Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

The Essential Role of Water Molecules in the Reaction Mechanism of Protein *O*-Fucosyltransferase 2

I. Sanz-Martínez, A. García-García, T. Tejero, R. Hurtado-Guerrero, P. Merino**

Supporting Information

Contents

Experimental Methods

Cloning, expression and purification of the fusion proteins

The design of the protein fusion was made as shown before for pPICZαA*cepofut2-linker-HsTSR1*.[1] 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 the *Rattus norvergicus* F-spondin 1, defined as *Rn*TSR4 (residues 613-666), was made synthetically and codon optimized by GenScript for expression in *Pichia 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 *Sac*II, was cloned into the pUC57 vector (GenScript). Following digestion with *Xho*I and *Sac*II the construct was subcloned into the protein expression vector pPICZαA (Invitrogen), resulting in the expression plasmid pPICZαA*cepofut2-linker-Rn*TSR4. The plasmid pPICZαA*cepofut2-linker-RnTSR4* was used as a template for introducing the double mutant E10C in the *Rn*TSR4 (corresponding to Glu622 in F-spondin 1 with Uniprot code P35446) and the mutant A418C in the *Ce*PoFUT2 by site-directed mutagenesis. The resulting plasmid was named as pPICZαA*cepofut2*^{A418C}*-linker-RnTSR4*^{E10C}. All constructs were made by GenScript.

All plasmids were isolated from *E. coli* strain DH5α, linearized with *Sac*I and used to transform the *Pichia pastoris* strain X-33 by electroporation. Transformants were selected and cells were grown as described before.[1] Both fusion proteins (*Ce*PoFUT2-linker-*Rn*TSR4 and *Ce*PoFUT2^{A418C}-linker-*Rn*TSR4^{E10C}) were purified according to the purification protocol described for *Ce*PoFUT2-linker-*Hs*TSR1.[1]

Crystallization

Crystals of the *Ce*PoFUT2^{A418C}-linker-*Rn*TSR4^{E10C} 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 25 mM Tris/HCl pH 7.5) with 0.5 µl of precipitant solution (0.1 M BIS-Tris pH 5.5, 0.2 M calcium chloride dihydrate, 45 % v/v MPD and ~20 mM ammonium sulfate) against 60 µl of precipitant solution. The crystals were obtained by sitting drop vapour diffusion at 18ºC. 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

The data was collected in the beamline XALOC of ALBA at a wavelength of 0.97 Å and a temperature of 100 K, respectively. The data was processed and scaled using the XDS package[2] and CCP4 software.[3] Relevant statistics are given in Table S1. The crystal structure was solved by molecular replacement with Phaser[3] and using the PDB entry 5FOE as the template that

corresponds to the fusion protein *Ce*PoFUT2-linker-*Hs*TSR1. Initial phases were further improved by cycles of manual model building in Coot[4] and refinement with REFMAC5. [5]After *Ce*POFUT2^{A418C} was built and refined, ARP/wARP[3] was used to build the *Rn*TSR4^{E10C}. The final model was validated with PROCHECK, model statistics are given in Table S1. The asymmetric units of the primitive monoclinic crystals contain 2 molecules of the binary complex in a stoichiometry of 1:1 because GDP was not present in the active site. The Ramachandran plot shows that 89.6%, 10.1%, 0.3% and 0.0% of the amino acids are in most favored, additional allowed, generously allowed and disallowed regions, respectively. The structure of the protein has been deposited at PDB with PDB ID: 8AY1

Table S1. Data collection and refinement statistics.

	*Ce*PoFUT2^{A418C}-*Rn*TSR4^{E10C}
Data collection	
Space group	$P6_3$
Cell dimensions	
a, *b*, *c* (Å)	156.85 156.85 80.45
a, b, g (°)	90, 90, 120
Resolution (Å)	20-2.13 (2.24-2.13*)
R_{merge}	0.087 (2.069)
R_{pim}	0.032 (0.765)
$I / \sigma I$	10.6 (1.1)
Completeness (%)	99.6 (98.1)
Redundancy	8.5 (8.1)
Mn(I) half-set correlation CC(1/2)	0.998 (0.491)
Refinement	
Resolution (Å)	2.13
Total number reflections	534790
Total unique reflections	62973
R_{work} / R_{free}	0.1935/0.2404
No. atoms	
*Ce*PoFUT2^{A418C}	6388
*Rn*TSR4^{E10C}	573
GlcNAc	28
Glycerol	6
Waters	156
Ethylenglycol	48
Sulfate	60
B-factors (Å^2)	
*Ce*PoFUT2^{A418C}	62.54
*Rn*TSR4^{E10C}	82.84
GlcNAc	66.05
Glycerol$^+$	77.30
Waters	58.73
Ethylenglycol	74.69
Sulfate	97.47
R.m.s. deviations	
Bond lengths (Å)	0.00608
Bond angles (°)	1.3892

One crystal was used to determine the crystal structure. *Values in parentheses are for highest-resolution shell.
There are two molecules in the asymmetric unit of the fusion protein

Supplementary Figures

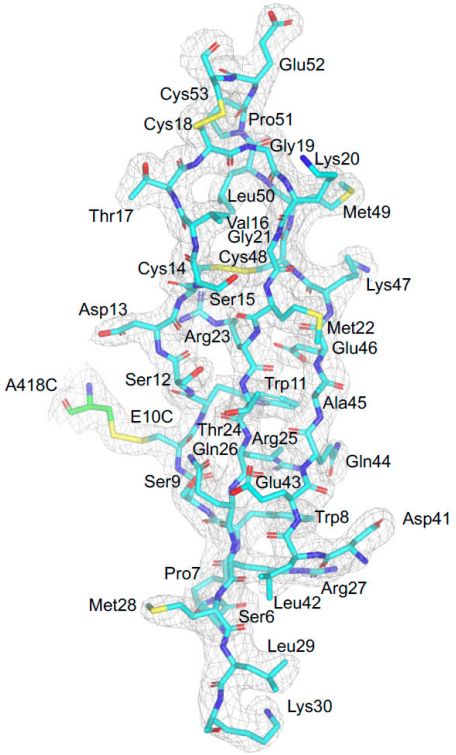

Figure S1. Electron density map of the *Rn*TSR4^{E10C}. The residues of *Rn*TSR4^{E10C} are colored as cyan carbon atoms. Electron density map is $2F_O–F_C$ (grey) contoured at 1.0 σ. The electron density for the disulfide bridge formed by E10C and A418C is also displayed.

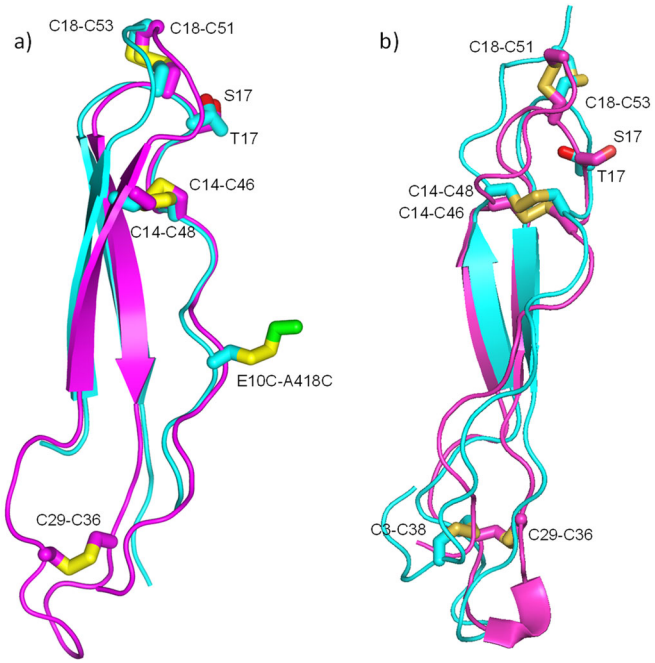

Figure S2. Superposition of *Rn*TSR4^{E10C} with *Hs*TSR1. (a) X-ray. (b) Final snapshots of 1 μs MD simulations. *Hs*TSR1 and *Rn*TSR4^{E10C} are shown in magenta and cyan, respectively. The disulfide bridge formed by E10C and A418C is also displayed in panel a.

Computational Methods

We started from the crystal structure of human PoFUT2 in complex with GDP-fucose previously obtained (PDB ID 4AP6), and we included *Hs*TSR1 and *Rn*TSR4 using as templates the structures of the ternary complex obtained with *Ce*PoFUT2, GDP and *Hs*TSR1 (PDB ID 5FOE) for the former, and the one reported in this work for the latter (X-ray coordinates in all cases). We selected human PoFUT2 because we consider more relevant the results with this protein. Both enzymes show a high degree of conservation, in particular at the active site and crucial sections of the enzyme so, it should be expected essentially the same results for both proteins. Since we have good crystal data for the human protein in complex with GDP-fucose and the apo form, we selected the former for our initial studies. On the other hand, the structures reported in both the previous manuscripts and this current one allow us to confirm the recognition pattern of the TSRs (with a second example in our current manuscript reinforcing the recognition mode of PoFUT2 on different TSRs), and offer a crucial tip for the rearrangement of the TSRs in the putative ternary complex. So, we took *Hs*PoFUT2-GDP-fucose complex and incorporated the *Hs*TSR1 and *Rn*TSR4 as indicated in the main text. The protein was prepared using the Schrödinger software package and the best pose of the docking, in each case, was immersed in a rectangular water box at least 10 Å thick from the protein, and the required Na^+/Cl^- ions were added to render the system electrically neutral.

The corresponding ternary models PoFUT2-GDP-fucose-*Hs*TSR1 and PoFUT2-GDP-fucose-*Rn*TSR4 were built for simulations by first equilibrating the systems according to the following steps: (1) 5000 steps of energy minimization (2500 steps using the steepest descent algorithm followed by 2500 steps of conjugate gradient) with the protein and the ligands restrained; (2) 5000 steps of energy minimization (2500 steps of steepest descent + 2500 steps of conjugate gradients) with no restraints; (3) heating from 0 to 300 K during 100 ps of MD with the protein and the ligands restrained, and (4) equilibration of the system for additional 100 ns of MD with no restraints. Classical MD simulations were then performed using periodic boundary conditions, with long-range interactions treated with particle mesh Ewald (PME)[6] and short-range interactions truncated at a cutoff radius of 8 Å. Chemical bonds involving hydrogen atoms were kept rigid with SHAKE,[7] and a time step of 2 fs was employed to solve the equations of motion. The temperature was kept constant at 300 K with the Langevin thermostat using a collision frequency of 1.0 ps−1. Pressure was as kept constant only during the 100 ps equilibration step via the

Berendsen barostat with a relaxation time of 1.0 ps. Production simulations were performed at constant volume. The simulations were carried out with Amber20[8] using the Amber ff14SB40[9] and gaff2 force fields during 250 ns. Water was modeled with the TIP3P model.[10]

After constructing and preparing the corresponding ternary complexes by replacing the mutated atoms and residues, classical unrestrained MD simulations of up to 1 microsecond were performed to obtain equilibrated systems. Those in which the reacting residues were within a sphere of maximum 4 Å were further optimized without any restraint at QM/MM level. Restrained simulations keeping the reacting residues in a sphere of 3 Å were also carried out to generate molecular configurations depicting the different possible scenarios: (i) deprotonation of Ser17 for further attack to the anomeric carbon of the fucose (a two-step process) or (ii) a concomitant proton-transfer from Ser17 to Glu54 while phosphate (aided by Arg294) is leaving (a concerted reaction). All the generated structures were equilibrated by at least 500 ns and then were further optimized without any restraint at QM/MM level.

The equilibrated systems obtained as described above were used as starting structures for QM/MM calculations. A sphere of waters of 24 Å, centered on the anomeric carbon of the fucose unit was added to simulate the solvent. The system was partitioned into quantum and classical regions. The quantum region comprised the side chain of Ser17 and Glu50 for *Hs*TSR1, and of Thr17 for *Rn*TSR4, respectively; the side chains of Glu54 and Arg294 of the enzyme and the L-fucose linked by the pyrophosphate unit. Seven water molecules coordinating the pyrophosphate unit were also included in the quantum region which showed a total charge of -3e (result from adding +1e (Arg294), -1e (Glu54), -1e (Glu50), -2e (phosphates of GDP)) for *Hs*TSR1 and -2e (result from adding +1e (Arg294), -1e (Glu54), -2e (phosphates of GDP)) for for *Rn*TSR4. The QM region was treated with BP86 hamiltonian,[11] combined with SVP basis set using correction for the dispersion. The molecular mechanics (MM) part was described using the Amber ff14SB[9] and gaff2[12] force fields, along with the TIP3P water model.[10]

We defined an active region of 12 Å radius centered at the anomeric carbon of the fucose unit. Every residue within the active region was set free, while the remaining residues were frozen during the optimizations. This partial freezing method was recently found to be appropriate for the treatment of large systems.[13] The procedure afforded an optimized

reactant structure that was used as a starting point for the initial reaction path calculation. This initial guess was built using a Reaction Coordinate (RC) that allowed the conversion of the reactant to the fucosylated intermediate.

All cuts between QM and MM regions were applied in aliphatic C-C bonds. The latter was treated at the MM level in order to keep the QM region small and computationally affordable. An electronic embedding scheme was used for the QM/MM treatment. All QM/MM calculations were performed with the modular program package ChemShell,[14] using Gaussian09 to obtain the quantum mechanical (QM) energies and gradients at the DFT level. In previous works, it has been shown that the BP86 functional was a good choice for an initial approach to glycosylation reactions with Leloir's nucleotides and that the QM(BP86/SVP)/AMBER level of calculation provided good geometries and energies at reasonable computational time. The energies and gradients for the MM region were evaluated by DL_POLY,[15] which was accessed through the ChemShell package using the AMBER20 force field. The transition states were located using either the dimer approach or PRF-O optimization procedure, and they were characterized by the vibrational frequency calculation on the QM/MM system.

Reaction coordinates

The reaction mechanism was monitored using reaction coordinates constructed from four distances: *d1*($C1_{Fuc}$–O_{THr17}), defined as the distance between the anomeric carbon C1 and the nucleophile oxygen O_{Ser} of the acceptor hydroxyl group. This distance represented the nucleophilic attack of the serine acceptor on the anomeric carbon of the donor GDP-fucose, resulting in the formation of the new β-glycosidic linkage. The second distance, *d2*($C1_{Fuc}$–$O1_{Phos}$), was defined as the distance between the anomeric carbon C1 and glycosidic oxygen O1 of the pyrophosphate unit of GDP-fucose, representing the release of the leaving group. The third, *d3*(H_{Thr17}–O_{Thr17}), and fourth *d4*(H_{Thr17}–O_{Glu52}), distances represented the H-transfer from the nucleophile hydroxyl of Thr17 to the oxygen of the catalytic base, Glu54 (Figure S3a). Four reaction coordinates were defined as rc1 = d4-d3; rc2 = d2-d1; rc3 = d2+d4-d3-d1 and rc4=d1. The energy profile of the reaction path is determined by the potential energy surface scan. The reaction coordinates rc1 and rc4 were changed by +0.2 Å increments (between 0.9 and 1,8 Å for rc1 and between 1.7 and 2.9 Å for rc4). Reaction coordinates rc2 and rc3 were changed by +0.2 Å increments. In the reverse scan, the reaction coordinate was varied by 0.2 Å increments. All atoms in

both QM and MM layers were free to move in the geometry optimization calculations using BP86-G3/SVP level of theory.

Energies

Table S2. QM/MM Energy barriers (kcal mol^{-1}) for the S_N2 reaction pathway calculated at various levels of theory using QM6 and with the MM subsystem characterized with AMBER20

Entry	level of theory	barrier *Hs*TSR1	barrier *Rn*TSR4
1	bp86-d3/6-311+G(d,p)	17.2	13.0
2	bp86-d3/cc-pvtz	17.7	14.2
3	bp86-d3/def2tzvp	17.9	14.0
4	bp86-d3/aug-cc-pvtz	18.3	13.8
5	tpsstpss-d3/6-311+G(d,p)	18.2	13.1
6	tpsstpss-d3/cc-pvtz	19.0	14.5
7	tpsstpss-d3/aug-cc-pvtz	19.1	13.9
8	tpsstpss-d3/def2tzvp	19.0	14.2
9	blyp-d3/6-311+G(d,p)	17.1	11.5
10	blyp-d3/cc-pvtz	17.5	12.6
11	blyp-d3/aug-cc-pvtz	18.2	12.7
12	blyp-d3/def2tzvp	17.8	12.4
13	bpw91/6-311+G(d,p)	18.6	12.5
14	b3lyp-d3/6-311+G(d,p)	23.4	16.1
15	b3lyp-d3/cc-pvtz	23.4	17.3
16	b3lyp-d3/def2tzvp	23.6	17.1
17	m06-2x/6-311+G(d,p)	30.0	22.3
18	m062x/cc-pvtz	30.2	23.6
19	m062x/aug-cc-pvtz	30.0	23.1
20	m062x/def2tzvp	29.7	22.9

Supplementary Figures

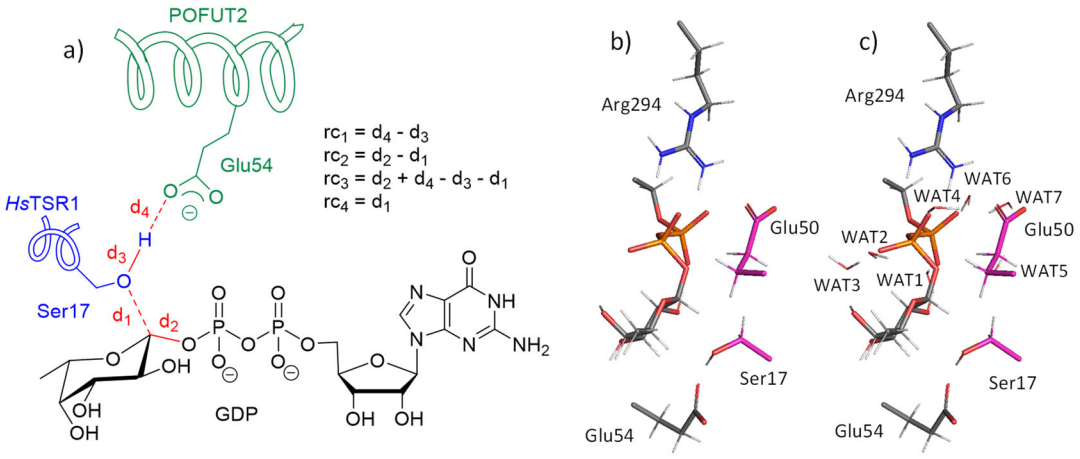

Figure S3. (a) Reaction coordinates. (b) QM region without water molecules for *Hs*TSR1 (82 atoms). (c) QM region incorporating seven water molecules for *Hs*TSR1 (103 atoms).

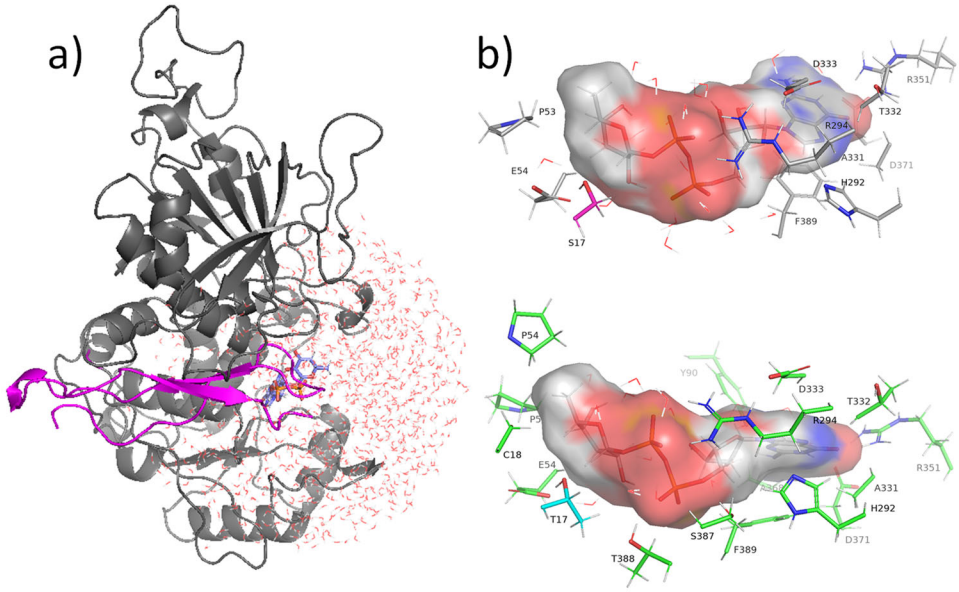

Figure S4. **Optimized reagents**. (a) Full view of the optimized reagent for the glycosylation reaction of *Hs*TSR1 (magenta). (b) Detail of residues of PoFUT1 interacting with GDP-fucose in the optimized reagent corresponding to glycosylation of *Hs*TSR1 (top) and *Rn*TSR4 (bottom. The Connolly surface for GDP-fucose reflects the high accessibility (red) of the solvent to the pyrophosphate unit in both cases in agreement with X-ray and MD simulations. Note that interactions are very similar for optimized reagents (this figure) and MD simulations of ternary complexes (Figure 2b,c of the main text).

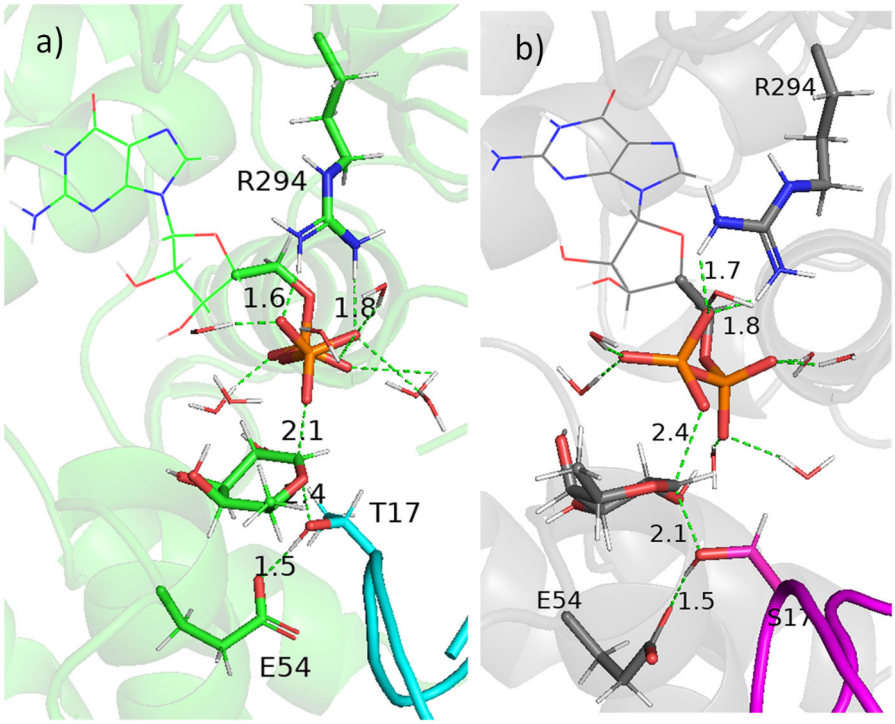

Figure S5. (a) Optimized transition structures for the glycosylation reaction of (a) *Rn*TSR4 (cyan) and (b) *Hs*TSR1 (magenta)

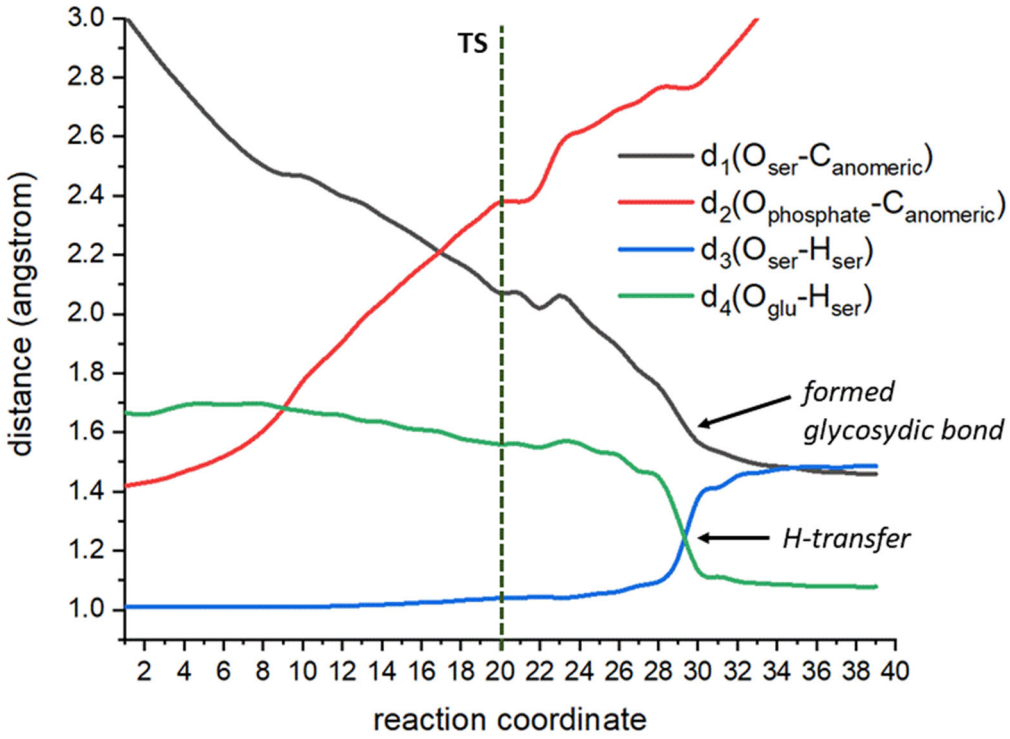

Figure S6. Evolution of the relevant distances along the reaction coordinate

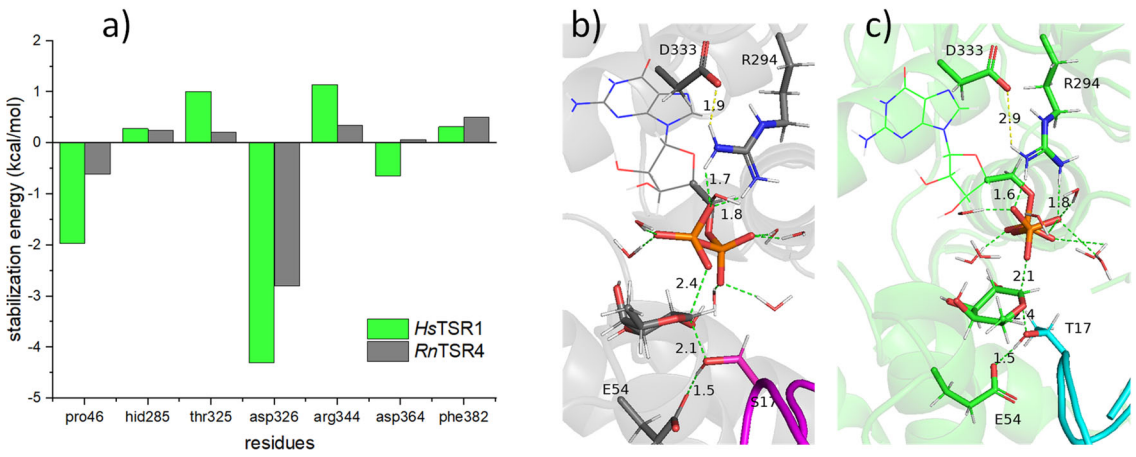

Figure S7. (a) Charge deletion analysis (b) and (c) transition structures of *Hs*TSR1 (a) and *Rn*TSR4 (b) showing unfavourable interaction of Asp333.

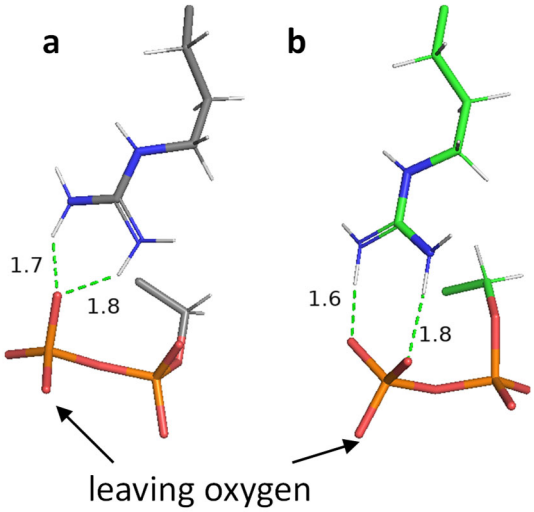

Figure S8. Detail of the different coordination of Arg294 to the leaving pyrophosphate unit. a) for the glcosylation of *Hs*TSR1. b) for the glycosylation of *Rn*TSR4.

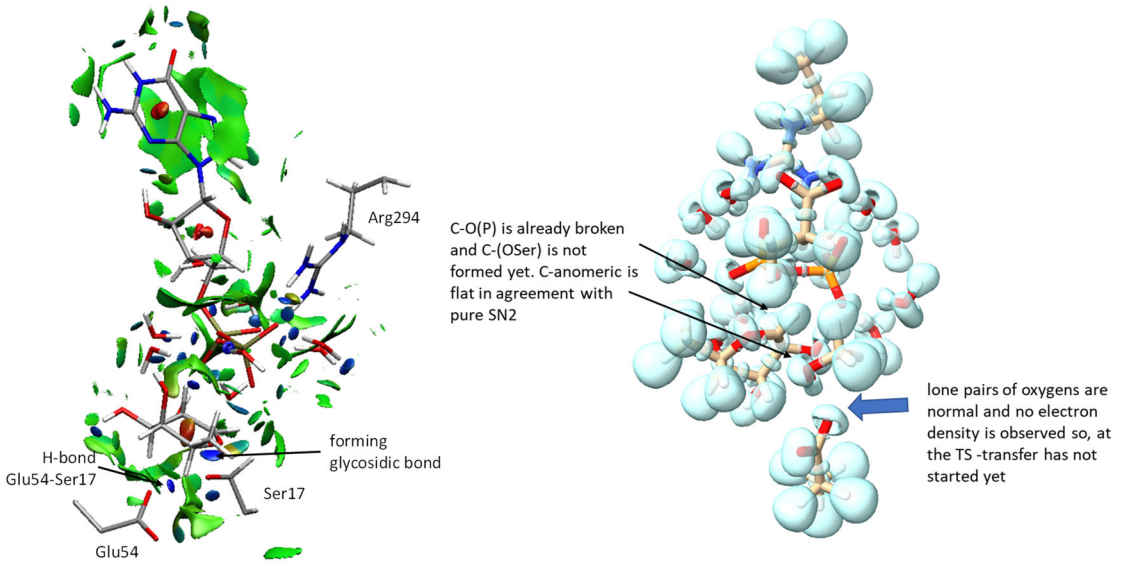

Figure S9. **Topological analyses,** Left: NCI calculations on the optimized transition structure **TS1** for *Hs*TSR1. Blue area corresponds to strong NCI (H-bond, halogen, etc.); green area corresponds to van der Waals interactions and red area corresponds to repulsive interactions (steric clashes, etc.). Right: ELF analysis of the optimized transition structure **TS1** for *Hs*TSR1. The clear localization of lone pairs of oxygen atoms of Glu50 corroborated that H-transfer has not started yet.

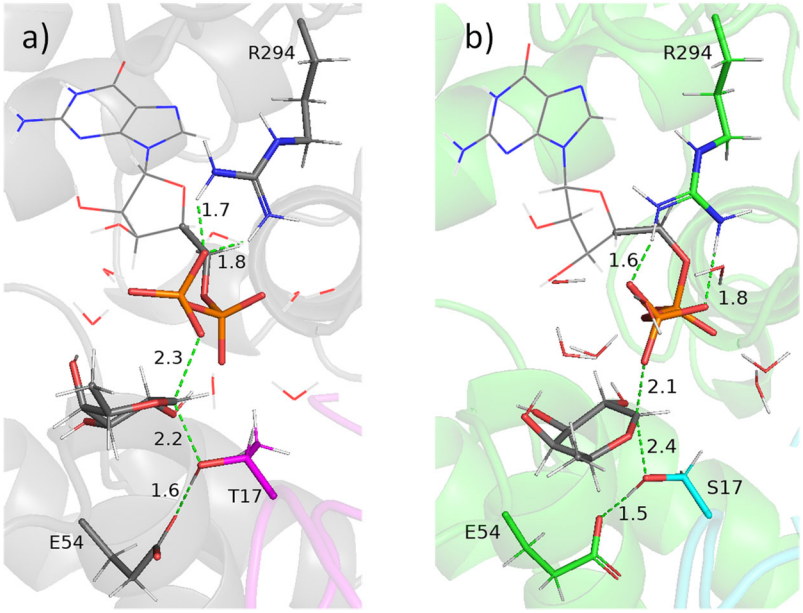

Figure S10. Optimized transition structures for the glycosylation reaction of *Hs*TSR1 S17T mutant (magenta) and *Rn*TSR4 T17S mutant (cyan)

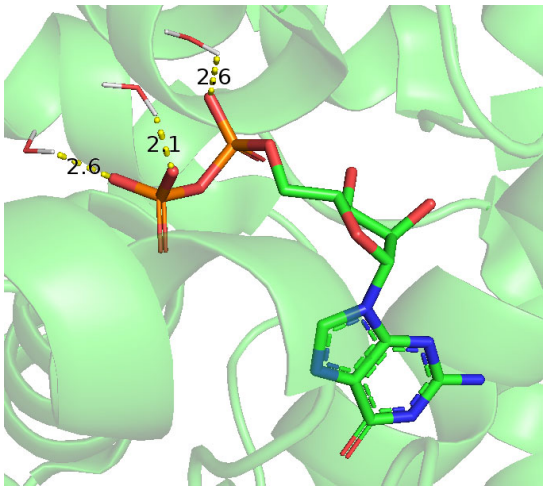

Figure S11. X-ray structure of complex formed between *Ce*PoFUT2, GDP and *Hs*TSR1 (PDB ID: 5FOE). Water molecules are located coordinating phosphate oxygens in the same way as that found in the computational model, supporting the exposure of the phosphate unit and the role of waters in the catalytic mechanism.

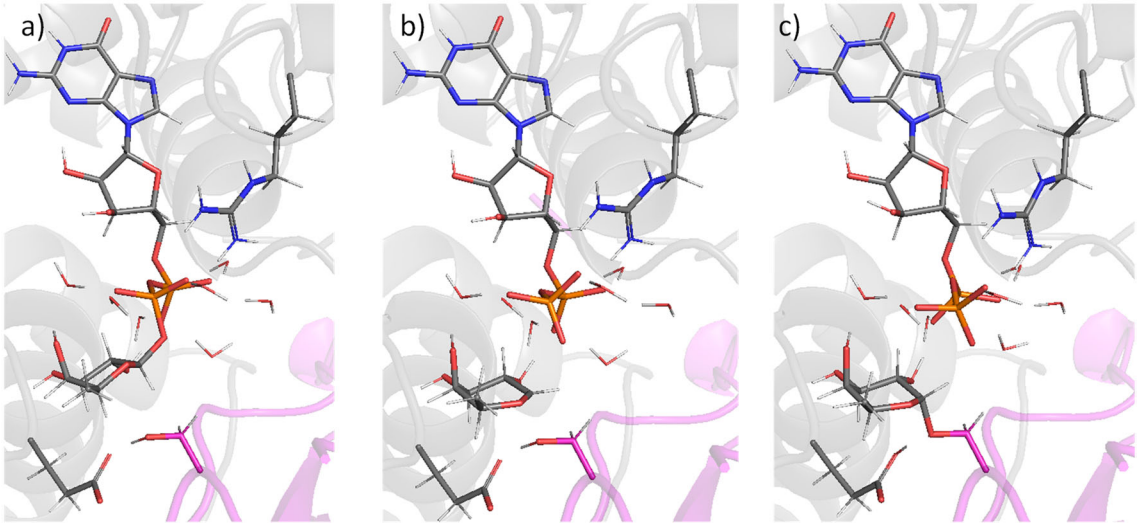

Figure S12. Optimized reactant (a), transition structure (b) and product (c) for the glycosylation reaction of *Hs*TSR1 (magenta).

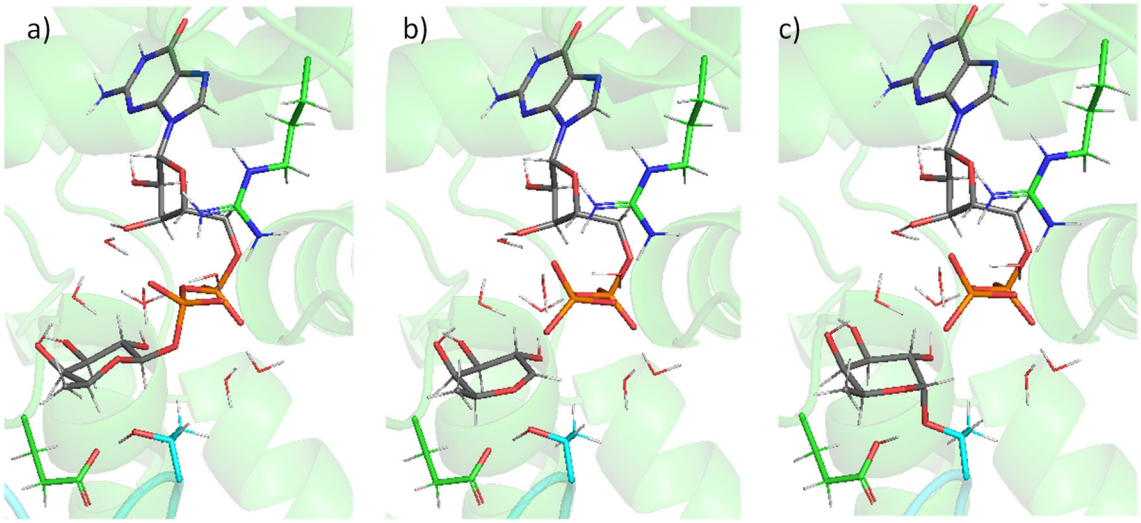

Figure S13. Optimized reactant (a), transition structure (b) and product (c) for the glycosylation reaction of *Rn*TSR4 (cyan)

References

[1] Valero-Gonzalez, J.; Leonhard-Melief, C.; Lira-Navarrete, E.; Jimenez-Oses, G.; Hernandez-Ruiz, C.; Pallares, M. C.; Yruela, I.; Vasudevan, D.; Lostao, A.; Corzana, F.; Takeuchi, H.; Haltiwanger, R. S.; Hurtado-Guerrero, R. A proactive role of water molecules in acceptor recognition by protein O-fucosyltransferase 2. *Nat. Chem. Biol.* **2016**, *12*, 240-246.

[2] Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132.

[3] Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235-242.

[4] Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

[5] Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355-367.

[6] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N·logN) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

[7] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327.

[8] Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; A.W. Goetz, R. H.; Izadi, S.; Izmailov, S. A.;

Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. *University of California, San Francisco* **AMBER 2020**.

[9] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

[10] 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.* **1983**, *79*, 926.

[11] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.* **1988**, *38*, 3098-3100.

[12] a) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Mod.* **2006**, *25*. b) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general AMBER force field. *J. Comput. Chem* **2004**, *25*, 1157-1174.

[13] Calixto, A. R.; Ramos, M. J.; Fernandes, P. A. Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms. *J. Chem. Theory Comput.* **2017**, *13*, 5486-5495.

[14] a) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell - a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 101-110. b) Lu, Y.; Farrow, M. R.; Fayon, P.; Logsdail, A. J.; Sokol, A. A.; Catlow, C. R. A.; Sherwood, P.; Keal, T. W. Open-Source, Python-Based Redevelopment of the ChemShell Multiscale QM/MM Environment. *J. Chem. Theory Comput.* **2019**, *15*, 1317-1328.

[15] Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: an Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113*, 11856-11865.