

Prompted by these results, we examined the significance of the X positions in the consensus sequence C^1-X^a-[X^b]-X^c-S/T-C^2-X^d-X^e-G (**Fig. 1c** and **Supplementary Table 4**). The additional (bracketed) residue, X^b, has been found in only one TSR—TSR1 of rat F-spondin (**Fig. 1c**)—and this TSR was poorly glycosylated[19]. We generated 20 *Hs*TSR3 mutants to examine the effects on recognition by POFUT2 (Online Methods and **Supplementary Tables 3** and **4**). The mutations included substitutions in structural positions (W11, C14 and C18), the X^a–X^e positions and the conserved glycine in the consensus sequence (**Fig. 1c**). Mutations were generated in the X^a–X^e positions to evaluate whether certain types of amino acids not yet found in a mapped O-fucosylation site affect the efficiency of fucosylation (**Supplementary Table 4**). Mutants were evaluated for effect on secretion (**Supplementary Fig. 14**) and fucosylation efficiency (**Fig. 3e** and **Supplementary Fig. 15**) in cell-based assays.

There is a relationship between TSR folding, fucosylation, and secretion. Mutations predicted to cause defects in folding (for example, C14A, C18A or both (C14A/C18A)) caused large decreases (>50%) in secretion (Online Methods and **Supplementary Fig. 14**) and fucosylation (**Fig. 3e**), consistent with the idea that a properly folded TSR is required for recognition by POFUT2. For these mutants, fucosylation could be measured only in cell lysates. Several other mutations had similar effects: S15L, 16S17 (insertion of an additional S in the X^b position), G19A/G21A, G19A/G20A/G21A and G19A/G20A/G21A/V22G. Fucosylation was detectable on many of these mutants, suggesting that the TSR was folded sufficiently for recognition by POFUT2.

Two mutations in the X^a position, S15D and S15Q, enhanced secretion (**Supplementary Fig. 14**) but caused a significant reduction in fucosylation, suggesting that *Hs*POFUT2 is highly selective for amino acids in the X^a position (**Fig. 3e** and **Supplementary Fig. 15**). The S15 side chain is not engaged in direct interactions with *Ce*POFUT2, though it faces L221 (**Fig. 3d**). Thus, the significant reduction in activity by aspartic acid or glutamine can be rationalized by the proximity of a negatively charged or polar amino acid to a nonpolar residue, which might lead to a decrease of the binding of TSR on POFUT2. In contrast, the X^c, X^d and X^e positions were highly permissive for bulky amino acids (V16H, G19Q, G19R, G20H) for both secretion and fucosylation (**Fig. 3e** and **Supplementary Figs. 14** and **15**). For X^c (T16 in *Hs*TSR1, V16 in *Hs*TSR3) this is because the T16 side chain in the crystal structure is facing one of the water molecule networks that are also surrounded by N49 and E218 (**Fig. 3c**). In the crystal structure, G19, N20 and G21, which correspond to X^d, X^e and the glycine in the consensus sequence, respectively, are exposed to the solvent or surrounded by the network of water molecules, explaining why single mutations on these residues have a little effect upon catalysis (**Fig. 3c**,**d**). Notably, the conserved G21 was not essential for secretion or fucosylation (**Fig. 3e** and **Supplementary Figs. 14** and **15**). Our data also indicate that the tryptophan immediately upstream of the consensus sequence is not required for fucosylation (**Fig. 3e**), suggesting the requirement of tryptophans may vary with different TSRs. In the crystal structure, the W11 side chain is tethered by the R25 and R44 side chains and also faces the bulk solvent, explaining why this residue is important mainly for the stability of the TSRs (**Fig. 3c**). On this basis, we propose a simplified consensus sequence for POFUT2 modification, C-X^a-X^b-S/T-C within a TSR, where certain amino acids are excluded from X^a (for example, aspartic acid and leucine).

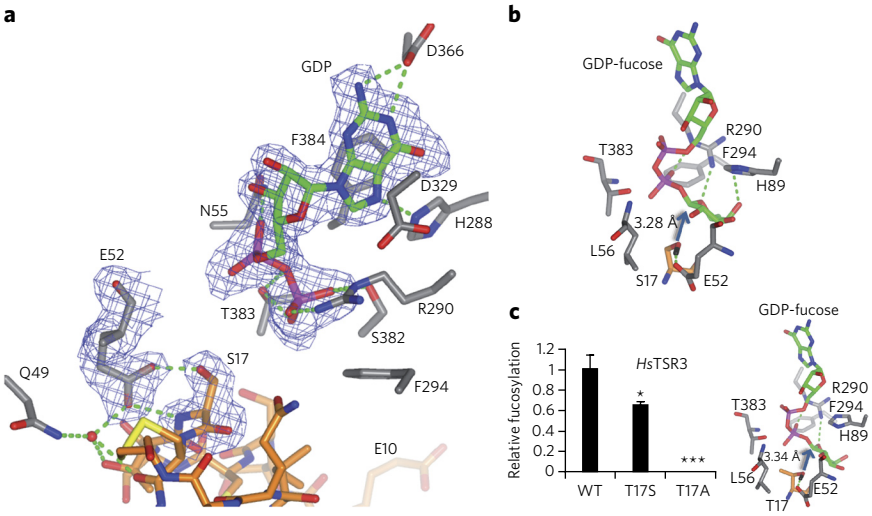

Figure 2 | Catalytic mechanism of POFUT2 and its preference of threonine over serine residues. (**a**) Close-up view of the complex active site. *Ce*POFUT2 residues are shown in gray; *Hs*TSR1 residues are shown in orange; GDP is shown in green; hydrogen bond interactions are shown as dotted green lines. Electron density maps are $F_O - F_C$ and $2F_O - F_C$ syntheses (blue) contoured at 2.2 and 1.0 σ for GDP and S17/E52, respectively. (**b**) Close-up view of *Ce*POFUT2 in complex with GDP-fucose and *Hs*TSR1, showing distance of hydroxyl group of S17 from the anomeric carbon of GDP-fucose. (**c**) Relative fucosylation of wild-type (WT) *Hs*TSR3 compared to the mutants T17S and T17A (left) and close-up view of *Ce*POFUT2 in complex with GDP-fucose and *Hs*TSR1 mutant S17T (right), showing distance of hydroxyl group of T17 from the anomeric carbon of GDP-fucose. All experiments were done in triplicate. Error bars, mean ± s.d. $^*P < 0.10$, $^{***}P < 0.01$ (ANOVA).

Water-mediated interactions

The limited number of direct interactions in the complex is likely to be insufficient to maintain the complex and cannot explain the plasticity required for POFUT2 to achieve recognition on such a diverse family of TSRs. As mentioned above, the protein-protein interface is mostly embedded by solvent, which provides ~50% of the interactions in the complex through an intricate network of water molecules (**Fig. 3a**,**c** and **Supplementary Table 6**).

Most of the hydrogen bonding interactions between the *Hs*TSR1 and water molecules occur through the backbone of nonconserved residues. This feature is probably required for the plasticity to recognize such dissimilar TSRs. In contrast, *Ce*POFUT2 residues interact with the bridging water molecules mainly through conserved side chains and to a lesser extent with the backbone of nonconserved residues, ensuring catalytic efficiency across different orthologs (**Fig. 3c** and **Supplementary Table 5**).

The existence of both direct and water-mediated stabilizing interactions agrees with the favorable enthalpy value measured for this

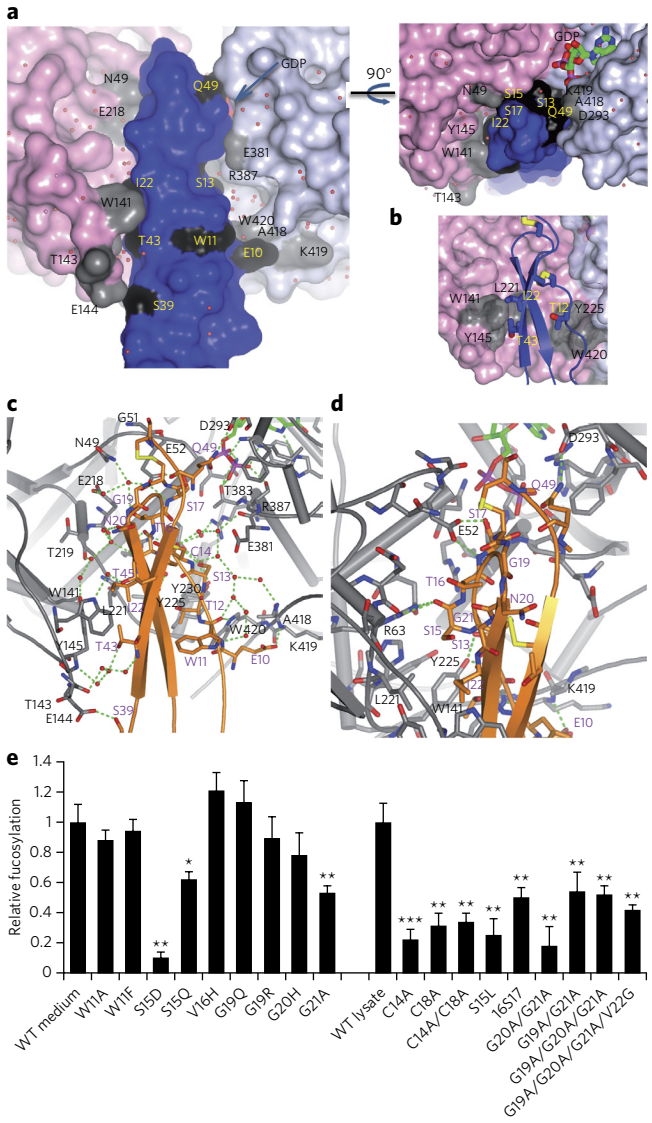

Figure 3 | Interactions in the interface of *Ce*POFUT2 in complex with GDP and *Hs*TSR1. (**a**) Surface representation of the complex interface in two different views. Colors are the same as in Figure 1b. residues shown in gray and black can engage in direct or water-mediated interactions. red spheres indicate water molecules. (b) Close-up view of the interface showing only the stacking interactions in the complex. (c) Close-up view of the direct and water-mediated interactions in the complex. CepoFUT2 and HsTSr1 are shown in gray and orange, respectively. Carbon atoms for each protein are color coded as in Figure 2a. (d) Close-up view of the few direct interactions in the complex. Colors are the same as in c. Hydrogen bond interactions are shown as dotted green lines. (e) relative fucosylation of wild-type (WT) HsTSr3 and different mutants in the interface. Activity is compared with WT HsTSr3 from lysates or medium (Supplementary Fig. 15). All experiments were done in triplicate. Error bars, s.d. $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$ (ANovA).

system (Online Methods and **Supplementary Table 1**); the formation of the complex is in turn entropically unfavorable owing to the persistence of ordered inter-protein solvent molecules that are not released to the water bulk and to the loss of conformational degrees of freedom on binding (**Supplementary Table 1** and **Supplementary Fig. 16**). A crucial role for water molecules in stabilizing protein-protein contacts was recently found in an antifreeze protein[20].

Macromolecular crowding causes osmotically induced protein dehydration[21] and affects water-mediated protein interfaces[22]. To assess the relevance of the interstitial water molecules in our crystallographic structure, we compared the effect of adding 15% and 20% PEG 400 on the K_d of the protein complex and found increases of ~8- and ~60-fold, respectively (Online Methods and **Supplementary Table 1**). This result suggests that PEG-induced water exclusion from the protein complex interface might substantially reduce the affinity between the two proteins, supporting the importance of water in the interaction between POFUT2 and TSR. The secondary structural elements of the complex did not differ substantially in the absence or presence of 20% PEG 400, suggesting that the integrity of the complex was maintained (**Supplementary Fig. 17**).

The role of interstitial water in the complex interface was examined through 0.5μs MD simulations performed on *Hs*TSR1 in complex with the enzyme and in the absence of the flexible linker (Online Methods). Exploratory MD simulations in the absence of explicit solvent but maintaining the crystallographic waters disrupted binding of *Hs*TSR1 and *Ce*POFUT2, indicating a crucial role for discrete water molecules in the molecular recognition process. Conversely, fully solvated complexes between *Ce*POFUT2 and *Hs*TSR1 (wild type, S17T and S17A mutants, in the presence of GDP or GDP-fucose), maintained their associated structure during the total simulation time (0.5 μs). Notably, the first solvation shell around the bound TSR was preserved on binding to *Ce*POFUT2 in the MD simulations, as shown by the similar number of water molecules within 3.4 Å of the whole peptide and the residues located at the complex interface (**Supplementary Table 6**). The solvation patterns of *Hs*TSR1 and S17T and S17A mutants are similar in the free and bound states. A detailed analysis of hydration around the acceptor residues revealed that S17 is, as expected, more solvated than T17 and A17; the number of water molecules in the first solvation of the *Ce*POFUT2-GDP-fucose-*Hs*TSR1 complexes was calculated as 9 ± 2 (s.d.), 6 ± 1 and 5 ± 2 for serine, threonine and alanine, respectively. The poorer solvation of T17 in the complex might increase the reactivity of its more hindered hydroxyl group compared to that of S17. This might explain, in combination with the higher affinity of *Ce*POFUT2 for threonine (**Fig. 2c**), the higher rate of the fucosylation reaction measured for threonine-containing TSRs.

We next calculated the spatial distribution functions for water around *Hs*TSR1 (ref. 23). Whereas the isolated *Hs*TSR1 presents very few structured water molecules, a substantial density of structured water can be observed for its complex with *Ce*POFUT2 throughout the simulation (**Fig. 4a**). The radial distribution functions (RDF) for different atoms of *Hs*TSR1 in both the free and bound states were also derived from the simulations. In general, for residues that are exposed to the solvent, such as the carboxylate group of D30 and the hydroxyl group of S31, these functions display a typical hydrophilic interaction with a well-defined first hydration shell showing a density peak at 2.8 Å and a second hydration shell around 4.7 Å (ref. 24) (**Fig. 4b**). In contrast, atoms involved in highly populated interprotein hydrogen bonds, such as the carbonyl group of Q49, show negligible water density in the first hydration shell. The existence of structured water molecules between *Ce*POFUT2 and *Hs*TSR1 was analyzed through normalized two-dimensional RDF functions (2D-RDF); these functions calculate the probability of finding a water molecule close to two selected atoms in comparison to the one obtained in the bulk[25] and quantify the magnitude of the localized water density (**Fig. 4c** and **Supplementary Fig. 18**). The 2D-RDF functions were calculated for selected pairs of heteroatoms involved in contacts with crystallographic water. The calculated pairwise values of shared water density were high (>8) in all cases, demon-strating the presence of persistent water pockets and the active role of solvent in binding. The highest shared water density (~32 g/ml) was obtained between the carbonyl group of E218 and the hydroxyl group of S15. The average distance between these two atoms was 4.6 Å

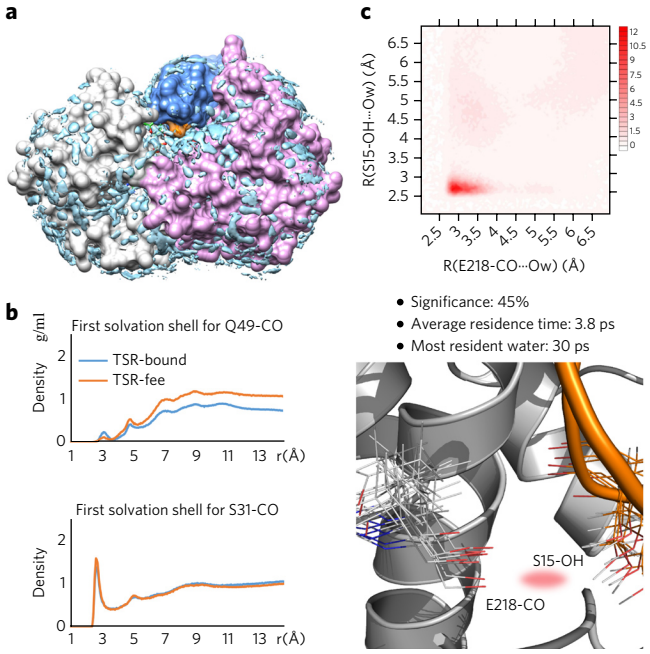

Figure 4 | Hydration structure and dynamics of *Ce*POFUT2-GDP-fucose-*Hs*TSR1 supports water-mediated binding. (**a**) Water oxygen density over 0.5 µs calculated for the ternary complex through MD simulations. The two domains of *Ce*POFUT2-GDP are shown in gray and pink; *Hs*TSR1 is shown in blue; GDP-fucose is shown in green. Reactive S17 is highlighted in orange. Regions with water density >1.5× the density of the bulk are shown as blue isosurfaces. (**b**) 1D-RDF functions calculated for buried (top) and solvent-exposed (bottom) atoms of *Hs*TSR1 in the isolated (orange) and bound (blue) states. (**c**) Top, 2D-RDF function for two selected atoms of the protein-protein interface hydrogen bonded to a crystallographic water; significance (population in the MD ensemble) and average residence time are also shown. Atoms are labeled according to AMBER force field nomenclature. Bottom, ensembles for the ternary complex obtained through MD simulations for the residues involved in the water pocket. Several conformations for S15 and E218 are shown for clarity.

(2.7–5.9 Å). This indicates that although these bridging water molecules are highly structured, they are also very dynamic, continuously changing positions and exchanging with the bulk. The average residence time for a water molecule implicated in these pockets was 3.8 ps, and the most persistent bridging molecule had a residence time of 30 ps. This notion of dynamic structured water is also reflected by the time evolution of the 2D-RDF. The density peak between two selected atoms gradually changed in intensity and position (**Supplementary Fig. 19**), leading to a new water pocket involving a different atom nearby.

These features stress how crucial the role of solvent is in the molecular recognition process of TSR. The hydration water could make the binding surface highly adaptable and thus somewhat promiscuous[26]. The lack of specificity in the complex is probably due to the large number of water-mediated interactions. The dynamic character of the water molecules of the first hydration shell has a pivotal role in the recognition event, allowing water molecules to adopt different positions for singular substrates, acting as an extension of the protein surface.

DISCUSSION

The results described here provide the molecular basis for the promiscuous behavior of POFUT2 acting on folded TSRs with different primary sequences, which might apply to other enzymes that modify a wide variety of peptide sequences. Our data support a general recognition mechanism in which water participates actively in the protein association. The presence of dynamically structured water allows the binding of different TSRs to POFUT2 in a sequence-independent manner that overrides the need for multiple direct protein-protein interactions. This shows a striking example of how nature has produced multispecific enzymes that overcome sequence recognition by harnessing nonspecific protein-protein interactions mediated by water molecules.

Received 8 June 2015; accepted 17 December 2015;
published online XX XX 2016

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: Coordinates and structure factors have been deposited in under accession code 5FOE.

References

1. Hurtado-Guerrero, R. & Davies, G.J. Recent structural and mechanistic insights into post-translational enzymatic glycosylation. *Curr. Opin. Chem. Biol.* **16**, 479–487 (2012).
2. Moremen, K.W., Tiemeyer, M. & Nairn, A.V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).
3. Luther, K.B. & Haltiwanger, R.S. Role of unusual O-glycans in intercellular signaling. *Int. J. Biochem. Cell Biol.* **41**, 1011–1024 (2009).
4. Sakaidani, Y. *et al.* O-linked-*N*-acetylglucosamine on extracellular protein domains mediates epithelial cell-matrix interactions. *Nat. Commun.* **2**, 583 (2011).
5. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
6. Vasudevan, D. & Haltiwanger, R.S. Novel roles for O-linked glycans in protein folding. *Glycoconj. J.* **31**, 417–426 (2014).
7. Luca, V.C. *et al.* Structural biology. Structural basis for Notch1 engagement of Delta-like 4. *Science* **347**, 847–853 (2015).
8. Vasudevan, D., Takeuchi, H., Johar, S.S., Majerus, E. & Haltiwanger, R.S. Peters plus syndrome mutations disrupt a noncanonical ER quality-control mechanism. *Curr. Biol.* **25**, 286–295 (2015).
9. Takeuchi, H. & Haltiwanger, R.S. Significance of glycosylation in Notch signaling. *Biochem. Biophys. Res. Commun.* **453**, 235–242 (2014).
10. Taylor, P. *et al.* Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. *Proc. Natl. Acad. Sci. USA* **111**, 7290–7295 (2014).
11. Chen, C.I. *et al.* Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. *EMBO J.* **31**, 3183–3197 (2012).
12. Lira-Navarrete, E. *et al.* Structural insights into the mechanism of protein O-fucosylation. *PLoS One* **6**, e25365 (2011).
13. Yan, C., Wu, F., Jernigan, R.L., Dobbs, D. & Honavar, V. Characterization of protein-protein interfaces. *Protein J.* **27**, 59–70 (2008).
14. Tan, K. *et al.* Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. *J. Cell Biol.* **159**, 373–382 (2002).
15. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
16. Ricketts, L.M., Dlugosz, M., Luther, K.B., Haltiwanger, R.S. & Majerus, E.M. O-fucosylation is required for ADAMTS13 secretion. *J. Biol. Chem.* **282**, 17014–17023 (2007).
17. Wang, L.W. *et al.* 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 (2007).
18. Lira-Navarrete, E. *et al.* Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide *N*-acetylgalactosaminyltransferase2. *Angew. Chem. Int. Edn Engl.* **53**, 8206–8210 (2014).
19. Gonzalez de Peredo, A. *et al.* C-mannosylation and O-fucosylation of thrombospondin type 1 repeats. *Mol. Cell. Proteomics* **1**, 11–18 (2002).
20. Sun, T., Lin, F.H., Campbell, R.L., Allingham, J.S. & Davies, P.L. An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science* **343**, 795–798 (2014).
21. Parsegian, V.A., Rand, R.P. & Rau, D.C. Macromolecules and water: probing with osmotic stress. *Methods Enzymol.* **259**, 43–94 (1995).

22. Rajapakṣha, A., Stanley, C.B. & Todd, B.A. Effects of macromolecular crowding on the structure of a protein complex: a small-angle scattering study of superoxide dismutase. *Biophys. J.* **108**, 967–974 (2015).
23. Naidoo, K.J. & Kuttel, M. Water structure about the dimer and hexamer repeat units of amylose from molecular dynamics computer simulations. *J. Comput. Chem.* **22**, 445–456 (2001).
24. Soper, A.K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **258**, 121–137 (2000).
25. Andersson, C. & Engelsen, S.B. The mean hydration of carbohydrates as studied by normalized two-dimensional radial pair distributions. *J. Mol. Graph. Model.* **17**, 101–105, 131–133 (1999).
26. Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108 (2008).

Acknowledgments

We thank S.B. Engelsen (University of Copenhagen) for providing the software to calculate 2D-RDF functions and residence times for water molecules. We thank synchrotron radiation sources DLS (Oxford) and beamline I02 (experiment number MX10121-2), and Institute for Biocomputation and Physics of Complex Systems (BIFI), (Memento cluster) for supercomputer support. This work was supported by Agencia Aragonesa para la Investigación y Desarrollo (ARAID), Ministerio de Economía y Competitividad (MEC; BFU2010-19504 to R.H.-G., CTQ2013-44367-C2-2-P to R.H.-G., CTQ2012-36365 to F.C.), the US National Institutes of Health (GM061126 and CA123071, both to R.S.H.), Diputación General de Aragón (DGA; B89 to R.H.-G.) and the EU Seventh Framework Programme (2007–2013) under BioStruct-X (grant agreement 283570 and BIOSTRUCTX 5186, to R.H.-G.).

Author contributions

R.H.-G. designed the crystallization construct and solved the crystal structure. J.V.-G., E.L.-N., C.H.-R. and R.H.-G. cloned the different constructs, purified the enzymes and crystallized the complex. R.H.-G. solved and refined the crystal structure. R.H.-G. and J.V.-G. performed the ITC experiments. G.J.-O. and F.C. performed the molecular dynamics experiments. C.L.-M., D.V., H.T. and R.S.H. cloned the different constructs for expression in mammalian cells and performed site-directed mutagenesis, analysis of enzymatic studies (including the mutants in this work), study of the nonprocessivity of *Ce*POFUT2 and studies in mammalian cells (both the secretion and the activity experiments). M.C.P. and A.L. performed the AFM studies. I.Y. and R.H.-G. performed the multiple alignment of the TSRs and POFUT2s. R.H.-G. wrote the article with contributions from H.T., R.S.H., F.C., A.L. and G.J.-O. All authors read and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to R.H.-G.

ONLINE METHODS

Cloning, expression and purification of *Hs*TSR1 and *Hs*TSR1-2-3. The DNA sequence encoding residues 378–550 of human thrombospondin 1 (*HsTSR1-2-3*), was synthesized and codon optimized by GenScript to be expressed in *E. coli*. The DNA, containing at the 5′ end a recognition sequence for EcoRI and a PreScission protease cleavage site, and at the 3′ end a sequence for SalI, was cloned into the pUC57 vector (GenScript). Following digestion with EcoRI and SalI the construct was subcloned into the protein expression vector pMALC2x (NEB), resulting in the expression plasmid pMALC2x*HsTSR1-2-3* (encoding A378–G550). The plasmid pMALC2x*HsTSR1-2-3* was used as a template for introducing a stop codon by site-directed mutagenesis after the triplet encoding D429 in order to express only *Hs*TSR1. The resulting plasmid was named pMALC2x*HsTSR1*. Site-directed mutagenesis was carried out following the QuikChange kit protocol (Stratagene), using the Phusion High-Fidelity DNA polymerase (Thermo Scientific). To facilitate the purification of *Hs*TSR1-2-3, we introduced a histidine tag preceding the PreScission protease cleavage site. *HsTSR1-2-3* was amplified from pMALC2x*HsTSR1-2-3* using the forward primer 5′-*CTGGAAGTTCTGTTCC AGGGGCCC*GGTAGCGGTAGCGCAGACGATGGTTGG-3′, which contains a sequence encoding a PreScission protease cleavage site (shown in italics), and the reverse primer 5′-CGGAATTC*GTCGAC*TCAGCCATCAATCGGGC AATC-3′, containing a SalI site (italics) and a stop codon (underlined). *HsTSR1-2-3* was amplified from the first PCR product using the reverse primer described above and forward primer 5′-CGGAATTCGAATTCAG CGGCGGC**CATCACCATCACCATCACCATCAC** GGCGGCAGCC*TGGA AGTTCTGTTCCAGGGGCCC*-3′, containing an EcoRI site (underlined), a sequence encoding a histidine tag (bold) and a PreScission protease cleavage site (italics).

The PCR product was then digested with EcoRI and SalI and cloned into pMALC2x, resulting in the expression plasmid pMALC2xHist*HsTSR1-2-3*. All plasmids were verified by sequencing (Sistemas Genómicos).

Both pMALC2x*HsTSR1* and pMALC2xHist*HsTSR1-2-3* were transformed into Rosetta-gami 2 (DE3) cells. Cells were grown and disrupted by sonication as described[27].

The supernatant containing the maltose binding protein (MBP) bound to *Hs*TSR1 was loaded into 1 × 5 ml MBP Trap HP column (GE Healthcare) equilibrated with buffer A (25 mM Tris-HCl, 150 mM NaCl, pH 7.5). The fusion protein was eluted with buffer A + 10 mM maltose, dialyzed with buffer A, and subsequently cleaved overnight with PreScission protease (PP; GE Healthcare) at 4 °C. *Hs*TSR1 was further purified by size exclusion chromatography using a Superdex 75 XK26/60 column (Sigma) equilibrated with buffer A. The eluted protein was dialyzed against buffer B (25 mM Tris-HCl, pH 7.5), concentrated and used for biophysical experiments.

The supernatant containing the MBP-*Hs*TSR1-2-3 was loaded into 1 × 5 ml HisTrap HP column (GE Healthcare) previously equilibrated in buffer C (25 mM Tris-HCl, 150 mM NaCl, 20 mM imidazole, pH 7.5). The fusion protein was eluted with a gradient of imidazole from 20 to 200 mM. Then, buffer exchange of the fusion protein into buffer A was carried out using a HiPrep 26/10 desalting column, and the fusion protein was subsequently cleaved overnight with PP at 4 °C. The PP and the maltose binding protein were removed by a 1 × 5 ml GSTrap (GE Healthcare) and MBP Trap HP columns, respectively. The unbound *Hs*TSR1-2-3 was further purified by size exclusion chromatography as described above. The eluted protein was dialyzed against buffer B, concentrated and used for biophysical experiments.

Cloning, expression and purification of *Ce*POFUT2. The DNA sequence encoding amino acid residues 40–424 of *Ce*POFUT2, defined as *cepofut2*, was synthesized and codon optimized by GenSript for expression in *P. pastoris*. The DNA, containing at the 5′ end a recognition sequence for XhoI and a KEX2 cleavage signal, and at the 3′ end a sequence for SacII, was cloned into the pUC57 vector (GenScript). Following digestion with XhoI and SacII the construct was subcloned into the protein expression vector pPICZαA, resulting in the expression plasmid pPICZαA*cepofut2*. Site-directed mutagenesis was carried out following the QuikChange kit protocol (Stratagene), using the KOD HotStart DNA polymerase (Novagene). The plasmid pPICZαA*cepofut2* was used as a template for introducing the following single and double amino acid changes by site-directed mutagenesis: R299D, R298K-R299K and R299K-A300V. The plasmid pPICZαA*cepofut2-R298K-R299K*, which encodes the double mutant R298K-R299K, is referred to as *Ce*POFUT2 in the manuscript.

The DNA sequence encoding the fusion protein formed by amino acid residues 40–424 of the *Ce*POFUT2 (this construct contains two triplets encoding for the double mutant R298K-R299K), the flexible linker (formed by a combination of 22 glycine and serine residues) and *Hs*TSR1 (**Fig. 1c**) was synthesized and codon optimized by GenScript for expression in *P. pastoris*. The DNA, containing at the 5′ end a recognition sequence for *Xho*I and a KEX2 cleavage signal, and at the 3′ end a sequence for SacII, was cloned into the pUC57 vector (GenScript). Following digestion with XhoI and SacII the construct was subcloned into the protein expression vector pPICZαA (Invitrogen), resulting in the expression plasmid pPICZαA*cepofut2-linker-HsTSR1*. The plasmid pPICZαA*cepofut2-linker-HsTSR1* was used as a template for introducing the mutant S17A in the *Hs*TSR1 by site-directed mutagenesis. All plasmids were verified by sequencing (Sistemas Genómicos).

All plasmids were isolated from *E. coli* strain DH5α, linearized with SacI and used to transform the *P. pastoris* strain X-33 by electroporation. Transformants were selected and cells were grown as described before[12]. Supernatant containing *Ce*POFUT2 was dialyzed against buffer D (25 mM Tris-HCl, pH 8.5) and loaded into 1 × 5 ml HiTrap Blue column. The protein was eluted with a NaCl gradient from 0 to 1 M. Supernatant containing the fusion protein was dialyzed against buffer D and loaded into 1 × 5 ml HiTrap QFF column. The protein was also eluted with a NaCl gradient from 0 to 1 M. Buffer exchange of both *Ce*POFUT2 and the fusion protein into buffer E (25 mM MES, pH 6.7) was carried out using a desalting column as described above. Both proteins were then treated with Endo H_f (NEB) over night at 18 °C. Samples were loaded into 1 × 5 ml MBPTrap HP column where both *Ce*POFUT2 and the fusion protein was collected in the flow-through. Finally, both proteins were further purified by size exclusion chromatography as described above. The eluted proteins were dialyzed against buffer B, concentrated and used for biophysical experiments.

For the formation of the different complexes, a 5:1 ratio of either *Hs*TSR1 or fucosylated *Hs*TSR1 and *Ce*POFUT2, and a 3:1 ratio of *Hs*TSR1-2-3 and *Ce*POFUT2 was purified by size exclusion chromatography. In these experiments, the column was first equilibrated with buffer F (25 mM Tris-HCl, 10 mM NaCl, 50 µM GDP, pH 7.5). Buffer exchange of the eluted complexes into buffer F without NaCl was carried out, and complexes of *Hs*TSR1-*Ce*POFUT2 and *Hs*TSR1-2-3–*Ce*POFUT2 were concentrated and used for biophysical experiments.

Crystallization. Crystals of the fusion protein in complex with GDP were obtained by mixing 0.5 µl of protein solution (a mix formed by 20 mg/ml of the fusion protein and 5 mM GDP in buffer C) with 0.5 µl of precipitant solution (16–20% PEG 3000, 100 mM sodium citrate, pH 5.5) against 60 µl of precipitant solution. The crystals were obtained by sitting drop vapor diffusion at 18 °C. The crystals were cryoprotected in the precipitant solution plus 20% ethylene glycol and frozen in a nitrogen gas stream cooled to 100 K.

Structure determination and refinement. Data were collected in the beamline I02 of DLS at a wavelength of 0.97 Å and a temperature of 100 K. The data were processed and scaled using the XDS package[28] and CCP4 (refs. 29,30) software. Relevant statistics are given in **Supplementary Table 2**. The crystal structure was solved by molecular replacement with Phaser[29,30] and using human POFUT2 (PDB 4AP5) as the template. Initial phases were further improved by cycles of manual model building in Coot[31] and refinement with REFMAC5 (ref. 32). After *Ce*POFUT2 was built and refined, ARP/wARP[29,30] was used to build the *Hs*TSR1. The final model was validated with PROCHECK; model statistics are given in **Supplementary Table 2**. The asymmetric units of the primitive monoclinic crystals contain two molecules of the ternary complex in a stoichiometry of 1:1:1 (**Supplementary Fig. 9**). The Ramachandran plot shows that 95.19%, 3.76% and 1.06% of the amino acids are in most favored, allowed and disallowed regions, respectively.

Spectroscopic characterization by Far-UV circular dichroism (CD). CD spectra were acquired at 25 °C in a Chirascan spectropolarimeter (Applied-Photophysics). Far-UV CD spectra were recorded from 190 to 250 nm using a 1-mm path-length cuvette. The experiments consisted of the fusion protein at 15 µM in 50 mM HEPES, pH 7.5, in the presence and absence of 20% PEG 400.

Radioactive assay for POFUT2 activity. POFUT2 assays were performed as previously described[33]. Briefly, *Hs*TSR3 from thrombospondin 1 (100 µM final,

doi:10.1038/nchembio.2019

purified from BL21 cells) and GDP-[^{3}H]fucose (10 µM final, 19.8 Ci/mmol, American Radiolabeled Chemicals) was incubated with purified *Ce*POFUT2 or *Hs*POFUT2 (11.2 ng) in 50 mM HEPES, 10 mM $MnCl_2$, pH 6.8, for 20 min at 37 °C. The reaction was stopped by adding 1,000 reaction volumes of 0.1 M EDTA. TSRs were then purified from the mixture using a C18 cartridge (100 mg, Agilent Technologies), and radioactivity was measured on a Beckman scintillation unit.

Fucosylation of *Hs*TSR1. *Hs*TSR1 was fucosylated in the presence of *Ce*POFUT2 and GDP-fucose. The reaction was carried out in 25 mM Tris-HCl, 150 mM NaCl, pH 7.5, at 18 °C overnight. *Hs*TSR1 was in five-fold excess compared to the enzyme whereas GDP-fucose was added at 700 µM. The fucosylated *Hs*TSR1 sample was then loaded into 1 × 5 ml HiTrap QFF column (GE Healthcare), previously equilibrated with 25 mM Bis-TRIS, pH 6.5. Fucosylated *Hs*TSR1 was collected in the flow through, concentrated, quantified and used in both ITC and size exclusion chromatography experiments.

Isothermal titration microcalorimetry (ITC). ITC was used to characterize the interaction of *Ce*POFUT2 with GDP, GDP-fucose, *Hs*TSR1 and fucosylated *Hs*TSR1. All experiments were carried out in an Auto-iTC$_{200}$ (Microcal, GE Healthcare) at 25 °C with *Ce*POFUT2 at 50 µM and concentrations of ligands at 300 µM, in both 50 mM HEPES and 50 mM PIPES, pH 7.0. To determine how PEG 400 might affect the K_d of *Hs*TSR1, we used the same conditions as above but using 15% and 20% PEG 400 in buffer HEPES, pH 7.0.

The buffer independent binding enthalpy ΔH_0 was estimated from linear regression using different ionization enthalpies ΔH_{ion} for each buffer (ΔH_{ion} (HEPES) = 4.88 kcal/mol and ΔH_{ion} (PIPES) = 2.74 kcal/mol)[27]. Data integration, correction and analysis were carried out in Origin 7 (Microcal). The data was fitted to a one-site equilibrium-binding model.

Atomic force microscopy (AFM). AFM enabled the direct visualization of single *Ce*POFUT2 molecules to characterize the different conformational states upon ligand binding. Methods and instrumentation were as reported[34]. 5× GDP, GDP-fucose, *Hs*TSR1 and $MnCl_2$ were added to *Ce*POFUT2. AFM images were analyzed with the WSxM software[35] as described[36]. Though *Ce*POFUT2 presents a slightly negative net charge at neutral pH, the enzyme was adsorbed electrostatically on the negative mica surface owing to the presence of Mn^{2+}, which inverts its polarity. Evaluation of adsorption of enzymes on mica showed that they preserve the enzymatic activity, even later desorbed[37]. *Ce*POFUT2 appears mainly as monomer, but the images allowed the identification of several distinguishable conformational states and compaction degree. *Ce*POFUT2 heights vary from 3 to 7 nm owing to the charged ligands, Mn^{2+}, GDP and/or GDP-fucose. These molecules attract the AFM tip, making a 'flattening' effect on the molecules. The addition of *Hs*TSR1 increases their overall volume between 7 and 10 nm (**Supplementary Fig. 10**). The *z* height has sub-nanometer resolution, and in the *xy* plane the scanned features show the AFM tip broadening effect that does not affect the comparative analysis of the width related to the size owing to proportionality. When incubating only with Mn^{2+}, the molecules clearly showed the two domains in a separate disposition, composing an open or extended structure (**Supplementary Fig. 10**). The negatively charged GDP-fucose binds strongly to the substrate, hindering visualization of enzyme molecules, but the analysis was possible. Binding to GDP or GDP-fucose introduces a conformational change that leads to an increase in volume of one of the modules and, in turn, in proximity of the domains (**Supplementary Fig. 10**). Simultaneous binding of GDP and *Hs*TSR1 leads to global reorganization of the structure, generating a feature of a higher volume in which the modules cannot be distinguished and composing a closed conformation (**Supplementary Fig. 10**).

Fucosylation of the fusion protein *Ce*POFUT2-*Hs*TSR1. The fusion protein *Ce*POFUT2-*Hs*TSR1 (wild type (WT) or S17A mutant, 0.11 mg/ml) was incubated in the presence of radioactive 2.5 µM GDP-[1-^{3}H]fucose (19.8 Ci/mmol, American Radiolabeled Chemicals). The reaction was carried out in 10 µl of 50 mM HEPES, pH 6.8, 10 mM $MnCl_2$, and 0.5% Nonidet P-40 at 37 °C for 20 min and stopped by adding 900 µl of 100 mM EDTA, pH 8.0. The sample was loaded onto a C18 cartridge (100 mg, Agilent Technologies). After the cartridge was washed with 5 ml H_2O, the bound fusion protein was eluted with 1 ml 80% methanol. Incorporation of [^{3}H]fucose into the fusion proteins was determined by scintillation counting of the eluate. Reactions without the fusion proteins were used as background control. Data are from three independent assays and are reported as mean ± s.e.m.

Fucosylation of *Hs*TSR1-2-3 to test processivity of *Ce*POFUT2. *Hs*TSR1-2-3 (10 µM) was fucosylated in the presence of 1.5 µg of *Ce*POFUT2 and GDP-fucose at concentrations of 200, 20 or 2 µM. The reaction was carried out in 30 µl of 50 mM HEPES, pH 6.8, 10 mM $MnCl_2$ at 37 °C overnight. An aliquot of the products were analyzed by LC-MS using an Agilent 6340 ion-trap mass spectrometer with a nano-HPLC CHIP-Cube interface as previously described[38]. Extracted ion chromatograms for the most abundant charge state of the unmodified form, or mono-, di-, or tri-fucosylated forms of *Hs*TSR1-2-3 were generated. The remainder of the products were purified by reverse phase HPLC (Agilent Technologies, 1200 Series) with a C18 column (4.6 mm × 150 mm, VYDAC) with a linear gradient of solvent B (80% acetonitrile, 0.1% trifluoroacetic acid (TFA) in water) from 10% to 90% in solvent A (0.1% TFA in water) for 30 min, monitoring absorbance at 214 nm. The lyophilized samples were reduced, alkylated, digested with trypsin, and analyzed by nano-LC/MS/MS using an Agilent 6340 ion-trap mass spectrometer with a nano-HPLC CHIP-Cube interface auto sampler as reported previously[38].

Analysis of secretion and fucosylation of *Hs*TSR3 mutants using cell-based assays. Mutations in the variable X positions of the *O*-fucose consensus sequence (**Supplementary Table 4**) were generated by site-directed mutagenesis using the pSecTag-HsTSR3 (with C-terminal Myc-His_6 tag) and the primers listed in **Supplementary Table 3**. Plasmids encoding wild-type and mutant forms of *Hs*TSR3 were coexpressed with hIgG controls in HEK293T cells by transient transfection. Briefly, 6 µL of polyethylenimine (PEI) was added to 0.8 µg pSecTag-*Hs*TSR3 and 0.2 µg pRK5-hIgG in 100 µL 150 mM sodium chloride. Transfection mixtures were vortexed, incubated for 20 min at room temperature, and added dropwise onto 80% confluent HEK293T cells in 1 mL DMEM in 32-mm dishes. Four hours after transfection, cells were washed with PBS and incubated in 1 mL OptiMEM containing 200 µM peracetylated alkynylfucose (Invitrogen) for 72 h as previously described[39]. Media was collected for analysis and cells lysed in RIPA buffer (0.1% SDS, 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% sodium doxycholate, 1% Nonidet P-40). Samples were then analyzed for protein production and secretion by western blot with anti-Myc antibodies. Fucosylation was analyzed after cycloaddition ('click') reaction of fucosylated proteins with azide-biotin as previously described[39]. Briefly, media or lysates containing alkynyl fucose–tagged glycoproteins were incubated with 1 mM copper sulfate and 2 mM sodium ascorbate in PBS with 0.1 mM azide-biotin and 0.1 mM tris-(benzyltriazolylmethyl)amine for 1 h at room temperature. Clicked samples were centrifuged for 5 min at 14,000 r.p.m. before analysis by immunoblot using IRDye 800–labeled streptavidin.

For western blotting, proteins were separated by 15% SDS-PAGE then transferred to 0.45-µm nitrocellulose membrane (BioRad). Blots were blocked 1 h at room temperature in 5% milk in PBS, 0.1% Tween (PBST), then probed with 1:2,000 anti-Myc antibody (clone 9E10 (ref. 39)) in PBST. Samples containing unclicked protein were incubated with Alexa Fluor 680 goat anti-mouse IgG (Invitrogen, catalog number A-21057 (ref. 39)) and IRDye800 anti-human IgG (Rockland Immunochemicals, catalog number 609-432-017 (ref. 8)), both 1:2,000 in PBST, for 1 h in the dark at room temperature. Clicked samples were incubated with 1:2,000 Alexa Fluor 680 goat anti-mouse IgG[39] and 1:20,000 streptavidin IRDye800 (Rockland Immunochemicals, catalog number S000-32 (ref. 39)) in PBST for 1 h at room temperature in the dark. Western blots were visualized using an Odyssey Imager (LI-COR). Intensities of bands were quantified from captured images using Odyssey Software. Relative mutant secretion was determined by calculating the ratio of Myc signal to that of the IgG control in the media, with wild-type protein normalized to 1. Relative fucosylation was calculated from captured images by dividing the intensity of streptavidin by the intensity of Myc signal, and expressing as a percentage of wild type ratio. s.e.m. was calculated for each sample mean comparing the mutant streptavidin to Myc ratio to wild-type ratio.

Molecular dynamics simulations. Parameters for the substrates were generated with the antechamber module of Amber14 (ref. 40) using a combination of GLYCAM06 (ref. 41) parameters for the fucose unit and the general Amber force field (GAFF) for GDP, with partial charges set to fit the electrostatic potential generated with HF/6-31G(d) by RESP[42]. The charges are calculated according to the Merz-Singh-Kollman scheme using Gaussian 09 (ref. 43).

Each protein was immersed in a truncated octahedral box with a 10 Å buffer of TIP3P[44] water molecules and neutralized by adding explicit counter ions (Na^+Cl^-). All subsequent simulations were performed using the ff14SB force field[45]. A two-stage geometry optimization approach was used. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal/mol were applied to the solute, and the Andersen temperature coupling scheme[46] was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages. Water molecules are treated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects are modeled using the particle-mesh-Ewald method[47]. An 8-Å cutoff was applied to Lennard-Jones and electrostatic interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 0.5 µs under the same simulation conditions. Water density properties were derived from the production trajectories using a cubic grid consisting of 160 × 160 × 160 bins with 0.5-Å spacing, through the grid command available in the cpptraj module of AmberTools14 (ref. 40). Minimizations in explicit solvent for the modeled complex (**Supplementary Fig. 11**) was performed running 2000 steepest-descent steps followed by 2000 conjugate gradient optimization steps using the AMBER force field. MD simulations were performed on *Ce*POFUT2 in complex with *Hs*TSR1. Two additional calculations were run with mutants of *Hs*TSR1 (S17T, S17A). The simulations were duplicated to study the influence of GDP or GDP-fucose. In all cases, the MD calculations were run without the linker between the protein and the corresponding TSR. In addition, MD simulations on *Hs*TSR1 ligand in the free state were run.

Volume maps of the water density were obtained with UCSF Chimera[48]. The radial distribution functions (RDF) of water (O atom) for selected protein atoms were measured using the radial command in cpptraj. 2D-RDF functions and the residence times were calculated using in-house programs developed in collaboration with S.B. Engelsen (University of Copenhagen)[25].

27. Sahún-Roncero, M. *et al.* The mechanism of allosteric coupling in choline kinase α1 revealed by the action of a rationally designed inhibitor. *Angew. Chem. Int. Edn Engl.* **52**, 4582–4586 (2013).
28. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
29. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
30. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
31. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
32. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
33. Luo, Y., Nita-Lazar, A. & Haltiwanger, R.S. Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J. Biol. Chem.* **281**, 9385–9392 (2006).
34. Lira-Navarrete, E. *et al.* Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. *Nat. Commun.* **6**, 6937 (2015).
35. Horcas, I. *et al.* WSXM: a software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.* **78**, 013705 (2007).
36. Lostao, A., Peleato, M.L., Gómez-Moreno, C. & Fillat, M.F. Oligomerization properties of FurA from the cyanobacterium *Anabaena sp.* PCC 7120: direct visualization by *in situ* atomic force microscopy under different redox conditions. *Biochim. Biophys. Acta* **1804**, 1723–1729 (2010).
37. Marcuello, C., Arilla-Luna, S., Medina, M. & Lostao, A. Detection of a quaternary organization into dimer of trimers of *Corynebacterium ammoniagenes* FAD synthetase at the single-molecule level and at the in cell level. *Biochim. Biophys. Acta* **1834**, 665–676 (2013).
38. Leonhard-Melief, C. & Haltiwanger, R.S. O-fucosylation of thrombospondin type 1 repeats. *Methods Enzymol.* **480**, 401–416 (2010).
39. Al-Shareffi, E. *et al.* 6-alkynyl fucose is a bioorthogonal analog for O-fucosylation of epidermal growth factor-like repeats and thrombospondin type-1 repeats by protein *O*-fucosyltransferases 1 and 2. *Glycobiology* **23**, 188–198 (2013).
40. Case, D.A. *et al.* AMBER 14. (University of California, San Francisco, 2014).
41. Kirschner, K.N. *et al.* GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655 (2008).
42. Bayly, C.I., Cieplak, P., Cornell, W. & Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
43. Frisch, M.J. *et al.* Gaussian 09, Revision D.01. (Gaussian, Inc., Wallingford, CT, 2009).
44. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
45. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
46. Andrea, T.A., Swope, W.C. & Andersen, H.C. The role of long ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.* **79**, 4576–4584 (1983).
47. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
48. Pettersen, E.F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

doi:10.1038/nchembio.2019

A proactive role of water molecules in acceptor recognition by protein *O*-fucosyltransferase 2

Jessika Valero-González, Christina Leonhard-Melief, Erandi Lira-Navarrete, Gonzalo Jiménez-Osés, Cristina Hernández-Ruiz, María Carmen Pallarés, Inmaculada Yruela, Deepika Vasudevan, Anabel Lostao, Francisco Corzana, Hideyuki Takeuchi, Robert S Haltiwanger & Ramon Hurtado-Guerrero

EDITORIAL SUMMARY

AOP: A structural and functional study including a mutational analysis of *C. elegans* POFUT2 with GDP and a peptide substrate helps explain how this and other fucosyltransferases recognize protein substrates with diverse sequences.